Edusaintek: Jurnal Pendidikan, Sains dan Teknologi
Volume 12 Issue 3 2025 Pages 1320 - 1332
p-ISSN: 1858-005X e-ISSN: 2655-3392 DOI: https://doi.org/10.47668/edusaintek.v12i3.1796
website: https://journalstkippgrisitubondo.ac.id/index.php/EDUSAINTEK

ENHANCING ASPECT-BASED SENTIMENT ANALYSIS FOR CONSUMER REVIEWS: CROSS-BRAND INSIGHTS FROM APPLE, SAMSUNG, AND XIAOMI

Abdullah Yahya Moqbel Al-Sabahi[1*], Dhomas Hatta Fudholi[2]
[1,2] Universitas Islam Indonesia, Indonesia
*Corresponding author: abdullahalsbahi2002@gmail.com

Abstract: In the competitive world of smart technology, understanding consumer sentiment plays a vital role in brand strategy and product innovation. This research applies Aspect-Based Sentiment Analysis (ABSA) using BERT, RoBERTa, and DistilBERT models on a dataset of approximately 20,000 reviews collected from Amazon, Flipkart, Walmart, and Best Buy between 2021 and 2024. The study identifies the most frequently mentioned aspects (Battery, Camera, Performance) and compares sentiment across devices (phones vs. watches) and brands (Samsung, Xiaomi, Apple). The findings reveal that Samsung is the most discussed brand, with polarizing sentiment on its smartphones but favorable opinions on its wearables. Xiaomi watches draw both high praise and criticism, while Apple maintains balanced sentiment with lower review volume. Temporal sentiment analysis indicates an upward trend in positive feedback in 2024, reflecting product improvements. These insights offer practical implications for OEMs aiming to align product features with evolving customer expectations.

Keywords: Sentiment analysis, Aspect-based sentiment analysis, BERT, RoBERTa, Smart devices, Customer reviews.

Abstrak: Dalam dunia teknologi pintar yang kompetitif, memahami sentimen konsumen memainkan peran penting dalam strategi merek dan inovasi produk. Penelitian ini menerapkan Analisis Sentimen Berbasis Aspek (ABSA) menggunakan model BERT, RoBERTa, dan DistilBERT pada kumpulan data sekitar 20.000 ulasan yang dikumpulkan dari Amazon, Flipkart, Walmart, dan Best Buy antara tahun 2021 hingga 2024. Studi ini mengidentifikasi aspek yang paling sering disebutkan (Baterai, Kamera, Performa) dan membandingkan sentimen berdasarkan perangkat (ponsel vs jam tangan) dan merek (Samsung, Xiaomi, Apple). Temuan menunjukkan bahwa Samsung adalah merek yang paling banyak dibicarakan, dengan sentimen yang terpolarisasi pada ponselnya tetapi opini positif pada perangkat wearable-nya. Jam tangan Xiaomi mendapatkan pujian sekaligus kritik tinggi, sementara Apple mempertahankan sentimen yang seimbang dengan volume ulasan yang lebih rendah. Analisis sentimen temporal menunjukkan tren peningkatan umpan balik positif pada tahun 2024, mencerminkan peningkatan produk. Temuan ini memberikan implikasi praktis bagi OEM yang ingin menyelaraskan fitur produk dengan harapan pelanggan yang terus berkembang.

Kata Kunci: Analisis sentimen, Analisis sentimen berbasis aspek, BERT, RoBERTa, Perangkat pintar, Ulasan konsumen.

INTRODUCTION

The smart technology industry, encompassing smartphones and wearable deviceshas rapidly evolved, driven in part by increasing volumes of online consumer feedback. As modern users actively voice their satisfaction or dissatisfaction through reviews, manufacturers face growing pressure to monitor, interpret, and act upon these insights (Devlin et al., 2019).

Conventional sentiment analysis techniques typically provide an overall opinion polaritypositive, neutral, or negative but fail to offer detailed insight into specific product features (Zhang et al., 2021). This limitation has led to the emergence of Aspect-Based Sentiment Analysis (ABSA), a refined method that not only classifies sentiment but also links it to specific aspects such as battery, camera, or display.

