Sentiment Analysis of Amazon Electronics Product Reviews

Yi Zu & Nishtha Sawhney
Khoury College of Computer Sciences
DS 5110: Introduction to Data Management and Processing
Final Project Report

December 12, 2024

Contents

1 Introduction

Understanding customer sentiment is crucial in today's electronics market, where consumer feedback directly influences product success and market trends. This project conducts sentiment analysis of Amazon Electronics product reviews using Natural Language Processing (NLP) techniques, specifically VADER and TextBlob sentiment analyzers, to extract meaningful insights from customer feedback. The study analyzes 5,000 reviews from the Electronics category on Amazon until July 2014, aiming to uncover patterns in customer satisfaction and identify key factors influencing consumer opinions.

The project objectives focus on analyzing sentiment distribution using both VADER and TextBlob analysis tools, comparing their effectiveness in understanding customer feedback patterns, and creating interactive visualizations of sentiment trends. By extracting and analyzing frequent keywords, the project identifies common themes in customer feedback to provide actionable insights for product development and customer satisfaction improvement.

The methodology comprises several key components, beginning with comprehensive data processing of Amazon Electronics product reviews. This includes text standardization, handling missing values, and metadata formatting. The analysis phase employs both VADER and TextBlob for sentiment classification, supplemented by keyword extraction and frequency analysis. The study tracks sentiment patterns across brands and product categories to understand how different factors influence customer opinions.

The findings are presented through an interactive dashboard featuring various visualization techniques, including sentiment distribution charts, word clouds for positive and negative reviews, and brand-specific sentiment analyses. The deliverables encompass a processed dataset, visualization dashboard, analysis report, and technical documentation. This research enhances the field of consumer behavior analysis by providing data-driven insights into customer sentiment patterns in the electronics market, benefiting both manufacturers and consumers through an improved understanding of satisfaction drivers and product performance factors.

2 Literature Review

Sentiment analysis of product reviews has become increasingly important in understanding customer behavior and product performance. This review examines key developments in sentiment analysis techniques and their application to online product reviews, specifically in the electronics sector.

2.1 Dataset and Foundational Work

McAuley and Leskovec (2013) made a significant contribution to the field by establishing the largest public Amazon review dataset (1996-2014), which became an industry standard for e-commerce research. Their work was groundbreaking in its integration of ratings and text analysis, enabling researchers to analyze user sentiment from unstructured, informal text. This dataset and methodology provided a framework for extracting real-world insights beyond structured surveys, particularly valuable for understanding consumer behavior in the electronics sector.

2.2 Sentiment Analysis Methodologies

The development of sentiment analysis tools has evolved significantly since this foundational work. The VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer, developed by Hutto and Gilbert (2014), demonstrated particular effectiveness in analyzing informal text, achieving correlation scores above 0.9 with human raters. Their work showed that lexicon-based approaches could effectively handle the nuances of user-generated content, particularly relevant for analyzing product reviews.

2.3 Feature Extraction and Analysis

Hu and Liu (2004) established fundamental methods for extracting product features from reviews and determining sentiment orientation. Their approach to mining and summarizing customer reviews laid the groundwork for modern sentiment analysis techniques, providing methods that remain relevant for analyzing product-specific feedback and features.

2.4 Research Gaps

Significant gaps exist in current research:

- While McAuley and Leskovec established a comprehensive dataset, there is limited research combining multi-dimensional text analysis (quality, performance, reliability, value, emotional, usability) with sentiment analysis for electronic product reviews
- Although individual text analysis techniques exist, there is insufficient research on combining whitelist approaches, polarity-aware analysis, and context-sensitive sentiment analysis specifically for technical product reviews
- Despite the availability of various analysis tools, there is limited research integrating structured keyword analysis with sentiment patterns across product categories and brands in the electronics sector

2.5 Application to Current Project

- Implementing a multi-phase analysis approach that combines structured sentiment analysis with VADER and TextBlob, and comprehensive text analysis using whitelist categorization across six key dimensions: quality, performance, reliability, value, emotional response and usability
- Developing an innovative polarity-aware system that considers word context, negations, and modifiers while analyzing technical product reviews, enhanced by spaCy-based natural language processing
- Creating an integrated analysis framework that connects product-level keyword analysis with brand and category sentiment patterns through interactive visualizations and detailed metrics

3 Methodology

3.1 Data Collection

The data collection for this sentiment analysis project utilized the Amazon Product Reviews dataset made available by McAuley and Leskovec through UC San Diego's database. The dataset spans from 1990 to 2014 and was accessed via `https://jmcauley.ucsd.edu/data/amazon/index_2014.html`. The reviews were provided as a compressed JSON file (reviews_Electronics_5.json.gz) and loaded using Pandas' read_json function with the 'lines=True' parameter to properly parse the JSON Lines format.

The team focused on the "Small subsets for experimentation" section, collecting 5,000 product reviews from the Electronics category. Each review entry contained essential information including unique reviewer IDs, product ASINs (Amazon Standard Identification Numbers), review text, star ratings, review summaries, and timestamps. Additional product metadata was collected including category classifications, product descriptions, titles, brands, and pricing information.

3.2 Data Preprocessing

3.2.1 Raw Data Processing

The data preprocessing phase began with the extraction of 5,000 recent reviews from a larger Amazon Electronics dataset containing over 1.6 million reviews. We focused on reviews from July 2014, creating a manageable yet representative sample for our sentiment analysis. Initial data cleaning revealed and removed 32 duplicate reviews, bringing our dataset to 4,968 unique entries. Our text preprocessing pipeline implemented several critical cleaning steps to prepare the reviews for sentiment analysis. The process began with standardizing all text to lowercase and removing potential sources of noise, including HTML tags, URLs, special characters, and numbers. We then implemented a custom stopword removal system to eliminate common English words that don't contribute to sentiment meaning, while carefully preserving the semantic content of each review. To enhance our dataset's analytical capabilities, we engineered additional features and standardized temporal data. Review timestamps were converted from Unix format to readable datetime objects, and we calculated metrics such as review length and word count. These new features revealed that our reviews averaged 514 characters and 93.7 words in length, with significant variations ranging from very brief to extensive reviews of up to 2,734 words.

Final cleaning steps addressed data quality issues, including handling missing reviewer names by replacing them with 'Anonymous' and removing two empty reviews. The resulting dataset contained 4,966 reviews with a notable distribution across rating categories: 61.5% five-star reviews, 19.8% four-star reviews, and the remaining 18.7% distributed across one to three stars. This clean, structured dataset provides a solid foundation for subsequent sentiment analysis, with verified data integrity and consistent formatting throughout.

3.2.2 Metadata Integration

In our data preprocessing phase, a critical step was merging our cleaned reviews with product metadata to provide essential context for our analysis. While our initial review dataset contained only product IDs (ASINs) and review text, the metadata integration allowed us to incorporate valuable product information including titles, brands, categories, and pricing. This additional context was crucial for conducting meaningful analysis across different product types and market segments.

The merging process began with converting the metadata from JSON format to CSV for improved processing efficiency. Using the product ASIN as the joining key, we merged this metadata with our review dataset. We found matching product information for 3,128 reviews (63.0% of our dataset), while 1,838 reviews lacked corresponding metadata. Rather than removing reviews without metadata, we retained all 4,966 reviews and marked missing metadata fields as "Unknown." This decision preserved our full dataset while indicating where product information was unavailable. The merged dataset provided a complete view of customer sentiment across various product categories, allowing us to analyze both reviews with and without detailed product context.

