Abstract

This paper examines customer feelings about electronic products sold on Amazon and uses data from the Amazon Product Reviews Dataset 2020. By examining consumer satisfaction and behavior, researchers uncover key elements that influence both positive and negative customer experiences. The research investigates necessary questions regarding how sentiment shifts with product updates and how consumers commonly express their complaints while also identifying the product features that are most frequently appreciated by users. Machine learning models provide a layered understanding of consumer perspectives through both detailed and concise reviews, and methodologies involve thorough data cleaning and sentiment analysis. The study compares customer satisfaction for purchases made on Amazon and purchases made in brand-specific stores. Results show that the length of reviews and their sentiment provide understanding that these go beyond basic star ratings. Time-based trends and helpful vote analyses reveal that customers engage seasonally and prefer giving brief and impactful feedback. The findings provide actionable understandings for e-commerce and they inform product development strategies while also improving the overall Amazon shopping experience.

1. Introduction

The rapid development of electronic products, from smartphones to laptops, has reshaped consumer expectations, habits, and overall satisfaction. Gaining a deeper understanding of customer satisfaction regarding these products, especially when purchased through Amazon—one of the largest online marketplaces—can reveal valuable insights into customer behavior and preferences. This study leverages data from the Amazon Product Reviews Dataset 2020, which contains in-depth customer feedback on various electronic products. By examining these reviews, the goal is to better understand the factors that drive consumer satisfaction and dissatisfaction, providing meaningful information for manufacturers, retailers, and customers alike.

Analyzing customer satisfaction with electronic products is important for several reasons. First, customer reviews offer a wealth of information, providing unfiltered, genuine perspectives directly from users. These reviews highlight both positive and negative aspects of the products, revealing potential areas for improvement. Second, understanding how consumers perceive electronics can illuminate broader market trends within the tech industry. Lastly, Amazon's unique environment, with third-party sellers and varied delivery options, may influence customer satisfaction differently compared to purchasing directly from brand-specific stores. By focusing on electronics sold via Amazon, this research aims to explore these distinct dynamics.

The analysis is centered around several key research questions. These questions are intended to uncover patterns in customer satisfaction, identify common issues, and ultimately offer a clearer picture of consumer preferences:

➔ Does customer satisfaction change with each new product release? As companies launch new versions of electronic products, it is important to determine whether customer sentiment shifts over time, which may indicate how well these releases meet or fail to meet consumer expectations.

➔ What are the most frequent complaints customers have about electronic products? Identifying common issues raised by consumers can help pinpoint product weaknesses and areas in need of improvement, potentially guiding future developments.

➔ What are the most frequent compliments customers give about electronic products? Positive feedback reveals which product features resonate most with customers, such as design quality, user experience, or specific functionalities. Recognizing these highlights can support branding and marketing initiatives.

➔ What insights can be drawn from long reviews versus short reviews? Detailed reviews often provide more comprehensive insights, such as specific examples or personal stories, whereas shorter reviews may highlight key themes or concerns. Analyzing both types of reviews helps provide a fuller understanding of consumer sentiment.

➔ Is there a difference between customers who purchase electronics from Amazon versus those who buy directly from brand stores? This question seeks to determine if purchasing from Amazon results in different levels of satisfaction compared to buying directly from a brand-specific store. Factors such as product availability, pricing, and service quality may contribute to these differences.

The following sections will discuss the existing literature on sentiment analysis, outline the methods used for data cleaning and processing, and present an exploratory analysis of the reviews, employing various visualization and statistical techniques. Ultimately, this analysis aims to identify patterns, challenges, and opportunities that can guide both manufacturers' product strategies and Amazon's service quality, enhancing the overall customer experience.

