Global Scientific and Academic Research Journal of Economics, Business and Management
ISSN: 2583-5645 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journals-gsarjebm-home/

Leveraging Online Customer Reviews for Market Research: A Framework Based on Amazon Data

By

Sang-Uk Jung*
* Department of Marketing, Hankuk University of Foreign Studies, Korea.

Article History

Received: 11/07/2025
Accepted: 22/07/2025
Published: 26/07/2025

Vol –4 Issue – 7

PP: -99-102

Abstract

This study proposes a conceptual framework for leveraging online customer reviews, specifically Amazon product reviews, as a scalable and authentic data source for market research. Traditional methods such as surveys and interviews often involve high costs, time delays, and subjective bias. In contrast, Amazon reviews offer real-time, voluntary feedback from a large pool of consumers, making them a valuable source of market intelligence. The framework presented in this paper integrates approaches from user-generated content (UGC) research and text analytics to derive actionable consumer insights. Through illustrative examples, including 3M adhesive products and K-beauty cosmetics, we demonstrate the framework's potential to support product development, competitive analysis, and consumer segmentation.

Keywords: *Amazon reviews, consumer insight, market research, text analytics, user-generated content, K-beauty*

1. Introduction

In the era of digital transformation, market research is undergoing a profound shift. Traditional methods, such as surveys, interviews, and focus groups, have long served as reliable instruments for gathering consumer insights. However, they face growing limitations in scalability, timeliness, and authenticity. These challenges are further compounded by high costs, low participation rates, and the tendency of respondents to provide socially desirable answers, which is particularly problematic in today's fast-moving digital environment.

In contrast, the rise of user-generated content (UGC) has opened new avenues for understanding consumer sentiment at scale. Online platforms such as Amazon host millions of product reviews each year, providing spontaneous, experience-based feedback from real users [1][2]. According to recent estimates, Amazon hosts over 30 million product reviews in the beauty category alone [3]. These reviews represent an invaluable reservoir of organic, real-time market intelligence that, if properly analyzed, can complement or even substitute traditional market research techniques [4].

Despite their potential, online reviews remain underutilized in structured marketing decision-making. While businesses often skim through anecdotal feedback, few systematically analyze review data to extract actionable insights. This gap presents an opportunity to formalize a method for transforming unstructured consumer text into strategic intelligence.

This study introduces the Consumer Review Insight Extraction Framework (CRIEF), a structured approach for analyzing online reviews using text analytics techniques. Grounded in information adoption theory and market sensing capability, CRIEF provides a four-stage process: Signal Identification, Thematic Structuring, Perceptual Mapping, and Strategic Translation. These stages are designed to guide marketers and researchers in converting review data into actionable strategies.

To demonstrate the framework's applicability, we apply it to two product categories with distinct consumer engagement profiles: 3M sticky adhesive products and K-beauty skincare items. By doing so, we aim to show that online reviews are not only rich in qualitative depth but also scalable in analytical utility.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature on user-generated content and text analytics. Section 3 introduces the CRIEF framework. Section 4 presents two illustrative applications, and Section 5 concludes with implications and directions for future research.

2. Literature Review

Market research has traditionally relied on structured methods such as surveys, interviews, and focus groups to gather

*Corresponding Author: **Sang-Uk Jung** .

consumer insights. These approaches are well-suited for hypothesis-driven inquiries and targeted feedback collection. However, they are often constrained by issues such as small sample sizes, high implementation costs, long data collection cycles, and potential response bias [5]. In rapidly evolving digital markets, these limitations reduce their effectiveness in capturing dynamic consumer preferences.

As a response to these challenges, researchers have increasingly turned to user-generated content (UGC) as an alternative data source. UGC refers to consumer-created information publicly shared via platforms such as Amazon, Yelp, and TripAdvisor. These sources offer unstructured but highly authentic accounts of consumer experiences and opinions. Studies have shown that UGC can provide richer and more nuanced consumer insights than structured methods, particularly in terms of real-time relevance and spontaneous expression [5][6].