3.2.3 Final Data Structure

Building upon our initial text preprocessing, we enhanced the dataset by incorporating and cleaning product metadata to provide a richer context for sentiment analysis. Starting with 4,966 reviews and 34 metadata columns, we first streamlined the dataset by removing 12 non-essential columns such as technical specifications and image URLs, retaining 22 columns that directly contributed to our analysis objectives.

The metadata cleaning process focused on standardizing product information across multiple dimensions. We converted price data from string format to numerical values, revealing that 1,712 products had valid prices ranging from $0.47 to $1,628.88, with a mean price of $36.39. For missing price data, we used NaN values to ensure accurate statistical analysis. We also standardized category-related information by replacing generic "Unknown" values with more descriptive placeholders - using "Uncategorized" for missing categories, "Unknown Brand" for missing brand information, and "No description available" for missing product descriptions.

To enhance the analytical value of user feedback data, we created a new "helpful_ratio" feature calculated from the helpful votes data, providing a normalized measure of review helpfulness. We also improved the overall data structure by renaming columns for consistency (such as changing "reviewerID" to "reviewer_id") and organizing columns in logical groups. The final dataset captured 19 unique main categories in the electronics product space, including major categories like Camera & Photo, All Electronics, and Computers, providing a comprehensive foundation for analyzing sentiment patterns across different product types and price points.

3.3 Analysis Techniques

3.3.1 Sentiment Analysis

For our sentiment analysis, we first implemented the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer, a lexicon and rule-based tool specifically designed for social media text. We chose VADER because it performs well with informal text, handles negations, punctuation, and emojis effectively, and is especially adept at understanding the intensity of sentiments, all characteristics that are common in customer reviews. The VADER library assigns a compound score between -1 (most extreme negative) and +1 (most extreme positive) to each text, which we then classified into three categories using standard thresholds: positive for scores ≥ 0.05, negative for scores ≤ -0.05, and neutral for scores in between. These thresholds provide a small buffer around zero to ensure clear sentiment classification.

Our combined sentiment analysis process addresses these challenges by integrating TextBlob with VADER in a systematic approach that better handles the nuances of product reviews. Our method analyzes each review in several steps. First, we run each review through both VADER and TextBlob to get two independent assessments. VADER gives us a score indicating how positive or negative the text is, while TextBlob provides both sentiment direction and how confident it is in its assessment. We then combine these tools' opinions, giving more weight to whichever tool shows more confidence in its assessment.

A key innovation in our approach is how we detect neutral reviews, a critical challenge when analyzing product reviews. Our system uses several checks to identify neutral content: it looks at whether VADER and TextBlob agree on their assessment, how strong their sentiment signals are, and whether the text seems more factual than opinionated. This multi-check approach helps catch neutral technical descriptions that might otherwise be misclassified as mildly positive or negative.

Our neutral detection also works well with common patterns in product reviews, such as feature listings, technical specifications, or basic functionality statements. By considering multiple indicators and weighing their importance appropriately, our system can better distinguish between actual opinions and neutral product information.

3.3.2 Data Visualization

For the visualization component of our analysis, we built an interactive web dashboard that makes our sentiment analysis results easy to explore and understand. Using Flask as our web framework and Plotly for creating charts, we designed a system where users can view sentiment patterns from different angles, whether looking at overall trends or drilling down into specific product categories and brands. The dashboard updates in real-time as users select different categories, showing how sentiments vary across different parts of the electronics market. To ensure the dashboard runs smoothly with our large dataset, we added data caching features. We kept the design clean and consistent, using the same color scheme throughout (blues for positive, purples for neutral, and teals for negative) to make the visualizations intuitive to read. This approach helps turn our complex sentiment data into clear, actionable insights about customer opinions in electronics reviews.

Our visualization strategy focused on two main types of analysis: overall sentiment patterns and categorical breakdowns. For the overall analysis, we created two key visualizations: a sentiment distribution pie chart and a rating-versus-sentiment bar chart. The pie chart gives a quick snapshot of how sentiments are distributed across all reviews, while the rating-versus-sentiment chart reveals interesting patterns in how customers' numeric ratings (1-5 stars) align with the sentiments expressed in their written reviews.

This comparison is particularly valuable as it shows cases where star ratings might not tell the whole story, for instance, some 5-star reviews might contain neutral language when describing product features.

We also analyzed the distribution of reviews across product categories through a dedicated visualization showing the top 10 categories by review volume. This helps understand which product types generate the most customer feedback and how sentiment varies across different product types. For example, categories like Computers and Cell Phone Accessories tend to have high review volumes, suggesting these products particularly engage customers to share their experiences. Each category's sentiment breakdown helps identify which product types tend to generate more positive or negative feedback. The brand analysis visualization offers crucial insights into manufacturer performance through a stacked bar chart showing sentiment distribution for the top brands. This helps identify which brands consistently receive positive feedback and which might have room for improvement. This information is valuable not just for manufacturers but also for retailers making inventory decisions and customers choosing between brands. The visualization shows both the volume of reviews each brand receives and the proportion of positive, neutral, and negative sentiments, providing a comprehensive view of brand performance in the market.

For our categorical analysis, we created specialized visualizations that examine sentiment from three critical business perspectives - brands, product successes, and potential issues. This categorical breakdown is essential because overall sentiment numbers alone can mask important patterns in customer feedback. For example, while our overall analysis might show generally positive sentiment across all electronics reviews, breaking it down by category reveals which specific brands, products, or categories drive those positive results or face challenges. Brand analysis is particularly crucial in the electronics market, where brand reputation significantly influences purchasing decisions. Our stacked bar visualization of brand sentiment helps identify which manufacturers consistently meet customer expectations and which ones show room for improvement. This information is valuable across the industry: retailers can optimize their inventory mix, customers can make more informed choices, and manufacturers can understand their market position relative to competitors.

For product-level insights, we developed an innovative approach to identifying the top 5 positive and negative products in each category. Rather than simply counting the number of positive or negative reviews, we calculated a sentiment ratio that considers both review volume and sentiment distribution. For positive products, we looked at the ratio of positive reviews to neutral and negative ones combined. This helps identify products that are genuinely well-received rather than just frequently reviewed. Similarly, for negative products, we examined the ratio of negative reviews to positive and neutral ones, helping spot products with consistent issues rather than just occasional complaints. This ratio-based approach provides a more nuanced understanding of product performance than simple review counts.

Breaking down sentiment by product categories helps understand how customer expectations and satisfaction vary across different types of electronics. This insight is valuable because different product categories often serve different customer needs and face different challenges. For instance, complex products like computers might naturally generate more mixed reviews due to their complexity, while simpler accessories might show more polarized sentiment patterns. Understanding these category-specific patterns helps businesses set appropriate expectations and develop targeted strategies for different market segments.

3.3.3 Text Analysis

The project began with a focused approach to understanding customer sentiment through word cloud analysis. This initial technique employed a whitelist methodology to systematically categorize meaningful terms across six key dimensions: quality, performance, reliability, value, emotional response, and usability. The function created specialized word clouds that visually represented these filtered terms, with the capability to focus on specific sentiment categories. This allowed for a clear visualization of how customers expressed their opinions across different aspects of product reviews.

Building on this foundation, the analysis incorporated a more sophisticated polarity-aware word cloud system. This second phase enhanced the sentiment analysis by considering the contextual meaning of words. The system was designed to recognize positive and negative word sets while accounting for negation words like "not" or "never" and context modifiers such as "but" or "however". This nuanced approach enabled a more accurate representation of customer sentiment by considering how words' meanings change based on their context.

