Amazon Reviews: VU DOPP Management-Report

Maximilian Kleinegger, Martin Wustinger, Tobias Grantner, Alexander Lorenz

1 User satisfaction based on Star Rating and Polarity Sentiment

Are Amazon product reviews generally more positive or negative for certain categories compared to others?

A reliable indicator of customer appreciation for a product is often found in positive reviews, while dissatisfaction is typically reflected in negative feedback. In exploring this aspect, we considered two metrics: star ratings, a direct measure of user satisfaction, and polarity sentiments derived from the *summary* and *reviewText* using VaderSentiment[1] and TextBlob[2], providing subtler insights into customer sentiment.

Analyzing the distribution of star ratings across various categories reveals that certain categories attract more positive reviews than others. However, even in the case of the categories with the lowest ratings, approximately half of the reviews still have a 5-star rating. Hence, reviews tend to be positive in general. A similar observation can be made from the sentiment analysis, where all categories exhibit a relatively positive average sentiment.

While star ratings and sentiment values show substantial similarity in most categories, some outliers are notable. Interestingly, many outliers in star rating distributions align with those observed in sentiment analysis, and vice versa. The category with the least favorable performance based on reviews is Software. This could be attributed to the challenges faced by software, such as a tendency to bugs, installation difficulties, and configuration complexities, leading to a less satisfactory user experience — issues that are typically not associated with simpler products like shirts or books.

Conversely, the categories that stand out as top performers, identified as outliers in both sentiment and star rating analyses, are Gift Cards and Digital Music. These categories typically function seamlessly upon purchase, minimizing the likelihood of customer dissatisfaction.

2 The subjectivity of reviews differs among categories

Are reviews more subjective for some categories of products than for others, based on sentiment analysis?

Our in-depth analysis reveals clear differences in how subjective customers express opinions across various product categories. This finding holds significance for managerial decisions and strategic planning.

Additionally, we notice a substantial contrast in how subjective the summary versus the review text tends to be in Amazon reviews. It's important to highlight that relying solely on the summary might not provide an accurate picture due to its inherent objectivity and uniform distribution across categories. To gain deeper insights, focusing on the review text is crucial. The case of Gift Cards, in particular, stands out as an outlier, where the summary contradicts the sentiments expressed in the review text.

Consider the example of Gift Cards versus Software categories, where we observe a significant discrepancy in subjectivity scores. Showing that the subjectivity does not depend on the user itself, rather it is a combination of product (-category) and user. Understanding these variations is valuable for managers as it offers a more nuanced perspective on market dynamics.

This subtle understanding of customer sentiments can directly impact decision-making, marketing strategies, and product development. By acknowledging the subjectivity differences, managers can tailor their approaches to better align with customer expectations, ultimately contributing to improved customer satisfaction and potentially driving higher sales.

[1] https://vadersentiment.readthedocs.io/en/latest/
[2] https://textblob.readthedocs.io/en/dev/

3 Identify most important aspects of product categories

Which aspects of different categories of products are the most important in the reviews in regard to their appearance in reviews?

To capture and identify valuable insights about products' characteristics, unique selling points, or flaws in various product categories, we evaluated word clouds for each category. This evaluation enables the management to quickly comprehend the most frequently mentioned characteristics. Only nouns, verbs, and adjectives are considered, as they are the most promising part of speeches encapsulating the most crucial information within sentences. Due to computional constraints 0.1% of each category which accumalates to roughly to overall count of 230.000 reviews.

The wordclouds reveal that certain characteristics are product-related and unique to specific categories, highlighting that fashion-related categories revolves around fit, size, and appearance, while beauty-related categories are more focused on facial care including hair, face, and lips. This insight suggests the potential for deriving domain-specific aspects from these observations. Interestingly, numerous positive emotions, including terms like good, amazing, great, and love, are frequently expressed in each category, thereby enforcing the observations outlined in section 2. The analysis also reveals that, in categories like books, cell phones, and accessories, UTF-8 encodings and HTML tags are present, suggesting possible embeddings of pictures. However, across all categories, there seems to be none crucial aspects that play a paramount role have been indentified. Therefore, further analysis is considered.

When performing named entity recognition using spaCy[3] and plotting the proportions of the entity labels for each category, it can be seen that entities contribute differently depending on the category. Based on accumulation of organizational values, gift cards, cell phones and accessories are more brand driven compared to products from appliances and amazon fashion. Consequently, organizational values related to brands can be a defining aspect within a product category. While spaCy has the capability to predict monetary entity values, it appears that none are detected across any category. One can conclude that this aspect seems to be not important. The fact that the dates are contributed in every category can help the management to identify and understand seasonal purchase behavior. Nevertheless, overall, the product labeled under certain categories exhibited a poor proportion. Named entity recognition helps to identify aspects, although it appears to have limitations,we might anticipate alternative outcomes.

4 Star Rating Prediction

To what extent can one predict the star rating from the text of a review?

Rating a product on a scale of one star to five stars is a way for customers to express their satisfaction with their purchase. As we have seen in the previous sections, the star rating of sales made through Amazon holds valuable information and can be used by sellers to gain insights about their products' demand and quality.

However, many sellers distribute their products through multiple channels and not all of them might have the same rating system. For example, a seller could simultaneously sell their products on Amazon and on their own website. While Amazon uses a five-star rating system, the seller's website provides the user only with the option to write textual reviews. In this case, it would be beneficial for the seller to be able to predict the star rating from the text of the review, in order to gain a universal rating system for all their product reviews across all channels.

We compared models based on different machine learning algorithms like logistic regression, random forests, naive bayes and multi-layer perceptrons. With the multi-layer perceptron model, we were able to train a model that can predict the star rating of a review based on the text of the review with a fairly high accuracy of around 70% and outperforms all other models. While it will not always perfectly predict the star rating of a specific review, it is able to predict the distribution of star ratings of a set of reviews quite well. Predicting the distribution of ratings from a collection of reviews can be invaluable to a seller, indicating the general satisfaction of their customers with a product when such a metric is not available.

[3] https://spacy.io/