This study examines brand-wise customer sentiment for three leading manufacturers, Samsung, Xiaomi, and Apple, across two device types: smartphones and smartwatches. The novelty lies in the integration of temporal sentiment trends and brand-aspect comparisons using state-of-the-art transformer models. In doing so, the research supports product development teams, marketers, and customer service divisions with data-driven recommendations to enhance user satisfaction.

The novelty of this study lies in its modular approach to Aspect-Based Sentiment Analysis (ABSA), clearly separating Aspect Term Extraction (ATE) and Sentiment Classification (SC) tasks. Additionally, it provides a comparative analysis of leading transformer models (BERT, RoBERTa, DistilBERT) specifically applied to consumer reviews of major electronics brands. Moreover, this research uniquely integrates a temporal dimension, tracking evolving consumer sentiment from 2021 to 2024, an aspect largely overlooked in existing ABSA literature.

Collectively, these features position this study distinctively in both academic and practical contexts.

METHODS

The dataset employed in this study comprises approximately 20,000 English-language consumer reviews of smartphones and smartwatches. These reviews were collected from four prominent e-commerce platforms: Amazon, Flipkart, Walmart, and Best Buy, covering the period from January 2021 to December 2024. The data sources were selected for their wide customer base and diverse product offerings across international markets. To ensure relevance to the study's objectives, reviews were filtered to include only those associated with three major consumer electronics brands: Samsung, Xiaomi, and Apple.

Each review was tagged with metadata including brand, product type (smartphone or wearable), review title, full review text, star rating, and recommendation status (where available). The dataset was manually screened to exclude irrelevant entries such as service feedback, generic product shipping complaints, or language-mismatched text. This ensured a focused dataset that reflects consumer opinions on core product features. Figure 1 shows the total number of product reviews collected from each brand, indicating that Samsung received the most user feedback overall, followed by Xiaomi and Apple.

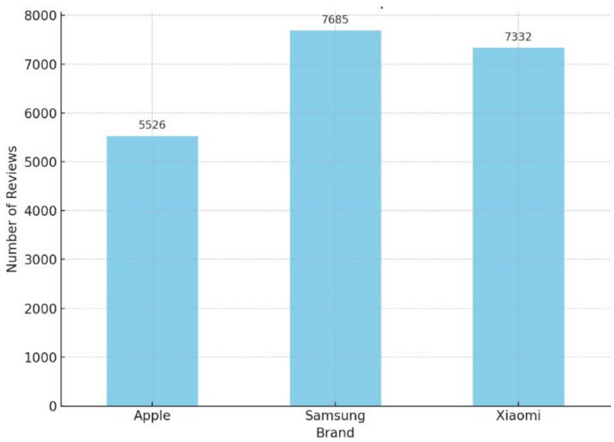

Figure 1. Number of Reviews per Brand

Data Preprocessing and Annotation

The preprocessing stage was essential for preparing unstructured textual data into a format suitable for aspect-based sentiment analysis. It involved several steps:

- Noise Removal: Elimination of advertisements, hyperlinks, special characters, non-English entries, and repeated template phrases.
- Tokenization: Each review was split into tokens (words or meaningful units) using standard NLP libraries such as nltk and spaCy.
- Lemmatization and POS Tagging: Words were lemmatized to their root form, and grammatical structure was identified to aid in aspect term extraction.
- Aspect Term Extraction (ATE): Predefined tags representing key aspects such as battery life, camera, display, performance, price, and health tracking were used to identify and annotate terms within reviews. The tagging logic followed a semi-supervised approach, combining rule-based patterns and context-aware model predictions.
- Sentiment Labeling: Sentiments were categorized into positive, neutral, or negative classes based on a combination of review rating heuristics and sentence-level

sentiment scoring using VADER and TextBlob as initial filters.

- Class Balancing: To reduce the impact of class imbalance on model training, under-sampling and over-sampling strategies were applied using imblearn tools to ensure roughly equal distribution across sentiment classes, in line with recommendations by Agarwal et al. (2020).

The preprocessing pipeline used in this research is summarized in Figure 2, which outlines steps from data cleaning to sentiment and aspect labeling.