The third phase introduced a comprehensive Sentiment Term Analyzer class that provided a structured approach to analyzing review content. This component performed detailed term analysis across various categories, calculating specific metrics such as positive and negative counts, percentages, and ratios. The analyzer generated multiple visualization types, including bar charts for category comparisons and distribution plots, offering different perspectives on the sentiment patterns in the reviews.

The project then implemented a focused keyword product analysis system. This component examined how specific keywords appeared across different products, calculating detailed metrics for each item and grouping results by categories and brands. This analysis provided insights into which products and brands were most frequently associated with particular keywords, offering a product-level view of customer sentiment.

The fifth phase employed advanced keyword analysis techniques using natural language processing. This comprehensive approach utilized the spaCy library for sophisticated text analysis, including dependency parsing and key phrase extraction. The system analyzed keyword contexts, calculated their impact on ratings, and identified significant word pairs through collocation analysis, providing a deep understanding of how words were used in reviews.

Finally, the project created a series of visualizations to represent the findings effectively. These visualizations included word pair associations, keyword impact on ratings, feature category distributions, and sentiment distribution charts. This visual representation of the data helped identify patterns and trends in customer sentiment across different products and categories.

4 Results

4.1 Sentiment Analysis Results

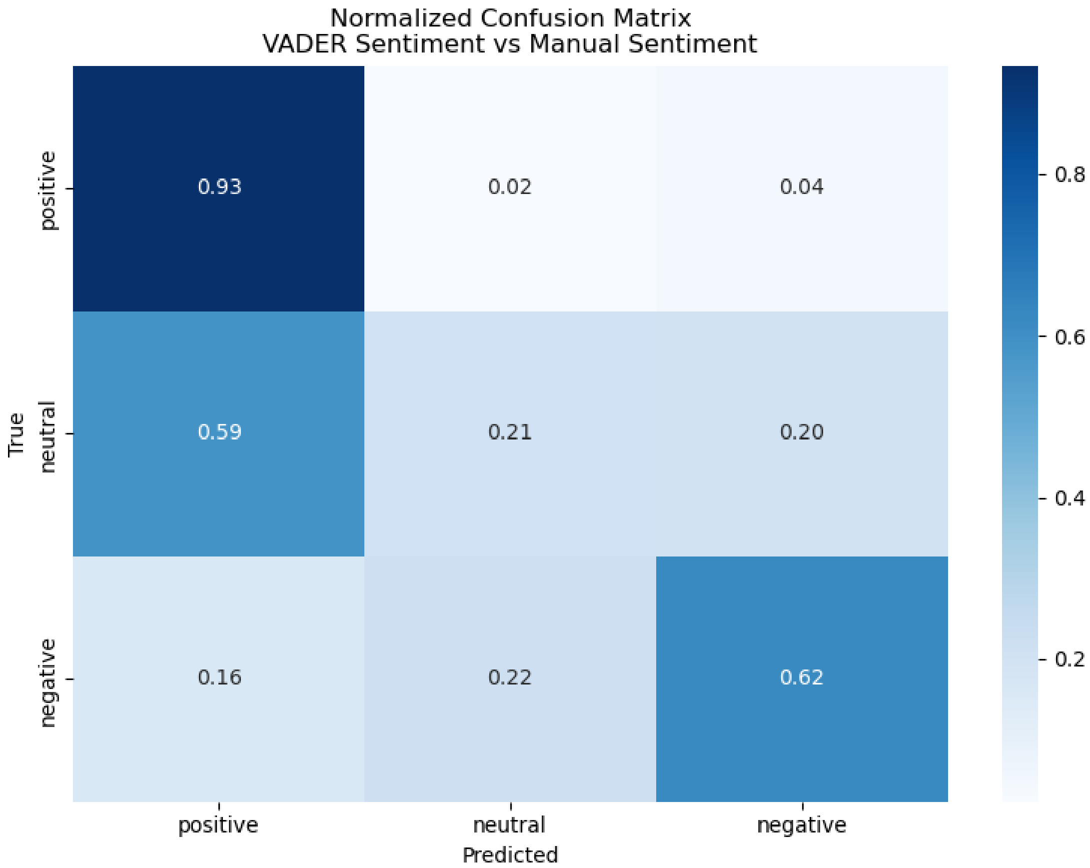

Figure 1: VADER Confusion Matrix Showing Classification Performance Across Sentiment Categories

```
Accuracy: 81.88%

Classification Report:
              precision    recall  f1-score

    positive       0.33      0.62      0.43
     neutral       0.49      0.21      0.29
    negative       0.89      0.93      0.91

    accuracy                           0.82
   macro avg       0.57      0.59      0.55
weighted avg       0.81      0.82      0.80


Sentiment Distribution:
positive: 847 reviews (84.8%)
negative: 93 reviews (9.3%)
neutral: 59 reviews (5.9%)
```

Figure 2: Initial VADER Classification Metrics and Sentiment Distribution Report

Our sentiment analysis results demonstrate a clear improvement from the initial VADER implementation to our combined VADER-TextBlob approach, as evidenced by the confusion matrices and classification reports. The initial VADER analysis of our review dataset showed moderately good overall accuracy at 81.88%, but revealed significant classification biases. In particular, while VADER accurately identified truly positive reviews (0.93 true positive rate), it struggled with neutral content, misclassifying 59.0% of neutral reviews as positive and showing a low neutral recall of 0.21. The sentiment distribution was notably skewed, with 84.8% positive, 9.3% negative, and only 5.9% neutral reviews.