1.2 Literature Review

This literature review looks at how different methods utilized in various studies of sentimental analysis can be applied to our research of 'Contextual Sentiment Analysis of Amazon Electronics Product Reviews to Enhance E-Commerce Experience'. It is evident that online reviews are becoming increasingly important for both businesses and consumers[1,2,4]. There is a lot of content available on the internet generated by different users and analyzing this data effectively can help organizations develop better product[1,5,7]. It can also help with targeted marketing, communication which would lead to better customer satisfaction[2,9].

Various studies show that machine learning and deep learning techniques are increasingly becoming popular for sentiment analysis due to their ability to handle large data with high accuracy[1,2,5,6]. Some popular algorithms preferred by researchers include Support Vector Machine (SVM), Naïve Bayes (NB), Long Short-Term Memory (LSTM) and BERT[1,4,8,11,15]. These algorithms can be utilized on labeled data obtained from product reviews to classify the sentiments into positive, negative or neutral[3,5,8]. Further, hybrid-methods which utilize machine learning with dictionary-methods can be deployed for increased accuracy.

After analyzing various papers, Python has been identified as the main programming language for sentimental analysis[1,8]. It is mainly because of the availability of different libraries specific for machine learning and natural language processing, and because of its flexibility[1,8,13,15]. Libraries such as NLTK and TextBlob are popular choices for tasks like breaking down text, tagging parts of speech, and scoring sentiment[1,8,13,15]. There are several other tools available, from commercial platforms to open-source libraries to help us with sentiment analysis[1,2,13,15].

Despite these advantages, research review also points out several challenges faced while analyzing product reviews[1,2,3,6,7]. One main challenge is that product reviews are short and often informal making it difficult to analyze[1,2,3,6,7]. This can be overcome by adapting models to specific domains and using external knowledge.[1,2,5,8,13] Another challenge is that the way people express sentiment can vary by product category[1,2,5,8,9]. A model trained for one product may fail for another product because of this[2,5,8,9,13]. This can be solved by using specific vocabulary and by domain adaptation[1,4]. Also, when the data set is dominated by one sentiment type, for example positive, there are chances that the results can get skewed[1,4]. This can be solved by techniques like oversampling, under sampling and cost-sensitive learning which can help balance the dataset.

The papers also talk about how these methods can be used to analyze product reviews effectively[1,2,3,4,6,7]. Customer satisfaction can be understood from overall sentiment identification, extracting product features helps understand the likes and dislikes of customers, helping product development and marketing[1,2,3,4,6,7]. Tracking sentiment trends can help us understand how opinions change and emerging issues.

Lastly, while providing a deep understanding of methods and challenges in opinion mining, it also provides direction for future research work. It talks about developing models that can adapt to the changing nature of online language, incorporating data from images and videos, and addressing ethical issues like reducing bias and protecting privacy.

2. Data Cleaning

Our data cleaning process focused on ensuring clarity and consistency across key review components to prepare for meaningful analysis. Working with Amazon's extensive review data, we encountered several quality issues, such as missing fields, irregular timestamps, and language inconsistencies. Addressing these was crucial to preserve data integrity and yield accurate insights.

To start, we loaded a subset of 1 million reviews to maintain computational efficiency while capturing a representative data sample. In our create_dict() function, we organized critical fields (rating, title, text, images, timestamps, helpful votes, and purchase verification status) into a structured dictionary. This step helped us systematically handle and access each field for further cleaning and analysis.

Next, we tackled missing values, particularly in the *title* and *text* fields, by using dropna() to exclude entries with null values in these columns. In the case of numerical fields, such as *helpful votes*, we imputed zeros where values were missing, assuming these reviews hadn't yet been rated as helpful by other users.

To enhance our sentiment analysis accuracy, we removed non-alphabetic content, such as emojis, numbers, and symbols, through a preprocessing function (preprocess_text). Using regular expressions and the Natural Language Toolkit (NLTK), we filtered out stopwords and short words, leaving only meaningful terms for analysis. By removing irrelevant characters and stopwords, this step enhances the clarity of the reviews, making it easier for sentiment models to classify sentiments accurately. These preprocessing steps are essential to maintain data consistency, improving the quality of subsequent analysis by focusing on meaningful words related to customer sentiment.

