Sentiment Analysis of Amazon Reviews: Insights from Consumer Feedback.

Heleyna Tucker
Department of Statistics, STATS 507
University of Michigan
Ann Arbor, United States
heleyna@umich.edu

***Abstract*—This paper presents a sentiment analysis of Amazon reviews for cell phone and accessory products. By leveraging the English DistilRoBERTa-base model, review sentiments were categorized into emotions such as joy, sadness, and anger. Various statistical and graphical techniques were employed to summarize the data and uncover patterns. The analysis revealed that positive reviews (ratings above 2.5) were primarily associated with joy and neutral emotions, while negative reviews (ratings below 2.5) exhibited higher proportions of sadness and anger. These insights demonstrate how sentiment analysis can help companies understand consumer feedback and improve product offerings. Additionally, specific subsets of reviews were examined to uncover detailed patterns using keyword-based filtering.**

I. Introduction

Online shopping has revolutionized consumer behavior, with platforms like Amazon offering diverse product ranges. Customer reviews play a crucial role in shaping purchase decisions, providing valuable insights into product performance. This paper explores how sentiment analysis can be employed to extract meaningful patterns from Amazon reviews, particularly for cell phones and accessories.

Sentiment analysis enables companies to gauge consumer sentiment, identify trends, and enhance product design. By analyzing review data, this study aims to uncover emotional patterns in feedback, providing actionable insights for affiliated businesses. Furthermore, this project aligns with professional development goals, simulating real-world data analysis tasks often required in industry roles.

A. Existing Literature

A variety of studies have leveraged sentiment analysis to interpret consumer opinions. For instance, in *Computers in Human Behavior*, Butt et al. conducted a psycholinguistic analysis of rumor and non-rumor tweets using the same sentiment analysis model. They found that “fear and sadness are the two most instigated emotions in the rumor tweets” and that responses exhibited “anger and surprise” [1]. While the study focused on Twitter data, its methodological framework provides inspiration for analyzing Amazon reviews. Unlike the Twitter study, this project emphasizes the use of visualizations, aiming to create comprehensive graphical displays that effectively convey insights to stakeholders.

II. Methods

The sentiment analysis employed the English DistilRoBERTa-base model, a pre-trained transformer-based model for multi-emotion classification. Given text input, the model categorizes emotions into anger, disgust, fear, joy, neutral, sadness, and surprise, with outputs representing proportional scores summing to one. Trained on a diverse dataset including social media, dialogues, and reports, the model achieves an evaluation accuracy of 66% (random chance is 14%).

The dataset, sourced from Kaggle, comprises approximately 195,000 Amazon reviews for cell phone and accessory products. To streamline analysis, reviews were grouped by ‘asin‘ (Amazon product codes). For each product, summary reviews were truncated to meet the model’s 512-token limit before sentiment classification. Data transformations, including the creation of helpfulness metrics and categorical grouping (e.g., positive and negative ratings), enabled a detailed exploration of emotional patterns.

Visualization techniques such as bar charts, boxplots, scatterplots, and heatmaps were implemented using matplotlib and seaborn libraries. Additionally, keyword-based filtering was applied to isolate subsets of interest, such as reviews mentioning ”headphones” or ”earbuds.”

III. Results

After running the model on the pooled summary reviews by asin, about 10,000 observations were in this compressed dataset. Numerous discoveries can be made about the data through graphical displays and numerical manipulation. For example, various products labeled as bad can be looked at with their individual emotional scores. Some bad products had more fearful emotions overall compared to another bad product that displayed more sadness in their review summaries. Below in Figure 1 is a bar chart that includes the average score across all emotions separated by good and bad reviews. Joy is the leading emotion for good reviews while sadness leads to bad reviews with anger, disgust, and fear following behind. Interestingly, surprise is more involved in good reviews than bad reviews. This implies surprise could be somewhat of a swing emotion for both bad and good reviews, explored further later in the project. It is important to note that there are significantly more good (10,310) than bad (119) products. Additionally, how