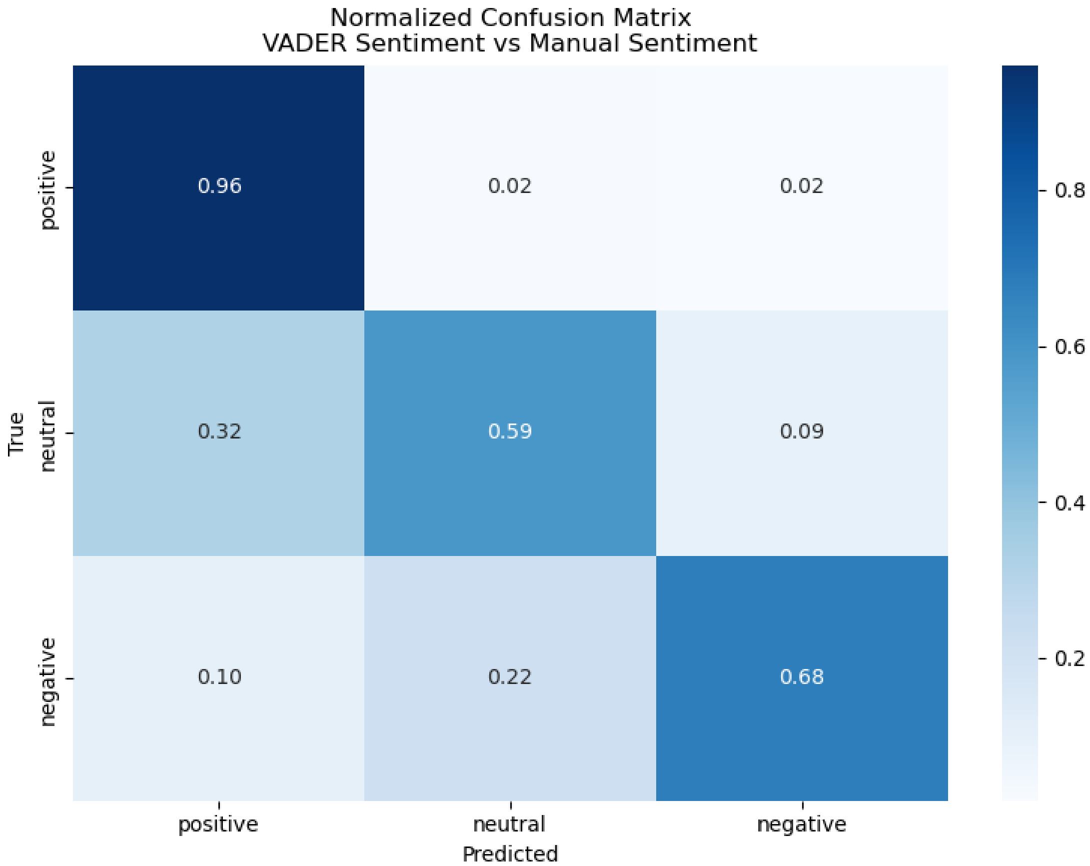

Figure 3: Enhanced VADER-TextBlob Combined Approach Confusion Matrix Demonstrating Improved Classification

```
Accuracy: 89.49%

Classification Report:
              precision    recall  f1-score

    positive       0.56      0.68      0.61
     neutral       0.74      0.59      0.65
    negative       0.94      0.96      0.95

    accuracy                           0.89
   macro avg       0.74      0.74      0.74
weighted avg       0.89      0.89      0.89


Sentiment Distribution:
positive: 828 reviews (82.9%)
neutral: 110 reviews (11.0%)
negative: 61 reviews (6.1%)
```

Figure 4: Enhanced VADER-TextBlob Classification Metrics and Updated Sentiment Distribution Report

Our combined VADER-TextBlob approach significantly improved these metrics, achieving an overall accuracy of 89.49%. The enhanced confusion matrix shows more balanced classification performance, particularly in neutral content detection. The true positive rate improved to 0.96, while the neutral classification accuracy increased substantially from 0.21 to 0.59. The combined approach also reduced misclassification rates, with only 32.0% of neutral reviews being incorrectly labeled as positive, compared to the previous 59.0%.

The classification metrics show comprehensive improvements across all categories. Positive review precision increased from 0.33 to 0.56, neutral precision from 0.49 to 0.74, and negative precision remained strong at 0.94. The F1-scores, which measure balanced performance between precision and recall, improved notably: positive reviews from 0.43 to 0.61, neutral from 0.29 to 0.65, and negative from 0.91 to 0.95. The final sentiment distribution also shows a more balanced representation: 82.9% positive, 11.0% neutral, and 6.1% negative reviews, indicating better alignment with human-annotated sentiments and more nuanced classification of neutral technical content.

These improvements are particularly significant given the technical nature of electronics reviews, where neutral statements about specifications and features are common. The enhanced performance in neutral classification suggests our combined approach better handles the nuanced language typical in product reviews, while maintaining strong performance in identifying clearly positive and negative sentiments.

4.2 Dashboard Analysis

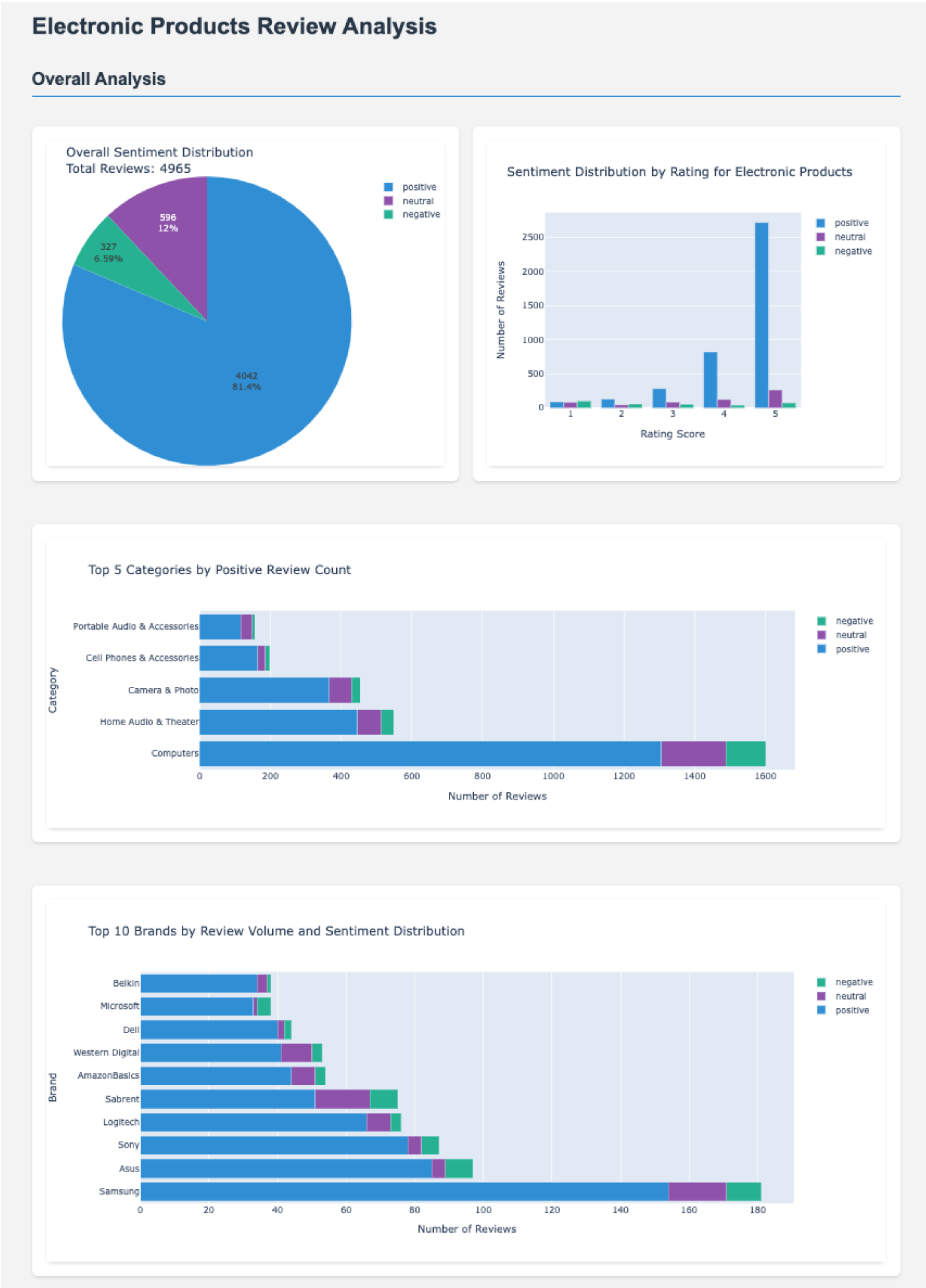

Figure 5: Dashboard displaying the overall sentiment analysis, top 5 categories by review count and top 10 brands by review volume and sentiment distribution.