One notable challenge we encountered was with timestamp inconsistencies. Some timestamps reflected dates in the distant past, likely due to data corruption or encoding errors. By filtering timestamps, we focused only on plausible entries within a realistic timeframe, from approximately 1996 onward. We converted valid timestamps to a standard date format using pd.to_datetime(), allowing us to analyze seasonal trends and shifts over time in review data.

3. Methodology

This study employs Natural Language Processing (NLP) techniques to analyze customer sentiment in Amazon electronics reviews, focusing on shifts in satisfaction, common complaints, compliments, and comparisons between long and short reviews. For sentiment analysis, we used TextBlob, a lightweight library that scores text polarity based on the presence of positive and negative words. TextBlob's simplicity enabled efficient processing of large volumes of review data, allowing us to capture general sentiment trends with minimal computational demands. To identify key themes in reviews, we applied word frequency analysis, which highlights commonly used terms to reveal prominent complaints and compliments. These NLP techniques were selected for their ability to process extensive text data quickly and to deliver results directly relevant to our research questions. Their simplicity allows for rapid analysis without the complexity of model training, making them ideal for handling over a million reviews while still yielding meaningful insights.

4. Exploratory Analysis

Customer reviews offer valuable insights into product performance, user satisfaction, and consumer behavior, making them an essential resource for businesses. Our group analyzed a large dataset of Amazon electronics reviews to explore how customers express their experiences, evaluate sentiment trends, and uncover connections between review characteristics and ratings. The goal was not just to crunch numbers but to understand patterns that could inform product development and ultimately improve the customer experience.

Throughout this analysis, we noticed two major benefits: it revealed common pain points and highlighted what customers appreciated the most. Reviews varied widely—from brief comments to in-depth narratives. One thing we learned was how both the length of reviews and specific words used provide insight into customers' priorities. For example, frequently used words in positive and negative reviews gave us a clearer sense of what mattered to users beyond just the star ratings. These details helped us see how customer feedback offers more depth than it might seem at first glance.

We also found that sentiment analysis captured emotional undertones that star ratings alone often missed. Sentiment trends—like slightly positive sentiment in 1-star reviews—showed us that constructive feedback doesn't always align with low ratings. These insights gave us a more nuanced picture of the relationship between ratings and customer sentiment. Given the scale of the dataset, we sampled 1,000,000 reviews to manage computational resources while still ensuring the sample was representative. The findings from this exploratory phase have given us practical insights and set the foundation for deeper research into customer behavior trends and product performance.

4.1 Distribution of Review Lengths

Review lengths varied widely. On average, they were 26 words long, with a median length of just 13 words. The longest review reached 2,658 words. Interestingly, the distribution showed that most reviews were concise—well under 100 words. This trend, shown in the figures below, suggests that users tend to leave quick feedback rather than detailed descriptions. This finding aligns with the research question on understanding customer satisfaction and dissatisfaction drivers—brief reviews likely capture top-of-mind reactions (positive or negative), while more detailed reviews may indicate significant investment in the product, revealing critical insights into customer preferences and concerns. For us, this reinforced the importance of developing tools that can extract insights efficiently from shorter reviews without neglecting the few, but meaningful, longer ones.