To analyze such large volumes of textual data, scholars have adopted computational methods including sentiment analysis, topic modeling, and natural language processing. Sentiment classification for movie reviews was first pioneered using supervised classifiers [7], and opinion summarization techniques were later applied to product reviews [8]. More recently, machine learning and deep learning approaches have enabled scalable extraction of consumer opinions from massive datasets [9].

Specific to Amazon reviews, several studies have examined how consumer feedback influences product perception, brand loyalty, and even sales performance [10][11]. Researchers have also used Amazon data to evaluate factors such as helpfulness of reviews, reviewer credibility, and linguistic markers of trustworthiness. However, many of these studies treat reviews primarily as predictors of purchase behavior, rather than as a source of strategic marketing insight.

This paper builds on and extends this body of work by proposing the Consumer Review Insight Extraction Framework (CRIEF), which positions online reviews not merely as behavioral predictors but as signals that support market sensing capabilities. Grounded in information adoption theory [12] and market sensing capability literature [13], the framework integrates review analytics into the broader context of strategic marketing decision-making.

Unlike prior studies that focus on isolated metrics or single-purpose sentiment analysis, this framework provides a systematic, multi-stage approach for extracting and translating consumer sentiment into actionable business decisions. It aims to bridge the gap between computational text analysis and managerial application, offering marketers a practical tool for leveraging unstructured data at scale.

3. Conceptual Framework: Consumer Review Insight Extraction Framework (CRIEF)

To address the underutilization of online customer reviews in structured marketing research, this study proposes the Consumer Review Insight Extraction Framework (CRIEF). The framework builds on existing theories in information adoption and market sensing capabilities, offering a structured process for extracting strategic value from unstructured textual data such as Amazon product reviews.

CRIEF is designed to guide both academics and practitioners in translating raw consumer narratives into actionable business intelligence. The framework is composed of four interrelated stages: Signal Identification, Thematic Structuring, Perceptual Mapping, and Strategic Translation.

3.1 Signal Identification

The first step involves identifying key signals embedded within consumer reviews. These signals may appear as frequently occurring terms, phrases marked by high sentiment polarity, or recently trending expressions. Reviews are filtered and pre-processed through natural language processing (NLP) techniques such as stop-word removal, tokenization, and lemmatization. Sentiment scores can be calculated using lexicon-based tools such as VADER and TextBlob, or through supervised machine learning classifiers trained on labeled review data. The outcome of this stage is a prioritized list of high-impact keywords or phrases that reflect consumer concerns, desires, or satisfaction drivers.

3.2 Thematic Structuring

Once the signals are identified, the framework employs unsupervised topic modeling techniques such as Latent Dirichlet Allocation (LDA) to cluster related terms and organize them into higher-order themes. These themes might include functionality, such as adhesive strength; experiential elements, such as skin irritation; and values, such as sustainability. This stage provides a mid-level abstraction layer that helps marketers understand what topics are emerging from the data without reading thousands of reviews manually.

3.3 Perceptual Mapping

The third stage translates thematic insights into perceptual maps or semantic networks. This is achieved by analyzing co-occurrence patterns among keywords and calculating semantic similarity using word embeddings or cosine similarity matrices. Sentiment heatmaps or two-dimensional perceptual maps can be used to visualize how product attributes such as hydration and scent are associated with specific brands or sentiment levels. This visual representation enables a comparative understanding of how consumers perceive different aspects of a product or brand.

3.4 Strategic Translation

The final stage converts analytical findings into actionable marketing strategies. This may involve identifying unmet needs, refining product features, tailoring communication to specific consumer clusters, or adjusting pricing strategies based on perceived value. For instance, if a segment of consumers frequently mentions “absorbs quickly” alongside positive sentiment, marketing messages can emphasize that attribute for targeted campaigns. Additionally, themes extracted in earlier stages can inform customer segmentation, persona development, and product roadmap planning.

*Corresponding Author: **Sang-Uk Jung** .

This four-stage CRIEF framework offers a replicable, theoretically grounded pathway for transforming unstructured consumer feedback into structured market insights. Unlike existing approaches that treat review data as secondary or anecdotal, CRIEF provides a holistic view that aligns with contemporary needs in data-driven marketing strategy.