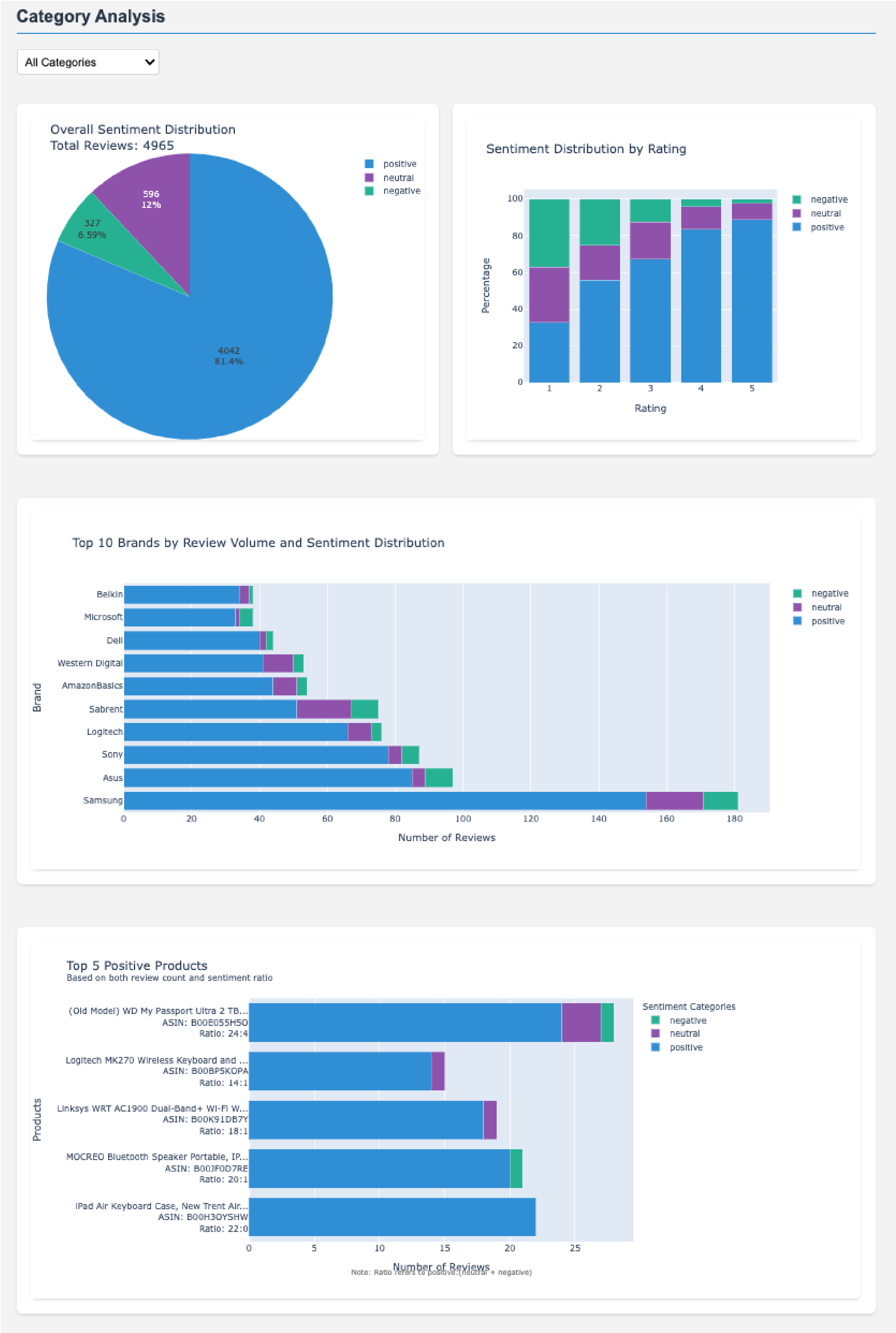

Figure 6: Dashboard with product category sentiment analysis

The most notable finding is the overwhelming dominance of positive sentiment, with approximately 83‰of reviews expressing favorable opinions. Neutral reviews account for about 11‰of the total, while negative reviews make up only 6‰. This heavily skewed distribution suggests either broad customer satisfaction with electronics products or a potential bias in review submission behavior.

The relationship between rating scores and sentiment shows a strong correlation in the data. Five-star ratings appear with the highest frequency, aligning with the predominant positive sentiment found in the text analysis. There is a marked decrease in review frequency for lower ratings (1-3 stars), which could indicate a reporting bias where satisfied customers are more likely to submit reviews than dissatisfied ones. This pattern raises interesting questions about the representativeness of online review data in capturing the full spectrum of customer experiences.

When examining category-specific patterns, the **Computers** category emerges as the most frequently reviewed category with a significant lead in review volume. **Home Audio & Theater** follows as the second most reviewed category, while **Camera & Photo** maintains moderate review numbers. **Cell Phones & Accessories** shows consistent review patterns, and **Portable Audio & Accessories** receives comparatively fewer reviews. This distribution might reflect either market size differences or varying customer engagement levels across product categories.

Brand performance analysis reveals consistently positive patterns across major manufacturers, with **Samsung** demonstrating particularly strong performance in positive sentiment. **Amazon's** branded products maintain high positive sentiment ratios, suggesting successful product development and customer satisfaction. Most major brands exhibit similar patterns characterized by high proportions of positive reviews (shown in blue segments), small proportions of neutral reviews (purple segments), and minimal negative reviews (green segments). This consistency across brands could indicate either strong quality control in the electronics industry or systematic patterns in how customers engage with review systems.

