Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
June 2024 5(6), 3548–3563

READING BETWEEN THE LINES: A QUALITATIVE ANALYSIS OF ONLINE PRODUCT REVIEWS

Medha Sood [1], Dr. Girish Taneja [2]

[1] Department of Commerce Business Management and Economics, DAV University Jalandhar Punjab India & KLB DAV College for Girls Palampur, Himachal Pradesh, India
[2] Department of Commerce Business Management and Economics DAV University Jalandhar, Punjab, India

Corresponding Author

Dr. Girish Taneja,
tanejagirish@gmail.com

DOI
10.29121/shodhkosh.v5.i6.2024.6412

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Overview: The rise of e-commercehas created a more interactive and competitive environment for both sellers and buyers. In recent times, a heavy reliance on online product reviews posted on the e-commerce website by existing and verified users has been witnessed. From the extant literature, it is evident that the impact of non–textual or quantitative aspects of online product reviews (OPR), such as volume, valence, star rating, helpful votes, etc., on buying behaviour, purchase intention, sales, etc. has been vastly studied. But since the past decade, the importance of analysing and understanding the textual data of OPR has been strongly recommended and explored.
Aim: Reviewers express their opinions through non-structured and unfiltered reviewswhere the information presented goes beyond mere words.The review text has embedded emotions and sentiments also. The current research aims to investigate and comprehend the implicit information conveyed by consumers in the unstructured and heartfelt reviews on Amazon.in. The study seeks to identify key semantic aspects affecting OPRs' assessment by highlighting latent topics, sentiments and intricacies within the textual content of the reviews.
Methodology: Amazon is the most popular e-commerce site in India, and its review system is widely seen as reliable, clear, and standardised. Its multi-dimensional format—textual reviews(review statements and title statements), aggregate star ratings which guarantee consistency across many product categories and offers a solid foundation forqualitative research. This study employed a multistage sampling technique to extract 5,900 reviews from 59 top-selling products across three best-selling categories: beauty, fashion, and electronics from Amazon India. Web scraping has been used to extract the review data using Python as a programming language. A qualitative content analysis, topic modelling, sentiment analysis and sentiment score analysis have been employed using various packages and functions in R Studio.
Findings: The research reveals that both negative and positive sentiments have significant effects on product ratings. The word count analysis indicated a predominant use of positive words such as 'good', 'product', 'quality', 'nice', and ' price'. The three distinct topics identified through topic modelling demonstrate that reviews are shaped by a combination of functional utility, sensory experience, and detailed product attributes. The sentiment analysis revealed that positive sentiment is more common than negative sentiment. Additionally, emotions like trust, anticipation and joy were predominant, while negative emotions such as disgust and anger were less frequently observed. A key finding is the asymmetrical effect, where negative sentiment has a notably stronger negative impact than positive sentiment.Although statistically significant, the relatively low R-squared value suggests that sentiment scores alone account for only a small part of the variance in product ratings, indicating that ratings are a complex outcome influenced by multiple factors beyond expressed sentiment.
Implications: The study highlights the multifaceted role of online product reviews in shaping consumer behaviour on e-commerce platforms. The predominance of positive emotions emphasises the value of building strong consumer confidence. At the same time, the asymmetrical effect of negative sentiment urges the need for businesses to address

the dissatisfaction promptly. For practitioners, the findingssuggest that effective product management involves not only mitigating negative reviews but also actively leveraging positive consumer emotions and experiences to strengthen brand loyalty. Strategically, e-commerce businesses should adopt a dual approach—proactively managing negative feedback while amplifying positive narratives—to enhance customer satisfaction and trust. Academically, the research contributes by evidencing the interplay of emotional, semantic, and many other factors in review analysis, offering pathways for future studies to incorporate richer variables and advanced text mining techniques.

Keywords: Online Product Reviews, Text Analysis, Sentiment Analysis, Content Analysis, Predictive Analysis, Topic Modelling, E-Commerce Platform

1. INTRODUCTION

The digital transformation of commerce has fundamentally altered the landscape of consumer decision-making, with online product reviews emerging as one of the most influential factors in buyingbehaviour.Traditional word-of-mouth recommendations have evolved into dynamic digital ecosystems of user-generated content. It has positioned reviews as a primary source of information in consumer evaluations. Recent evidence shows that 98% of consumers consult online reviews most frequently, while only 2% report never engaging with them, a sharp decline from 13% in 2020 (BrightLocal, 2023). This trend highlights the increasing rise of OPRs as a dominant medium of digital word-of-mouth.

The scale of review usage further demonstrates their significance. Google accounts for 73% of global online reviews, with 81% of consumers relying on the platform to assess local businesses, increasing from 63% in 2020 (BrightLocal, 2024). This growth reflects the increasing importance of reviews in consumer evaluation processes. Beyond shaping purchase intentions, reviews also generate broader business implications.U.S. retailers incurred $428 billion in return costs in 2020, and review-based analytics are now widely employed to identify category-specific causes of returns and mitigate operational inefficiencies (National Retail Federation, 2020). Advanced analytical approaches, such as topic modelling, have demonstrated the ability to extract actionable insights by linking narrative feedback to concrete product-related outcomes, reinforcing the strategic value of review analysis.The evolution of online reviews from simple rating systems to sophisticated multidimensional information sources has demanded advanced analytical techniques. This multi-dimensional nature of review content requires analytical approaches that can effectively capture and interpret the rich and salient information embedded within consumer narratives. Qualitative analysis of text, sentiment analysis, and topic modelling have been successful in discovering latent patterns, affective tone, and thematic structure inherent in consumer reviews. Such techniques provide researchers with further insights into product performance, consumer decision-making, and market behaviour.However, platform-specific variations introduce additional complexity. Consumer trust in reviews differs across platforms: 67% for Google, 47% for Amazon, 41% for Yelp, and 34% for Facebook (Bizrate Insights, 2024). These disparities reflect the interplay between platform policies, verification processes, and consumer perceptions of authenticity, underscoring the importance of contextualising review analysis within platform-specific dynamics.