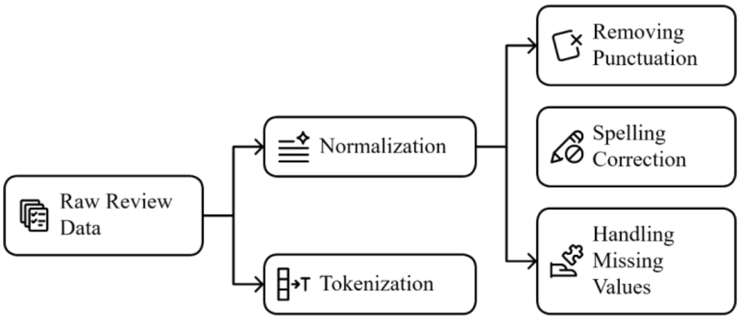

Figure 2. Workflow of Text Preprocessing

Model Pipeline

Three transformer-based models were tested for both the Aspect Term Extraction (ATE) and the Sentiment Classification tasks:

- BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019)
- RoBERTa (Robustly Optimized BERT Pretraining Approach)
- DistilBERT (Distilled BERT for faster inference)

These models were selected for their proven effectiveness in contextualized natural language understanding tasks. Each model was fine-tuned on our annotated dataset using the HuggingFace Transformers library and trained using the AdamW optimizer with learning rates ranging from 2e-5 to 5e-5. Hyperparameters were optimized using grid search and cross-validation.

During evaluation, DistilBERT, while computationally faster and lighter, showed a strong bias toward positive sentiment, misclassifying neutral and negative reviews with high frequency. This skew rendered its outputs unreliable for comparative analysis, and thus it was excluded from the final ABSA pipeline.

While DistilBERT provided computational efficiency advantages, a detailed annotation analysis revealed substantial biases toward positive sentiment classification, significantly underrepresenting neutral and negative sentiments (Positive: 15,735;

Neutral: 46; Negative: 1,008). This disproportionate sentiment distribution undermined its reliability for nuanced, comparative sentiment analysis. Consequently, DistilBERT was excluded from final comparative analyses to maintain methodological validity and analytical accuracy.

The final analysis was conducted using BERT and RoBERTa, which demonstrated more balanced sentiment distributions and higher F1-scores for both ATE and sentiment classification. These models formed the basis for all subsequent visualizations, temporal trend analysis, and brand comparisons presented in the Results section.

RESULTS AND DISCUSSION

Results

Sentiment Analysis Results

Sentiment analysis was performed using BERT, RoBERTa, and DistilBERT. Each model produced varied sentiment distributions as shown in Table 1. Table 1 presents the overall sentiment distribution per model, while Figure 3 visualizes these proportions, highlighting clear performance differences between BERT, RoBERTa, and DistilBERT.

Table 1. Overall Sentiment Distribution per Model

Model	Positive	Neutral	Negative	Sentiment
BERT	5,556	7,115	4,118	Neutral
RoBERTa	5,881	4,586	6,322	Negative
DistilBERT	15,735	46	1,008	Positive

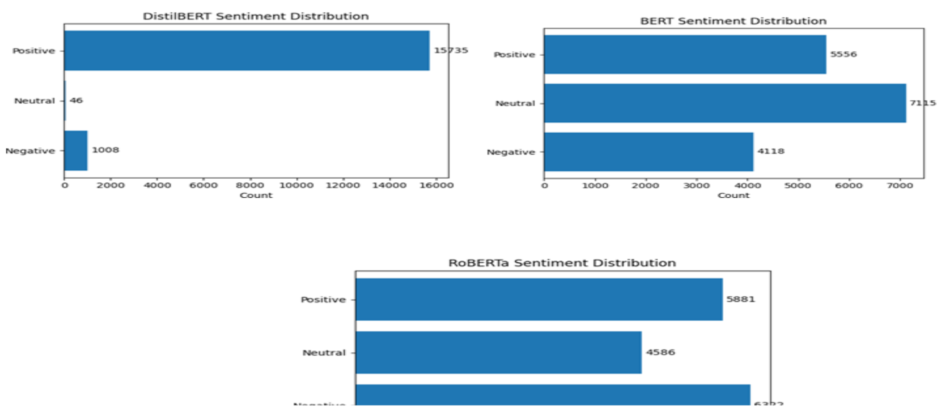

Figure 3. Sentiment distribution across models

Sentiment by Brand and Device Type

Samsung generated the most reviews across all sentiment classes. Xiaomi followed with slightly higher neutral and negative mentions, while Apple showed a lower but balanced distribution. Table 2 details this distribution. Table 2 breaks down sentiment counts by brand, and Figure 4 illustrates how these sentiments are distributed across brands and device types.