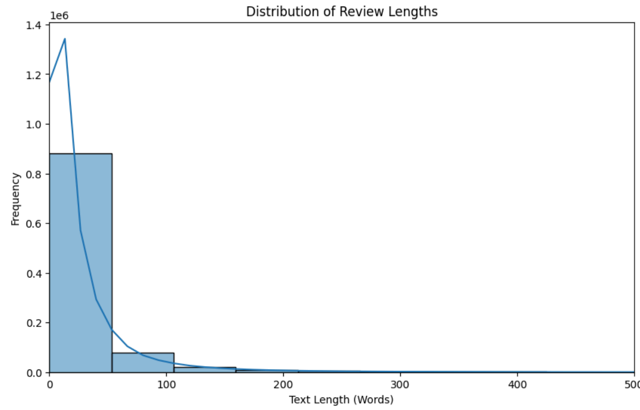

Figure 1

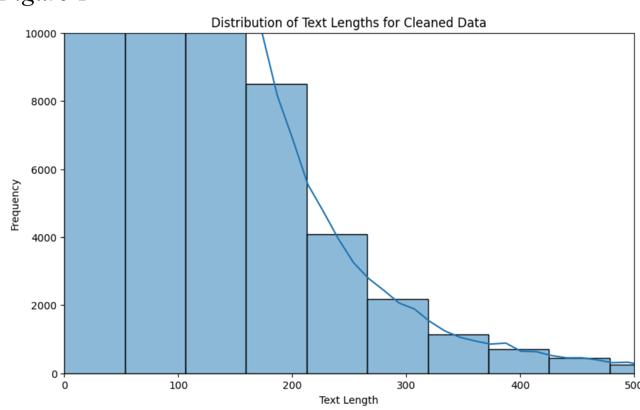

Figure 2

4.2 Comparison of Review Lengths and Ratings

We also compared review lengths across different ratings. Surprisingly, there wasn't a strong correlation between length and rating, shown in Figure 3. Long reviews appeared in both 1-star and 5-star categories, often either deeply criticizing or enthusiastically praising the product. This showed us that length alone doesn't predict sentiment, confirming the need for sentiment analysis alongside ratings. It was a good reminder that customer feedback comes in many forms, and even short reviews can carry valuable insights. This observation is closely related to the research question on identifying frequent complaints and compliments in electronic products. Shorter reviews are more likely to capture immediate impressions, while longer reviews provide in-depth feedback, helping manufacturers and marketers understand nuanced product strengths and areas for improvement.

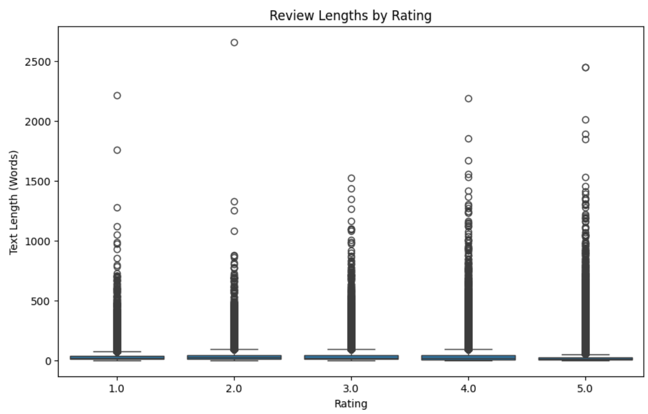

Figure 3

4.3 Most Common Words for Each Rating Category

We were curious about the words people used at different rating levels. For 5-star reviews, words like *"great," "use,"* and *"works"* appeared frequently, suggesting that functionality plays a big role in customer satisfaction, shown in Figure 4. The presence of emotional words like *"love"* pointed toward a personal connection with the products. On the other hand, negative reviews were less consistent in vocabulary, though terms related to defective or misleading products showed up repeatedly. This indicated potential issues with quality control or unmet product expectations, in addition to contributing to findings to answer our research questions related to frequent compliments and complaints from customers.