4. Applications and Case Examples

To demonstrate the utility and adaptability of the CRIEF framework, we apply it to two distinct product categories from Amazon: 3M sticky adhesive products and Korean beauty (K-beauty) skincare items. These categories were selected due to their differing consumer priorities, with functionality emphasized in the case of 3M products and emotional and sensory attributes prioritized in the case of K-beauty. This contrast allows us to showcase the versatility of the framework.

4.1 Application to 3M Sticky Adhesive Products

We collected Amazon reviews for 3M-branded sticky products using keyword filters such as "adhesive," "command strips," and "removable hooks." After preprocessing and sentiment scoring, five high-frequency keywords were identified based on co-occurrence and sentiment strength.

Table 1 summarizes the frequency and average sentiment score of each keyword. Keywords such as "strong adhesive" and "damage-free" were consistently associated with high sentiment scores, reflecting consumer satisfaction with performance reliability. In contrast, "surface compatibility" had a lower sentiment score, indicating a common pain point among users.

Table 1. Top Keywords and Sentiment Scores for 3M Sticky Sticker Reviews

Keyword	Frequency	Avg. Sentiment Score
strong adhesive	42	0.89
removable	31	0.76
damage-free	26	0.81
reusable	18	0.65
surface compatibility	12	0.48

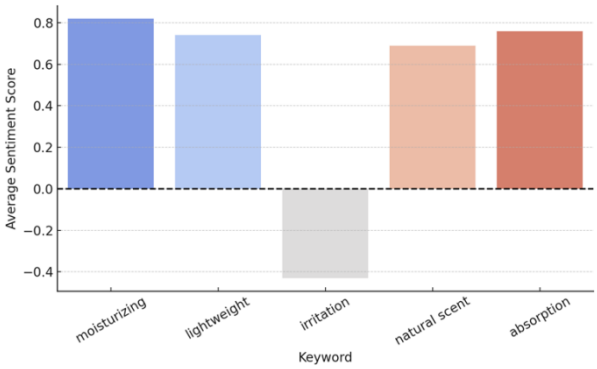
These insights can inform product refinement, such as improving compatibility with textured walls, and communication strategy, including the emphasis on damage-free removal in marketing messages.

4.2 Application to K-Beauty Skincare Products

Table 2. Top Keywords and Sentiment Scores for K-Beauty Product Reviews

Keyword	Frequency	Avg. Sentiment Score
moisturizing	58	0.82
lightweight	46	0.74
irritation	29	-0.43
natural scent	33	0.69
absorption	40	0.76

Figure 1. Sentiment Scores for K-Beauty Keywords



For K-beauty, we analyzed reviews for moisturizers and facial serums from top Korean skincare brands sold on Amazon. The data revealed five dominant keywords, with "moisturizing," "lightweight," and "absorption" receiving high sentiment scores. "Irritation," though less frequent, showed a strong negative sentiment and emerged as a potential risk area.

In Figure 1, we visualize sentiment distribution across these keywords, clearly distinguishing positive and negative themes. Figure 2 presents a wordcloud that highlights term prominence based on frequency.

These analyses help uncover not only product strengths, such as fast absorption and natural scent, but also areas that require caution or reformulation, including irritation reported by users with sensitive skin. CRIEF thus enables firms to map perceptual strengths and weaknesses and respond with targeted strategies.

Figure 2. Keyword Frequency in K-Beauty Product Reviews

5. Conclusion and Future Work

This study proposed a structured framework, the Consumer Review Insight Extraction Framework (CRIEF), for leveraging online customer reviews as a scalable, cost-effective source of market intelligence. Traditional market research methods require significant time, resources, and structured designs. In contrast, CRIEF enables marketers to

*Corresponding Author: **Sang-Uk Jung** .

extract real-time insights from unstructured review data in a more efficient and adaptive way.

By integrating theories from market sensing capability and information adoption, CRIEF delivers a theoretically grounded yet practically applicable process for extracting, structuring, and translating customer sentiments into strategic actions. The framework comprises four key stages: Signal Identification, Thematic Structuring, Perceptual Mapping, and Strategic Translation. Each stage is designed to guide marketers in making sense of consumer-generated data for better decision-making.