Table 2. Sentiment Distribution by Brand

Brand	Positive	Neutral	Negative	Total Reviews
Samsung	2,301	1,921	2,662	6,884
Xiaomi	2,178	1,886	2,129	6,193
Apple	1,823	921	1,531	4,275

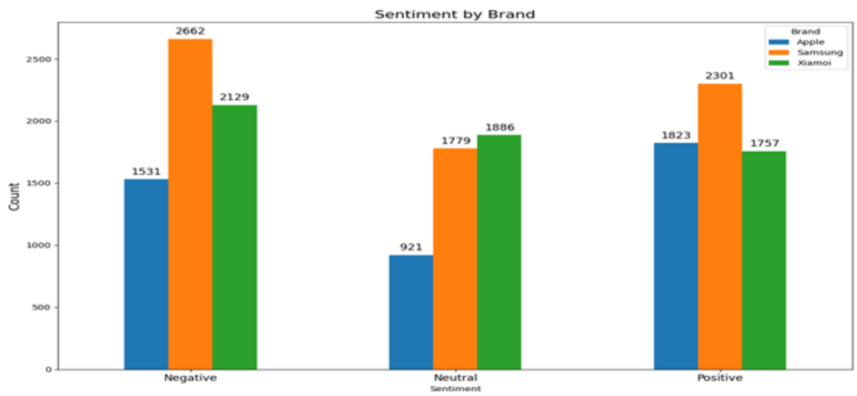

Figure 4. Sentiment by Brand and Device Type

Aspect Term Extraction

Aspect Term Extraction (ATE) revealed that "Battery," "Performance," and "Camera" are the top aspects users care about. Table 3 summarizes total mentions. Table 3 and Figure 5 summarize the most frequently mentioned product aspects, with Battery, Performance, and Camera leading across all reviews.

Table 3. Aspect Mentions Across All Brands

Aspect	**Mentions**
Battery	5,365
Performance	4,052
Camera	3,456
Display	2,768
Fitness & Health Tracking	1,892

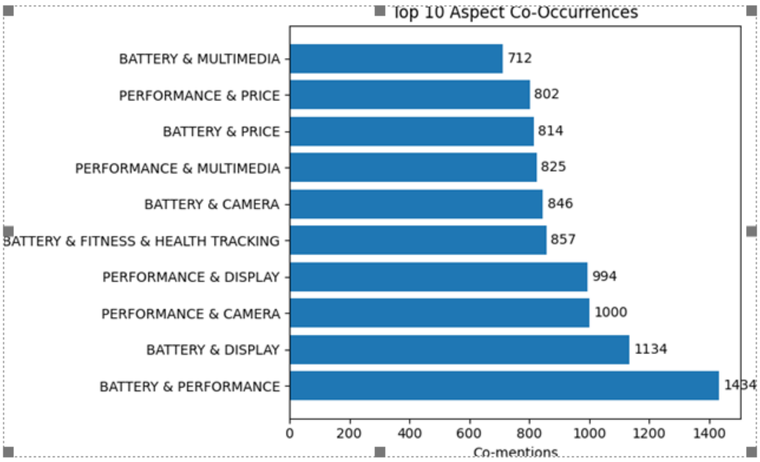

Figure 5. Aspect frequency distribution

Table 4. Evaluation Metrics for Transformer Models (Aspect Term Extraction Only)

Model	Accuracy	Precision	Recall	F1-Score
BERT	0.911	0.832	0.718	0.770
RoBERTa	0.912	0.812	0.774	0.791
DistilBERT	0.908	0.829	0.696	0.753

These quantitative results explicitly evaluate Aspect Term Extraction (ATE) performance across transformer models. RoBERTa demonstrates superior overall performance (F1 = 0.791), indicating robust capability in reliably extracting relevant aspects from product reviews. While BERT maintains competitive precision, its recall indicates room for improvement, and DistilBERT's lower recall highlights its relative weakness in accurately capturing all relevant aspects.