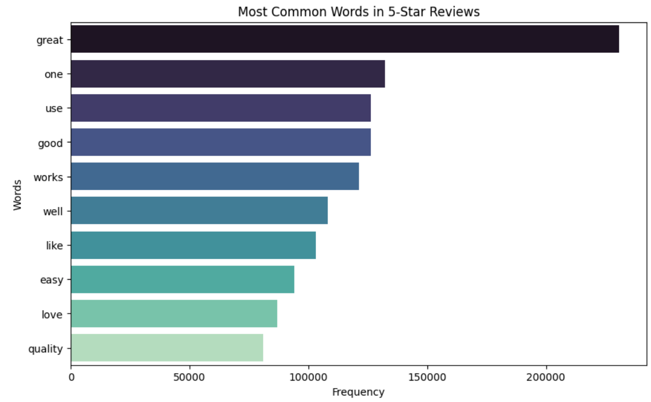

Figure 4

4.4 Most Common Words in Titles vs Reviews

A comparative analysis of common words in review titles versus review texts showed that similar terms appear in both, though titles tend to be more concise and focused on keywords. For example, words like *"great," "works,"* and *"product"* appeared frequently in both review titles and full texts, as shown in Figure 5, but titles contain fewer emotional or contextual terms. This indicated that while titles capture the core message, full texts provided more depth and personal experience. The disparity also suggested that analyzing both the title and review text together can give a more complete understanding of customer sentiment. This observation relates to our research question by understanding the comprehensive insight provided to customers when reading a longer versus shorter review.

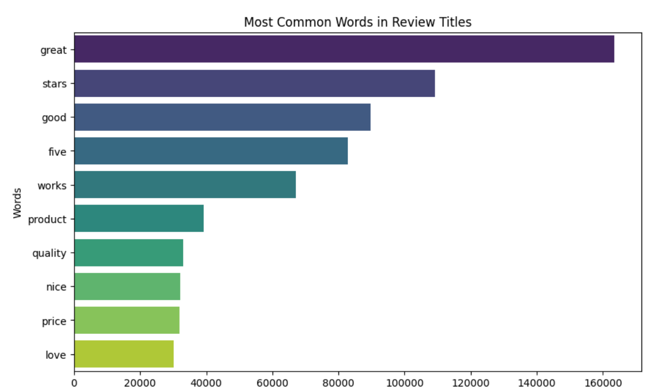

Figure 5

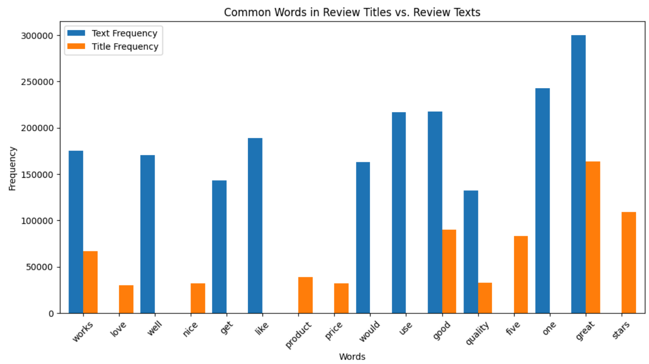

Figure 6

4.5 Sentiment Analysis vs Review Rating

Polarity-based sentiment scoring techniques assign sentiment values based on the polarity of words within a review, providing a basic metric for assessing customer sentiment. This analysis revealed that positive sentiment aligns closely with higher ratings, with a clear increase in sentiment polarity as ratings rise. Reviews with 5-star ratings show the highest sentiment scores, while those with 1-star ratings exhibit negative polarity. Figure 7 outlines a boxplot of sentiment scores by rating, which shows that even within the same rating category, there is variability in sentiment—some reviews with lower ratings still carry mildly positive sentiments, possibly indicating constructive feedback. This finding supports the idea that sentiment analysis provides a nuanced view that complements star ratings, highlighting the importance of textual insights in customer feedback. This approach aligns with research questions that explore the relationship between star ratings and the sentiment conveyed in the review text, offering a straightforward way to quantify sentiment trends over time.