The practical utility of the framework was demonstrated through two contrasting case examples: 3M sticky adhesive products and K-beauty skincare items. In the former, CRIEF identified key performance-related terms such as "strong adhesive" and "removable," which aligned with consumer satisfaction. In the latter, the framework highlighted emotional and sensory themes, such as "moisturizing" and "natural scent," along with risk indicators like "irritation." These cases validate the framework's flexibility in addressing both functional and experiential product domains.

Beyond these illustrations, CRIEF can be extended to other industries and digital platforms, including mobile app reviews, e-learning course feedback, and social media product mentions. The framework is also adaptable to multiple languages and market contexts, making it suitable for international firms seeking cross-cultural consumer insights.

Future research may explore several directions. First, CRIEF can be quantitatively validated using larger datasets and predictive analytics to test its business impact. Second, integrating multimodal data—such as image reviews, star ratings, or time-series trends—may enhance the richness of insights. Third, longitudinal studies can use CRIEF to track changes in consumer sentiment over time and inform brand positioning strategies.

This study contributes to digital market research by showing how computational techniques can be applied in ways that are directly relevant to marketing practice. As the influence of online reviews continues to expand, structured frameworks like CRIEF can help firms translate scattered consumer opinions into clear, strategic direction.

Acknowledgement

This study is supported by Hankuk University of Foreign Studies Research Fund.

References

1. Schindler, R., & Bickart, B. A. (2010). Perceived Helpfulness of Online Consumer Reviews: The Role of Message Content and Style. *Journal of Consumer Behaviour*, 11, 234-243.
2. Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to Your Customers: Insights into Brand Image Using Online Consumer-Generated Product Reviews. *International Journal of Electronic Commerce*, 20, 112-141.
3. Kang, M., Sun, B., Liang, T., & Mao, H. (2022). A study on the influence of online reviews of new products on consumers' purchase decisions: An empirical study on JD.com. *Frontiers in Psychology*, 13.
4. Cluley, R., Green, W., & Owen, R. (2019). The changing role of the marketing researcher in the age of digital technology: Practitioner perspectives on the digitization of marketing research. *International Journal of Market Research*, 62, 27-42.
5. Nunan, D. (2017). Reflections on the Future of the Market Research Industry: Is Market Research having its 'Kodak Moment'?. *International Journal of Market Research*, 59, 553-555.
6. Li, C., Kwok, L., Xie, K. L., Liu, J., & Ye, Q. (2021). Let Photos Speak: The Effect of User-Generated Visual Content on Hotel Review Helpfulness. *Journal of Hospitality & Tourism Research*, 47, 665-690.
7. Sharma, A., & Ghose, U. (2023). Toward Machine Learning Based Binary Sentiment Classification of Movie Reviews for Resource Restraint Language (RRL)—Hindi. *IEEE Access*, 11, 58546-58564.
8. Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).
9. Jin, J., Liu, Y., Ji, P., & Kwong, C. K. (2019). Review on recent advances in information mining from big consumer opinion data for product design. *Journal of Computing and Information Science in Engineering*, 19(1), 010801.
10. Kaur, K., & Singh, T. (2021). Impact of online consumer reviews on amazon books sales: empirical evidence from India. *Journal of theoretical and applied electronic commerce research*, 16(7), 2793-2807.
11. Rashid, A., & Huang, C. Y. (2021). Sentiment analysis on consumer reviews of Amazon products. *International Journal of Computer Theory and Engineering*, 13(2), 7.
12. Christofi, M., Khan, H., Zahoor, N., Hadjielias, E., & Tarba, S. (2024). Digital Transformation of SMEs: The Role of Entrepreneurial Persistence and Market Sensing Dynamic Capability. *IEEE Transactions on Engineering Management*, 71, 13598-13615.
13. Bharadwaj, N., & Dong, Y. (2014). Toward further understanding the market-sensing capability–value creation relationship. *Journal of Product Innovation Management*, 31(4), 799-813.

*Corresponding Author: **Sang-Uk Jung** .