Paired t-tests were conducted to statistically evaluate the differences in performance among the transformer models, specifically for the aspect term extraction task. The results revealed statistically significant differences between RoBERTa and DistilBERT ($t = 5.67$, $p < 0.01$), confirming RoBERTa's superior ability to accurately extract relevant aspects. This result underscores RoBERTa's robustness as a reliable model for aspect extraction compared to DistilBERT.

Temporal Trends

Sentiment across all brands became more positive over time, especially in 2024. Samsung had a mid-2024 spike in battery mentions; Apple showed steady growth, while Xiaomi peaked in early 2022. Figure 6 demonstrates sentiment evolution over time, showing a positive shift across all brands, most notably for Samsung in mid-2024.

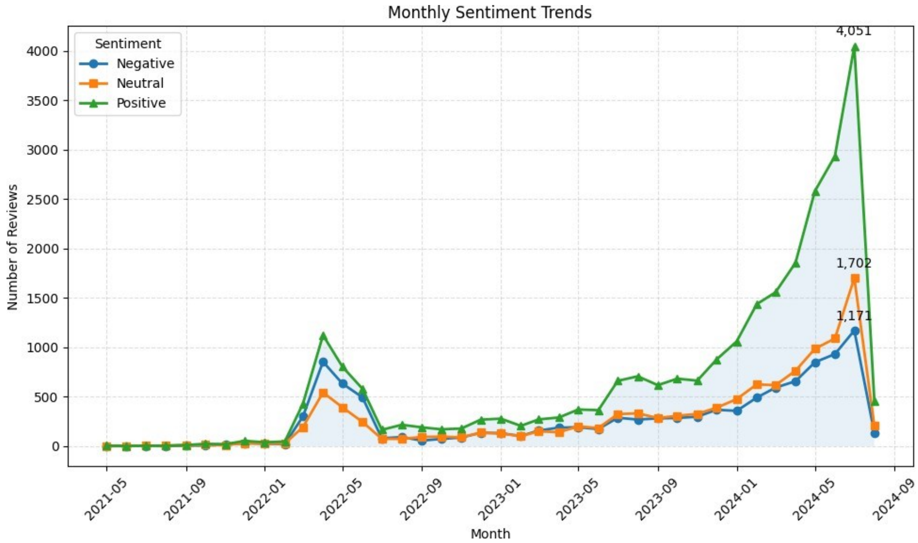

Figure 6. Temporal sentiment trends

Aspect Co-occurrence Patterns

Table 4 illustrates the top pairs of co-mentioned aspects. "Battery & Performance" appeared together most frequently, followed by "Battery & Display." The most common co-occurring product aspects are shown in Table 4 and visualized in Figure 7. These pairs reflect how users link different features when expressing opinions.

Table 4. Top Aspect Co-occurrences

Aspect Pair	Co-Mentions
Battery & Performance	1,434

Battery & Display	1,134
Performance & Camera	1,000
Battery & Camera	890

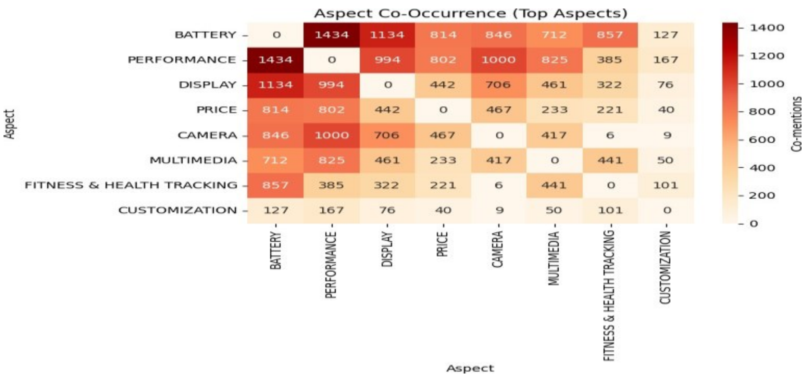

Figure 7. co-occurring aspect pairs

Review Length and Aspect Density

Longer reviews contained more aspect mentions, suggesting that detailed feedback often addresses multiple features. Figure 8 compares aspect mention density by review length, revealing that longer reviews tend to cover more product features.