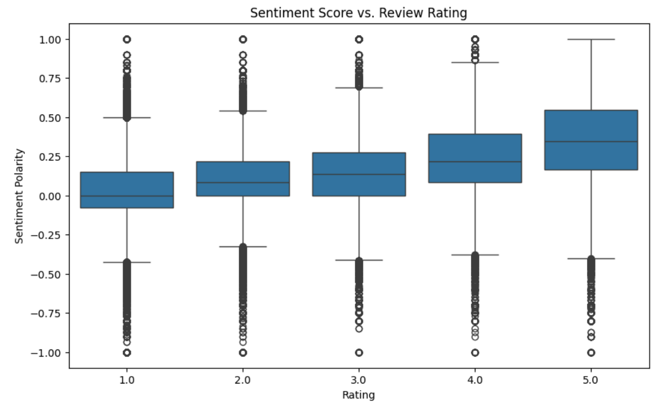

Figure 7

5. Time-Based Analysis and Trends

The time-based analysis revealed interesting patterns in the distribution of reviews over different years and months. The initial timestamps in the dataset contained anomalies, like dates far in the past, suggesting some reviews might have corrupted or incorrect time data. After filtering and cleaning the timestamps, we found that most reviews spanned from 1996 to 2023, with a large concentration of reviews in more recent years.

5.1 Distribution of Year/Month Reviews

This trend aligns with increased online shopping behavior over time, especially following the widespread adoption of e-commerce platforms like Amazon. Monthly trends show higher volumes during November and December, likely due to Black Friday and holiday shopping peaks. This suggests that customer engagement is closely tied to seasonal events, highlighting the importance of time-sensitive product campaigns and support strategies during these high-traffic periods. Moreover, the increasing number of reviews in recent years suggests that customer feedback has become an integral part of the online shopping experience, with customers relying more on previous reviews for purchasing decisions. This emphasizes the need for companies to monitor trends in feedback over time to respond promptly to shifting customer preferences and potential product issues.

5.2 Temporal Sentiment Analysis

To examine how customer sentiment has evolved over time, we conducted a temporal sentiment analysis, tracking the average sentiment polarity of reviews across different years. The plot in Figure 8 illustrates these trends, showing sentiment fluctuations from the early 2000s to 2023. Overall, the graph reveals notable sentiment changes, with peaks around certain periods that may correspond to product releases, updates, or broader industry trends. From 2000 to around 2015, the average sentiment polarity shows significant variability, likely reflecting customers' mixed experiences with early consumer electronics and ongoing advancements. Post-2015, there is a gradual increase in sentiment, reaching a peak in later years before a slight decline, which may correlate with heightened customer expectations or increased competition in the electronics market. This analysis highlights how temporal factors can influence customer satisfaction, suggesting that shifts in technology, product quality, and market dynamics play roles in shaping customer experiences over time.

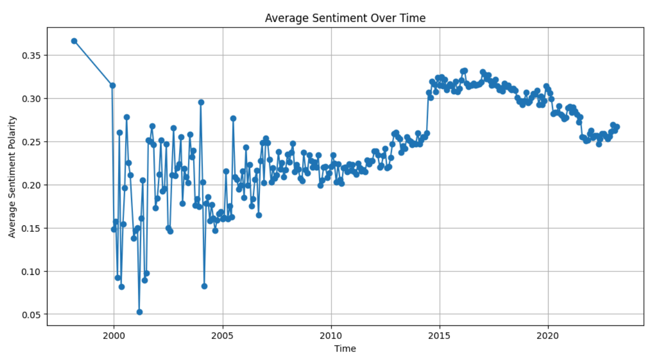

Figure 8

6. User Behavior and Helpful Votes

The relationship between helpful votes and text length offered intriguing insights into user behavior. As visualized in the scatterplot, there is a general trend where longer reviews tend to receive more helpful votes, suggesting that customers perceive detailed reviews as more informative and reliable. However, Figure 9 also revealed diminishing returns: while helpful votes increase with text length, extremely long reviews (above 500 words) show a decline in engagement, possibly due to reader fatigue.