4.3 Text Analysis

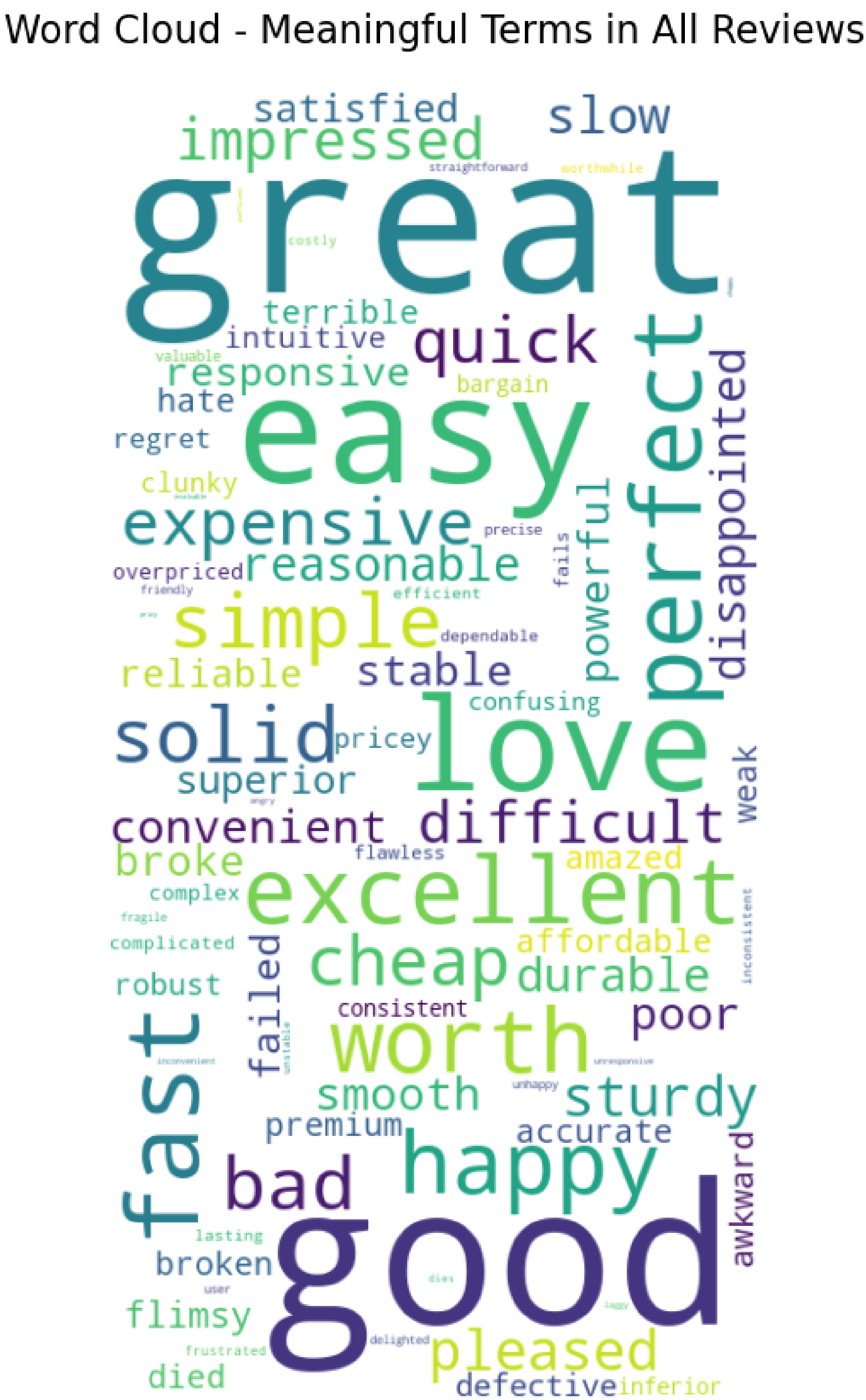

Figure 7: Word cloud for all terms

Table 1: Frequency analysis of key quality and performance terms in Amazon electronics reviews, showing quality-related words appearing significantly more often than performance descriptors

Category	Term	Frequency
Quality	great	1,389
Quality	good	1,225
Quality	perfect	237
Quality	excellent	214
Performance	fast	257
Performance	stable	42
Performance	broke	39
Performance	failed	35

The sentiment analysis reveals several noteworthy patterns beyond the raw frequency counts. First, there's a clear hierarchical structure in how customers evaluate electronics products, with general quality assessments ("great", "good") vastly outnumbering specific performance descriptors. This suggests customers tend to form overall impressions rather than focus on technical details.

The significant gap between positive and negative sentiment terms (for example, 1,389 mentions of "great" versus only 39 mentions of "broke") indicates a strong positive skew in reviews. However, this should be interpreted cautiously since negative experiences may be expressed through more varied vocabulary or detailed descriptions rather than single negative terms.

Performance-related feedback shows an interesting pattern: while "fast" appears 257 times, other performance terms like "stable" (42 mentions) and "failed" (35 mentions) are much less frequent. This suggests that speed is the primary performance metric customers choose to comment on, while other performance aspects may be mentioned only when they fall short of expectations.

The relative scarcity of technical performance terms compared to general quality assessments (214 mentions of "excellent" versus 42 mentions of "stable") might indicate a gap between how manufacturers market their products (often focusing on technical specifications) and how customers evaluate them (focusing on overall experience and basic functionality).

Finally, the frequency distribution of terms suggests a "satisfaction threshold" pattern - customers are more likely to use superlative terms ("great", "excellent") when extremely satisfied, but tend to use more moderate language ("good") or specific criticism when less satisfied. This indicates that achieving merely adequate performance may not be enough to generate strongly positive reviews in the competitive electronics market.

Figure 8: Word cloud visualization of sentiment terms in negative Amazon electronics reviews, where font size represents frequency and "NEG_" prefix indicates negated positive words (e.g., "NEG_great" represents phrases like "not great").

Category	Term	Frequency
Negative	bad	34
Negative	poor	17
Negative	difficult	15
Negative	terrible	14
Negative	slow	12
Category	Term	Frequency
Negated Positive	NEG_good	30
Negated Positive	NEG_great	16
Negated Positive	NEG_easy	9
Negated Positive	NEG_worth	8
Negated Positive	NEG_fast	5

The polarity-aware analysis highlights how customers express dissatisfaction and nuanced opinions in their reviews by capturing patterns such as negation and context modifiers. Negation words like "not," "never," and "no" are crucial in reversing the sentiment of positive terms, as seen in phrases like "not good" or "not worth it." This granular approach ensures that subtle shifts in sentiment are accurately identified, offering deeper insights into customer feedback. The prevalence of negated positive terms, such as "NEG_good" and "NEG_great," underscores how customers often communicate dissatisfaction less directly.

Context modifiers, such as "but," "however," and "though," further reveal how customers balance contrasting sentiments within their reviews. For instance, a statement like "The product is great, but it's not reliable" conveys both positive and negative sentiments, which traditional sentiment analysis might oversimplify. By accounting for these contrasts, the analysis captures the complexity of customer opinions and provides a more nuanced understanding of mixed reviews. These patterns demonstrate that customer feedback often includes both praise and criticism, reflecting a layered sentiment landscape.

Visualizations such as word clouds and frequency tables illustrate the prominence of terms like "difficult," "broke," and "NEG_good," providing a clear picture of how dissatisfaction is expressed. The findings reveal that customers frequently use negated positive terms and qualifiers to express nuanced dissatisfaction rather than outright negativity. This analysis offers actionable insights for businesses, allowing them to address specific concerns and improve their products and customer experiences. By understanding the underlying reasons for mixed or negative feedback, companies can better align their strategies to meet customer needs.

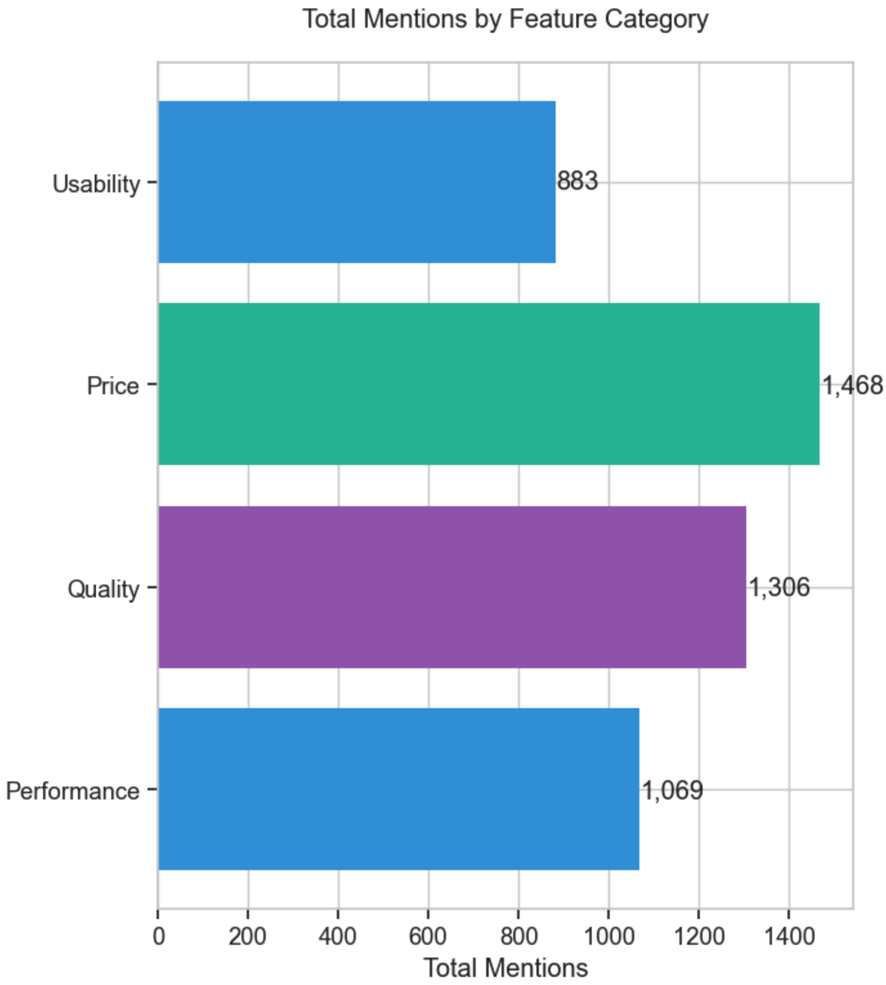

Figure 9: Distribution of feature-related terms in Amazon electronics reviews, showing price and quality as the most frequently discussed aspects, followed by performance and usability