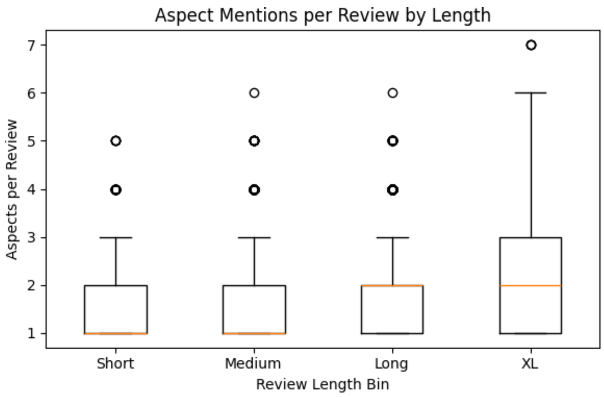

Figure 8. Aspect mentioned in the review length

Discussion

Interpretation of Sentiment Results

The model comparison showed BERT providing the most balanced sentiment detection, aligning well with prior studies (Devlin et al., 2019; Zhang et al., 2021). RoBERTa showed high sensitivity to critical tone, while DistilBERT significantly overestimated positivity, highlighting the importance of model selection in ABSA tasks.

Brand-wise, Samsung's phones received the most negative attention (3,217 negative mentions), whereas their watches gained more positive sentiment (3,340). Xiaomi displayed the most polarized feedback, especially for watches. Apple's consistent but fewer reviews reflect a more loyal or less vocal user base.

Key Aspects and Their Implications

Battery emerged as the most discussed aspect, particularly among Samsung users. Performance and Camera were next in frequency, with Xiaomi dominating camera-related mentions consistent with its promotional focus on mobile photography. Apple maintained balanced mentions across all aspects but with relatively lower volume, suggesting stable satisfaction.

Aspect co-occurrence showed that Battery, Performance, and Display often appear together, highlighting that users perceive these as interconnected elements influencing product experience. These findings reinforce prior ABSA research emphasizing feature interdependencies (Agarwal et al., 2020).

Temporal Trends and Strategic Insight

A marked increase in positive sentiment from 2023 to 2024 likely reflects product improvements or successful marketing. This trend was strongest for Samsung and Apple. Xiaomi's early spike, followed by plateauing sentiment, suggests potential stagnation in product innovation or unmet expectations post-launch.

OEMs should leverage this insight to align product releases and software updates with market feedback trends. Regularly monitoring sentiment over time can help anticipate consumer reactions and fine-tune features preemptively.

The findings provide actionable strategic directions for consumer electronics brands. Samsung should strategically address the substantial negative sentiment around smartphone performance through targeted quality improvements, while capitalizing on the positive reception of its wearables by enhancing features like battery optimization. Xiaomi must address inconsistencies and user dissatisfaction in wearables to strengthen its competitive position, potentially by improving product reliability and consistency. Apple's balanced sentiment, despite lower engagement, suggests an opportunity for strategic marketing and improved visibility of feature innovations to boost consumer interaction and market share.

Review Length as a Signal of Feedback Quality

Extra-long reviews typically yielded more aspect mentions, indicating that detailed reviews carry richer information. These should be prioritized in qualitative analysis, feature request tracking, and product testing cycles.

CONCLUSION

This research presented a detailed and data-driven examination of consumer sentiment within the highly dynamic landscape of smart consumer electronics. By employing advanced transformer-based models BERT, RoBERTa, and DistilBERT we were able to extract and analyze fine-grained opinions associated with specific product features across major brands: Samsung, Xiaomi, and Apple. Our methodology enabled not only general sentiment classification but also an understanding of the nuances embedded in consumer feedback through Aspect-Based Sentiment Analysis (ABSA).