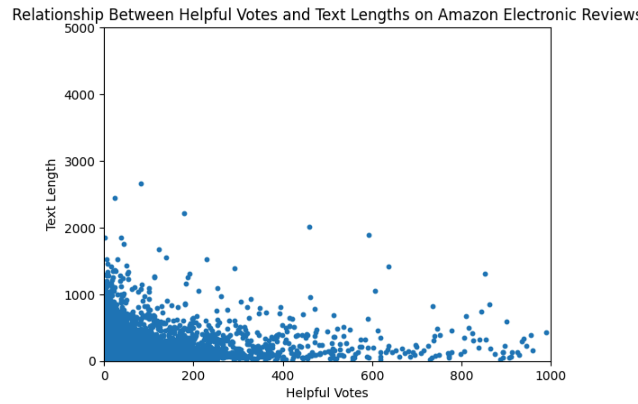

Figure 9

This relationship underscores the importance of concise, insightful reviews—striking a balance between length and substance is key for reviewers aiming to engage a wider audience. Interestingly, some shorter reviews still garnered a significant number of helpful votes, indicating that brevity combined with high relevance can be equally impactful. This observation aligns with behavioral trends in online platforms, where users prefer quickly digestible content unless deeply invested in the product or topic.

The analysis suggests that verified purchase reviews tend to receive more helpful votes compared to unverified ones, emphasizing the role of authenticity in influencing other shoppers. These insights are valuable for sellers and platforms alike, providing guidance on how review incentives and helpful vote mechanisms can be optimized to encourage meaningful user participation.

7. Challenges and Trade-offs

We encountered several challenges along the way. Handling corrupted timestamps was one of the first obstacles, and missing or incomplete text in some reviews limited our word-frequency analysis. Using a simple polarity-based sentiment model was a practical choice, but we recognized that it has limitations—it doesn't fully capture mixed or nuanced emotions.

Working with a large dataset also forced us to balance between sample size and computational limits. We ended up using a sample to ensure efficiency, but this introduced the risk of missing some insights that a full analysis might uncover. Analyzing a subset provided practical insights but may introduce sampling bias, limiting generalizability to the full dataset. Visualization choices also required trade-offs between clarity and complexity—simple plots efficiently captured trends but may have overlooked more intricate patterns within review content.

Moreover, our analysis methods faced limitations when dealing with the informal language prevalent in online reviews. Customers often use slang, abbreviations, misspellings, and emojis to express their sentiments, which our simple polarity-based sentiment model might not accurately interpret. This can lead to misclassification of sentiments, especially in cases of sarcasm or when emoticons carry significant emotional weight. Recognizing these nuances is crucial for a comprehensive understanding of customer feedback.

8. Potential Solutions and Next Steps

Moving forward, we plan to automate data validation and streamline preprocessing, ensuring cleaner datasets. Advanced sentiment models, such as fine-tuned BERT or transformers, could better capture mixed sentiments within reviews. Additionally, incorporating clustering techniques could unlock deeper insights, revealing hidden themes in reviews beyond word frequencies. Additionally, time-series analysis could identify trends in sentiment and review patterns over time, helping businesses anticipate future customer expectations. Expanding the dataset and refining analysis methods will ensure future iterations are more robust, insightful, and actionable.

To overcome these limitations, future work will consider incorporating advanced techniques such as emoji sentiment analysis and sarcasm detection to better handle informal language. By interpreting emojis and identifying sarcastic comments, we can enhance the accuracy of our sentiment analysis. Additionally, implementing aspect-based sentiment analysis will enable us to examine sentiments related to specific product features like battery life, durability, or user interface. This approach can provide more granular insights into customer preferences and pain points, aiding manufacturers and retailers in targeted product improvements.

9. Conclusion

This analysis provides valuable insights into customer sentiment, review patterns, and user behavior regarding electronics sold on Amazon. Key findings reveal the influence of review length on ratings and common themes across rating categories, highlighting patterns of both satisfaction and frustration. Sentiment analysis aligns well with customer feedback, though advanced methods could provide deeper insights.