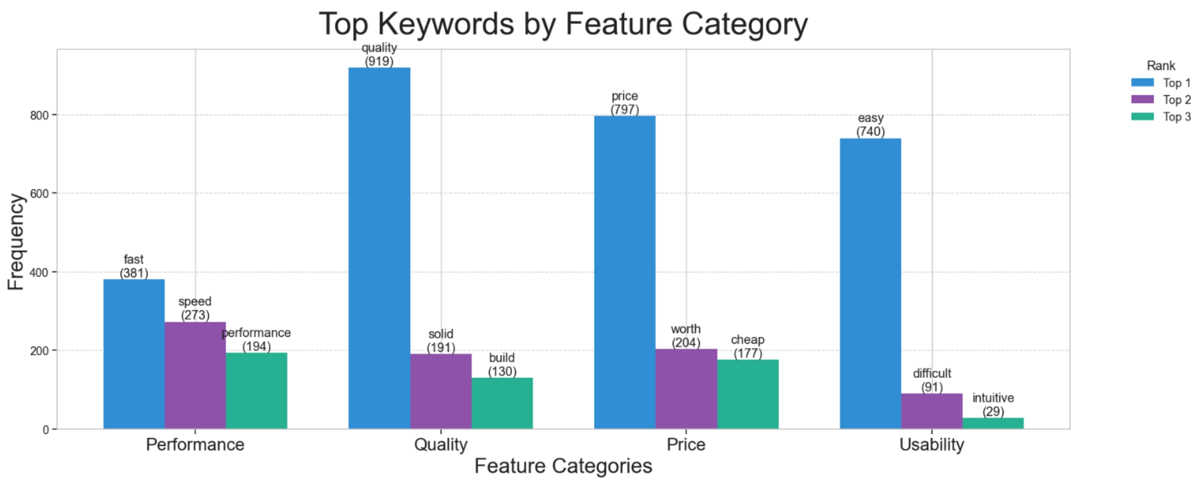

Figure 10: Frequency comparison of the top three keywords across feature categories (price, quality, usability) in Amazon electronics reviews, where blue bars indicate the most common terms, followed by purple and green for the second and third most frequent terms respectively.

Term analysis across feature categories revealed interesting distribution patterns. Price emerged as the most discussed feature category (1,468 mentions), followed by quality (1,306 mentions), while usability had the fewest mentions (883). Quality-related terms appeared 1,306 times, with “quality” being the most frequent specific term (919 mentions). Performance terms totaled 1,069 mentions, with “fast” leading at 381 occurrences. This suggests customers are particularly attentive to price-value relationships in electronics products while being more likely to express strong opinions when dissatisfied. Usability terms appeared 883 times, with “easy” being particularly common (740 mentions) suggesting ease of use as a factor for positive sentiment in product reviews.

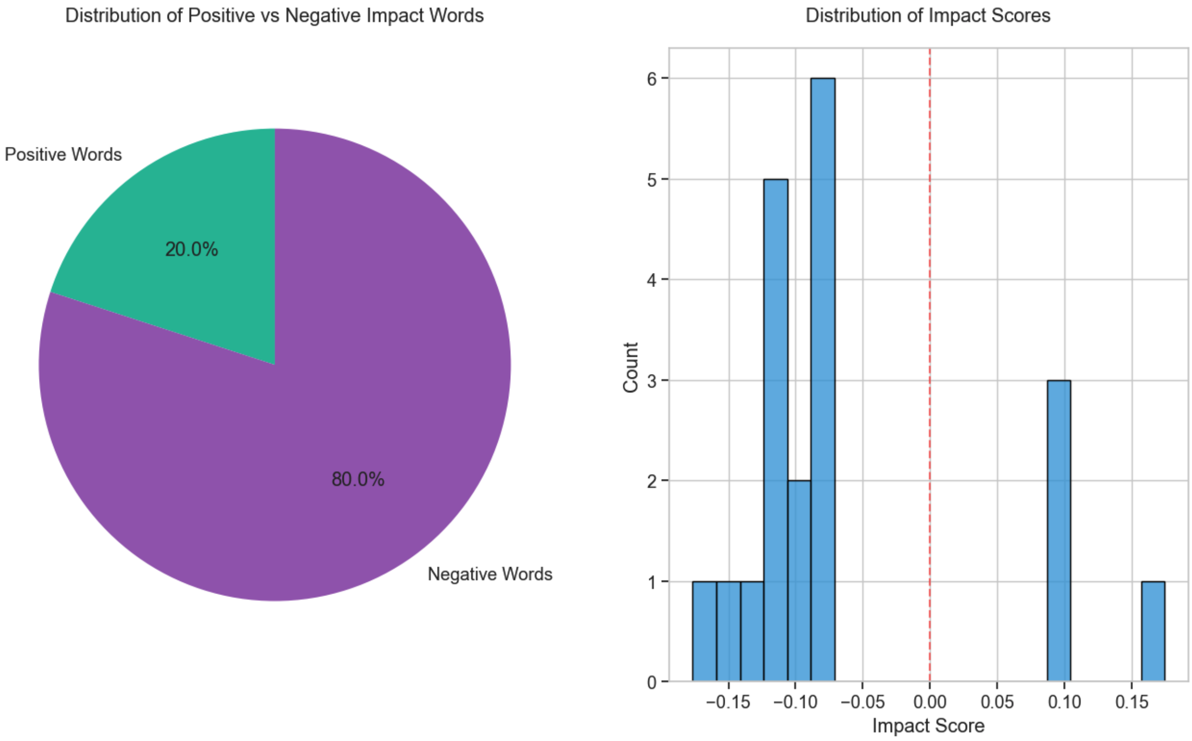

Figure 11: The pie chart and histogram reveal a stark contrast in review sentiment, with 80% of impactful words being negative despite moderate impact scores that cluster around -0.10, while the positive words (20%) show a more dispersed distribution of impact scores.