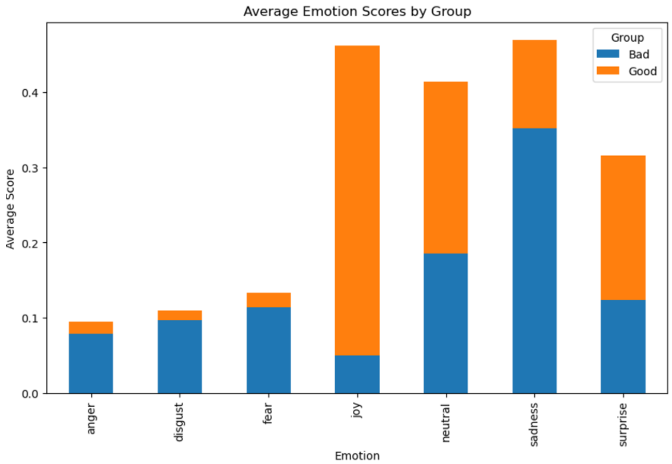

Fig. 1. Good and bad grouped emotion Scores for compressed asin review summary data for phone cases and accessories.

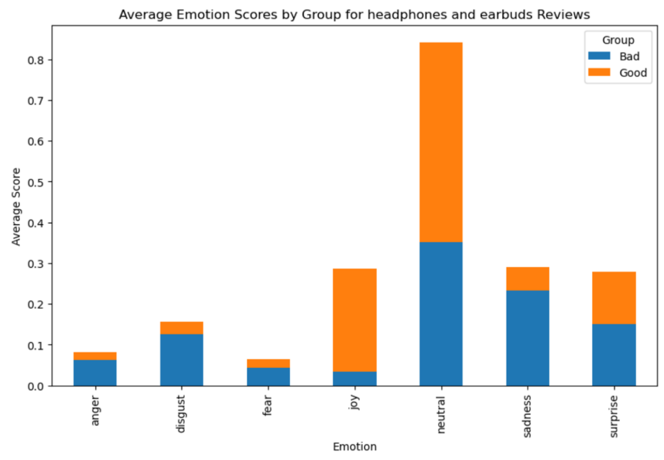

Fig. 3. Good and bad grouped emotion scores for headphone and earbuds specific Amazon reviews in phone case and accessory data.

each emotion affects each other is another point of interest shown in the heatmap below in Figure 2. It can be concluded

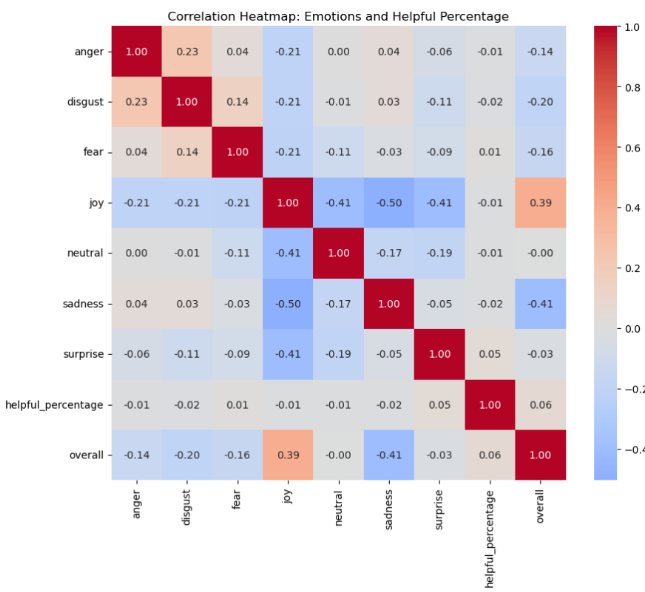

Fig. 2. Emotion Correlation heatmap.

that anger, for example, is the most correlated with disgust, while not being very correlated with sadness even though they are both categorized as generally bad emotions. Also noted, emotion did not seem to have a very large effect on helpfulness percentage for the product (how helpful each review was, on average).