The insights generated from this analysis have potential business implications, such as improving product features based on common customer complaints or enhancing marketing strategies aligned with positive feedback. By linking customer sentiment to business metrics like sales or customer retention, these findings could guide strategic decision-making for manufacturers and retailers.

This project also highlighted some challenges, like data inconsistencies and processing limitations, but it gave us valuable ideas for refining future analyses. Using more advanced models and tools will help uncover even deeper insights. Overall, this work lays a solid foundation for understanding customer feedback and using it to drive product improvements and business strategies.

References

Ahmad, M., et al. (2017). Hybrid Tools and Techniques for Sentiment Analysis: A Review. *International Journal of Multidisciplinary Sciences and Engineering, 8*(4). https://www.ijmse.org/Volume8/Issue4/paper5.pdf

Almahmood RJK, Tekerek A. (2022). Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E-Commerce Field. *Applied Sciences*.; 12(21):11256. https://doi.org/10.3390/app122111256

Alzahrani. M., et al. (2022, May 28) "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews." *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/3840071

Bunker, A. (2024, September 30). Net Promoter Score (NPS): The ultimate guide. *Qualtrics*. https://www.qualtrics.com/experience-management/customer/net-promoter-score/

Chinnalagu, A., & Durairaj, A. K. (2021). Context-based sentiment analysis on customer reviews using machine learning linear models. *PeerJ Computer Science,* 7, e813. https://doi.org/10.7717/peerj-cs.813

Daza, A., et al. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works. International Journal of Information Management Data Insights, 4(2), 100267. https://doi.org/10.1016/j.jjimei.2024.100267

Kumar, A., & Garg, G. (2020). Systematic literature review on context-based sentiment analysis in social multimedia. *Multimed Tools and Applications,* 79, 15349–15380. https://doi.org/10.1007/s11042-019-7346-5

Lee, Y., et al. (2023). Painsight: An Extendable Opinion Mining Framework for Detecting Pain Points Based on Online Customer Reviews. *arXiv*. arxiv.org/pdf/2306.02043.pdf

Malik, N., & Bilal, M. (2023). Natural Language Processing for Analyzing Online Customer Reviews: A Survey, Taxonomy, and Open Research Challenges. Preprints. https://doi.org/10.20944/preprints202312.2210.v1.

Morgan, B. (2022, November 8). Are Surveys Really Customer-Centric? *Forbes*. www.forbes.com/sites/blakemorgan/2022/10/20/are-surveys-really-customer-centric

Dey, S., et al. (2020, February)"A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews." *IEEE Xplore*.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9076924&tag=1.

Genc-Nayebi, N., & Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125, 207-219. https://doi.org/10.1016/j.jss.2016.11.027

Al-Moslmi, T., et al. (2017). Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. *IEEE Access*, 5, 16173-16192. https://doi.org/10.1109/ACCESS.2017.2690342

Suhaeni, C., & Yong, H.-S. (2023). Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences. *Applied Sciences, 13*(17). https://doi.org/10.3390/app13179766

The American Customer Satisfaction Index. (2022, April 12). *The Science of Customer Satisfaction*. https://theacsi.org/company/the-science-of-customer-satisfaction/

Wassan, S., et al. (2021) “Amazon Product Sentiment Analysis Using Machine Learning Techniques.” *International Journal of Early Childhood Special Education (INT-JECSE)*. www.revistaclinicapsicologica.com/archivesarticle.php?id=394.

Zhang, Lei, et al. (2018). Deep Learning for Sentiment Analysis: A Survey. *WIREs: Data Mining And Knowledge Discovery*, 8(4), e1253. https://doi-org.ezproxy.library.wisc.edu/10.1002/widm.1253.

Zhu, L., et al. (2022). Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science,* 8, e1044. https://doi.org/10.7717/peerj-cs.1044

Zucco, C., et al. (2020) Sentiment analysis for mining texts and social networks data: Methods and tools. *WIREs Data Mining And Knowledge Discovery*, 10, e1333. https://doi.org/10.1002/widm.1333