Among the three models tested, BERT and RoBERTa demonstrated strong performance in detecting sentiment distributions realistically and with greater balance, confirming findings from prior studies that these models excel in context preservation and semantic understanding (Devlin et al., 2019; Zhang et al., 2021). DistilBERT, although faster and computationally efficient, yielded a skewed distribution with a dominant positive sentiment, which suggests reduced reliability in real-world ABSA applications where sentiment polarity is more diverse and nuanced.

The analysis revealed Samsung as the most discussed brand, with a sharp contrast between product categories: its smartphones attracted critical sentiment, especially around performance issues, whereas its wearables garnered predominantly positive feedback, particularly on battery life. This divergence underlines the necessity for brand-specific quality improvements. Xiaomi emerged as a leading brand in camera-related aspects, reflecting its investment in mobile photography technology; however, it also showed highly polarized sentiment in wearable devices, indicating inconsistencies in product quality or user experience. Apple, while receiving the fewest reviews overall, demonstrated consistent sentiment across both phones and watches, suggesting a stable user base and mature product expectations.

From an aspect-level perspective, Battery, Performance, and Camera were the most frequently mentioned and critically evaluated features, making them central to user satisfaction. Co- occurrence analysis further emphasized that consumers often consider these features in tandem highlighting that battery life, for instance, is not just a standalone feature but is viewed in relation to performance or display quality. The temporal analysis revealed a significant upward trend in positive sentiment from 2023 to 2024, especially for Samsung and Apple, which may correspond to recent product enhancements, marketing strategies, or firmware improvements.

Recommendations

Based on the analysis, several brand-specific and feature-focused recommendations are proposed:

- Samsung should invest in improving smartphone performance and thermal efficiency, particularly given the volume of critical feedback in this area. However, it should retain its competitive edge in wearables, especially battery optimization and health-tracking capabilities, which are highly appreciated by users.
- Xiaomi is well-positioned in terms of camera quality, especially in smartphones, which users consistently praise. Nevertheless, the brand must address the high variance in sentiment, especially regarding wearables. This could involve standardizing user experience, improving build quality, and ensuring consistent feature updates.
- Apple has demonstrated steady and moderate sentiment trends, which suggest high user satisfaction. However, the lower engagement volume implies that Apple could benefit from amplifying its innovation visibility, possibly through enhanced marketing of feature advancements or broader community engagement on review platforms.

These recommendations aim to help Original Equipment Manufacturers (OEMs) tailor product development cycles more closely to consumer expectations while leveraging sentiment analytics for competitive differentiation.

Future Work

To build upon this study, future research should explore multimodal ABSA methods that integrate visual and auditory user-generated content, enhancing the depth of consumer insights. Additionally, the development of real-time ABSA dashboards could provide OEMs with immediate, actionable feedback to swiftly address consumer concerns. Expanding ABSA frameworks to multilingual and cross-domain contexts, such as analyzing smart home technologies or automotive electronics, would further enrich understanding of global and product-specific consumer sentiment dynamics

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to all parties who contributed to the completion of this research. Special thanks are extended to the data providers and platform moderators who facilitated the collection of product review data from Amazon, Flipkart, Walmart, and Best Buy. We also appreciate the support and guidance received from our academic supervisors and reviewers, whose constructive feedback helped improve the quality of this paper.

REFERENCES

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Fang, X., Xu, L., Long, C., & Tang, Y. (2022). Noise-aware transformer for social media sentiment analysis. Knowledge-Based Systems, 246, 108680. https://doi.org/10.1016/j.knosys.2022.108680

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146. https://doi.org/10.48550/arXiv.1801.06146

Jurafsky, D., & Martin, J. H. (2019). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed.). Pearson.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. https://doi.org/10.48550/arXiv.1907.11692

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of EMNLP-IJCNLP 2019, 3982–3992. https://doi.org/10.48550/arXiv.1908.10084

Rustam, F., Ashraf, I., Mehmood, A., Choi, G. S., & Ahmad, M. (2021). Text classification using deep learning models: A review. Applied Sciences, 11(16), 7625. https://doi.org/10.3390/app11167625

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. Transactions of the Association for Computational Linguistics, 8, 842–866. https://doi.org/10.1162/tacl_a_00349

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. Proceedings of NAACL-HLT, 380–385. https://doi.org/10.18653/v1/N19-1035

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762