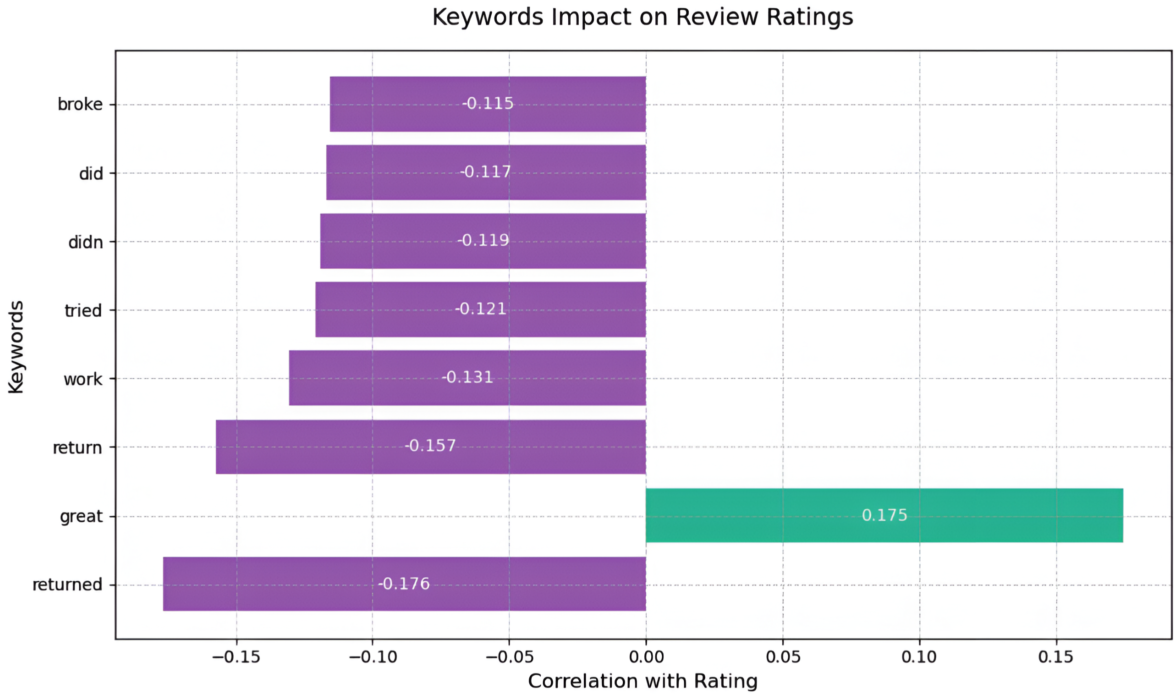

Figure 12: The horizontal bar chart shows keyword impact on review ratings, with "returned" having the strongest negative correlation (−0.176) and "great" having the strongest positive correlation (0.175), while other negative terms like "work," "tried," and "broke" show moderate negative correlations between −0.115 and −0.131.

The sentiment analysis visualizations reveal a compelling pattern in how customers express their experiences through reviews. While positive reviews may be more common overall, the pie chart shows that 80% of impactful words used are negative, with only 20% being positive. This asymmetry is further illuminated by the histogram, which shows a distinct clustering of impact scores around moderate negative values (−0.05 to −0.15), while positive impact scores, though fewer, are distributed across a wider range. This suggests that customers tend to be more specific and emotionally charged in their language when describing negative experiences.

This pattern is reinforced by the horizontal bar chart of keyword impacts, where action-oriented negative words dominate the strongest correlations with review ratings. The term "returned" shows the strongest negative correlation (−0.176), followed by other problem-indicating words like "work" (−0.131), "tried" (−0.121), and "broke" (−0.115). Only one positive term, "great" (0.175), appears among the most impactful words, matching the negative terms in strength but standing alone among them. This comprehensive view suggests that while customers may generally be satisfied with their purchases, they become significantly more articulate and specific in their word choice when expressing dissatisfaction, making negative reviews particularly valuable sources of actionable feedback despite their lower overall frequency.

5 Discussion

5.1 Methodological Advances and Challenges

Our combined VADER-TextBlob approach significantly improved sentiment analysis accuracy from 81.88% to 89.49%, with particular success in reducing neutral content misclassification from 59% to 32%. While this advances Hutto and Gilbert's (2014) work, our analysis revealed ongoing challenges with technical language in electronics reviews, where technical specifications and feature descriptions complicated sentiment interpretation, as evidenced by context-dependent terms like "stable" (42 mentions) and "fast" (257 mentions).

5.2 Customer Communication Patterns

Analysis of review text revealed a distinct hierarchy in how customers evaluate electronics products, with general quality assessments dominating over technical descriptions. This was demonstrated by the high frequency of broad evaluative terms like "great" (1,389 mentions) compared to technical descriptors. Notably, despite an overall positive review bias (83% positive reviews), 80% of impactful words were negative, suggesting customers are more specific and varied in expressing dissatisfaction than satisfaction.

5.3 Feature and Brand Analysis

Our categorical analysis revealed price as the most discussed feature (1,468 mentions), followed closely by quality (1,306 mentions), while usability garnered fewer but predominantly positive mentions (883). Brand analysis showed consistent positive sentiment across major manufacturers, with Samsung demonstrating particularly strong performance, indicating both successful brand positioning and industry-wide quality standards in the electronics sector.

5.4 Implications for Industry

The findings provide clear direction for industry stakeholders: the dominance of price and quality discussions in reviews suggests these should be primary focuses in product development, while the high frequency of usability terms indicates the importance of user-friendly design. Additionally, the detailed nature of negative reviews, despite their lower frequency, provides particularly valuable feedback for product improvement, suggesting companies should pay special attention to these more specific criticisms.

5.5 Research Contribution

This study advances sentiment analysis methodology by demonstrating the effectiveness of combining analysis tools for technical reviews, revealing the complex intersection between technical language and sentiment expression, and establishing a comprehensive framework for analyzing sentiment across six key dimensions: quality, performance, reliability, value, emotional response, and usability. This multi-dimensional approach provides a more nuanced understanding of customer feedback in technical product reviews than previously available.

6 Conclusion

This comprehensive study of Amazon Electronics reviews demonstrates the effectiveness of combining multiple sentiment analysis approaches to understand customer feedback. Our multi-layered methodology, integrating VADER and TextBlob, achieved a significant improvement in sentiment classification accuracy from 81.88% to 89.49%. The analysis revealed several key insights: a notable discrepancy between numeric ratings and written sentiment, the dominance of price considerations in customer reviews, and the asymmetric pattern in sentiment expression where negative experiences elicit more specific and varied vocabulary despite being less frequent.

The research uncovered complex patterns in how customers evaluate electronics products. While positive sentiments dominated the overall review landscape (83% positive reviews), the most impactful words were predominantly negative (80%), suggesting that dissatisfied customers provide more detailed and specific feedback. Our feature analysis revealed that customers prioritize price (1,468 mentions) and quality (1,306 mentions) over other aspects, while brand analysis demonstrated consistent positive sentiment across major manufacturers, with Samsung showing particularly strong performance.

Several limitations of this study warrant consideration. Our dataset from 2014 may not reflect contemporary review patterns, especially given the rapid evolution of e-commerce platforms and review systems. The strong skew toward positive reviews (83%) potentially limits our understanding of negative sentiment patterns, while our sentiment analysis tools still face challenges in handling technical language and subtle expressions of dissatisfaction. The mismatch between star ratings and written sentiment, particularly evident in 2-star reviews containing positive language, suggests that traditional sentiment analysis methods may not fully capture the nuanced way customers evaluate technical products.

Future research directions should address these limitations and extend our findings. Developing specialized sentiment analysis tools for technical product reviews, particularly in distinguishing between technical descriptions and sentiment expressions, would be valuable. Investigation into temporal changes in review patterns could provide insights into evolving customer behavior while exploring the relationship between technical complexity and sentiment expression could improve our understanding of customer feedback in the electronics market. Additionally, implementing advanced machine learning models like BERT could enhance sentiment analysis accuracy by better understanding context and technical terminology. Such a model could be trained to recognize domain-specific language patterns, such as distinguishing between "runs hot" as a complaint versus "high power" as a positive feature, thereby providing more nuanced insights into customer sentiment in technical product reviews.

The insights gained from this study have significant implications for both academic research and industry applications. For researchers, our methodology demonstrates the value of combining multiple sentiment analysis approaches and considering the technical context in sentiment classification. For industry practitioners, our findings provide actionable insights for product development, marketing strategies, and customer service improvements in the electronics sector. As e-commerce continues to evolve and customer review systems become increasingly sophisticated, the need for advanced sentiment analysis tools that can handle technical product reviews will only grow more crucial.

7 References

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. *Proceedings of the 7th ACM Conference on Recommender Systems*, 165-172.

Hutto, C.J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI.

Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, Seattle, Washington.