Next the data was subsetted into only reviews that included words 'headphones' or 'earbuds'. This quartered the amount of observations to around 50,000. A similar bar chart as the compressed dataset earlier was made, shown in Figure 3. It can be observed that both good and bad reviews were dominated by neutral words. Joy was a leading emotion for good reviews while bad reviews were led by sadness, surprise, and disgust after neutral. Interestingly, surprise is fairly equal on average for both good and bad reviews. In addition to that, this subset of data had more sadness in bad reviews than in summary review data for the whole sample size. The bad reviews also had less joy and more disgust compared to the compressed summary results.

The worst rated products can also be looked into within the subset (with a bad rating review ratio over 0.5). Fifteen products were found to be above this ratio. The emotion scores for each product were visualized using a heatmap in Figure 4. The results in Figure 4 show that each product

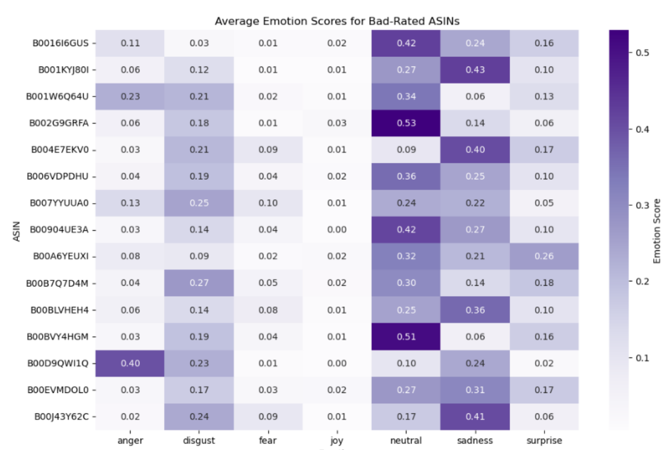

Fig. 4. Heat map of emotion proportions for 15 worse rated overall headphone and earbud products.

has various average emotion scores for their reviews. For example, B00D9QWI1Q has anger as its leading emotion. Neutral and sadness are the leading emotions for these bad reviews. Disgust and surprise are fairly consistent accross all of these products, with little to no fear or joy. Visualizations like this can help companies better understand their specific products and how they are being taken by consumers.

IV. Conclusion

To conclude, results of sentiment analysis can be used in numerous ways. Graphical displays can help better understand results for specific inquiries. For example, individual products of interest, or overall emotions for users. Data manipulation is also important to research all possible discoveries within a given dataset. In this Amazon case and accessory dataset, joy was found to be the leading emotion for good reviews while sadness took over bad reviews, on average. However, looking into specific products allows to get a closer look. Some reviews displayed more disgust or anger overall compared to the average, for example. Surprise was an emotion that was fairly equally divided between good and bad reviews. Researching these emotional scores for product reviews can help developers better understand their products and what they are doing well on and what could be improved to give their consumers the best experience. One notable setback was the word limit on the model, which could have cut some important word data. Possibly some more data manipulation techniques or model alterations could be looked at for future work.

References

[1] Butt, S., Sharma, S., Sharma, R., Sidorov, G., and Gelbukh, A. (2022). What goes on inside rumour and non-rumour tweets and their reactions: A Psycholinguistic Analyses. Computers in Human Behavior, 107345.

[2] Jochen Hartmann, ”Emotion English DistilRoBERTa-base”. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/, 2022.

[3] Kaggle dataset: https://www.kaggle.com/datasets/abdallahwagih/amazon-reviews

[4] Note: ChatGPT was used to help edit the Abstract, Introduction, and Methods section to be more concise.