Against this backdrop, the present study undertakes a comprehensive qualitative examination of online product reviews. By integrating qualitative text analysis, sentiment analysis, and topic modelling, the study aims to advance theoretical understanding of consumer behaviour in digital environments while generating practical insights for online businesses seeking to leverage review data more effectively. The current study makes a theoretical contribution to the understanding of online consumer behaviour, as well as gives operational insights into harnessing information fromonline review data.

2. REVIEW OF LITERATURE

Online commerce has dramatically transformed consumer behaviour, opening up new ways to understand how people assess, experience, and talk about products and services. Online product reviews are one of the most influential forms of user-generated content. The conceptual understanding of online product reviews has undergone significant evolution since their inception in 1999. Consumers use online product reviews not only to share their opinions and experiences but also to evaluate the product. Consequently, online reviews play a pivotal role in product marketing strategies, as companies use them to assess consumer attitudes and perceptions. The convergence of physical and virtual environments has given a chance to online marketers like Amazon to radically integrate the customer experience. The customer-generated content, as Electronic word of mouth (eWOM), not only acts as a value provider but also as a value creator for the customers.A customer refers to online reviews, considering them a relatively good source of a first–hand usage experience and believes them to be fruitful in assessing product quality (Salehan& Kim, 2016; Chen & Xie, 2008).This reliance stems from the perceived objectivity and authenticity of peer-generated content, which instils confidence and mitigates perceived risks associated with online shopping (Lee et al.,2011). Conversations from such user-generated content, along with other forms of marketing communications, can be both a precursor to and outcome of sales in an online marketing environment. Given their importance, online reviews have been studied across disciplines—by computer scientists, marketers, psychologists, economists, and information theorists—who aim to understand their broad and complex impact (Hennig-Thurau et al., 2004). Extensive research confirms the broader business value of OPRs, beyond simply influencing sales and business performance.

Content analysis has emerged as a fundamental methodological approach for extracting meaningful insights from online product reviews, with its application in consumer research spanning over four decades of methodological development and refinement (Vespestad & Clancy, 2021).Bell et al. (2022) defined content analysis as "an approach to documents that emphasises the role of the investigator in the construction of the meaning of and in texts" This definition highlights the interpretive nature of content analysis and the vital role of researcher expertise in extracting meaningful insights from textual data. The application of content analysis to online product reviews has been particularly valuable in the assessment of verbal and nonverbal messages within consumer communications. Kolbe and Burnett (1991) established foundational principles for content analysis in marketing and consumer behavior research, emphasising the importance of systematic categorisation and reliable coding procedures. The methodology has shifted from purely quantitative approaches that seek to "quantify content in terms of predetermined categories and in a systematic and replicable manner" to more advanced qualitative approaches that stress systematic and analytical interpretation while allowing flexibility in developing and refining categories (Bell et al., 2022; Hsieh & Shannon, 2005; Krippendorff, 2018).A mixed method approach is advocated by many studies utilising both quantitative and qualitative content analysis for understanding the drivers of purchase decisions and culturally specific context (Wahpiyudin et al., 2022). In hospitality, large-scale qualitative content analysis of booking-site reviews has been used as a quality management tool (e.g., SERVQUAL verification on 167,000 hotel reviews), demonstrating how qualitative categories can be effectively utilised (Wąsowicz-Zaborek, 2023). Alzate et al. (2022)have used the text mining analysis based on a lexicon-based approach, the Linguistic Inquiry Word Count (Pennebaker et al., 2007), which provides the researcher with insights intoemotional and psychological brand associations.Building on such lexicon-based approaches that decode content and psychological associations from textual data, researchers have increasingly explored more advanced computational techniques like topic modelling to capture deeper latent structures and thematic patterns within online product reviews.Application of topic modelling using the foundational technique Latent Dirichlet Allocation (LDA) has enabled researchers and practitioners to identify latent thematic structures within review datasets and extract actionable insights about consumer preferences and product performance (Blei et al., 2003).The systematic literature review of topic modelling techniques in business applications revealed a significant shift in research focus toward service quality and customer satisfaction based on online customer reviews (Kim et al., 2022; Lopez & Garza,2022).

Another very popular technique is Sentiment analysis (opinion mining), which is a computational technique used to systematically identify, extract, quantify, and study affective states and subjective information from text. In the context of e-commerce, sentiment analysis provides a robust methodology to process the vast amounts of unstructured textual data found in OPRs, transforming qualitative feedback into actionable insights. Early research in sentiment analysis focused on lexicon-based approaches, where predefined dictionaries of words with associated sentiment polarities (positive, negative, neutral) are used to score text (Mohammad& Turney, 2010).Many studies have given evidence (Li, Wu, &Mai, 2019; Ravi & Ravi, 2015; Jaichandran et al., 2019) to support the importance of sentiment analysis in understanding online reviews.Prior studies have consistently demonstrated the utility of sentiment analysis in understanding consumer preferences and predicting market trends. For instance, sentiment analysis has been applied to identify key product attributes that drive customer satisfaction or dissatisfaction, allowing businesses to prioritise product development efforts (Ghose &Ipeirotis, 2011; Ullah et al., 2016; Ghasemaghaei et al.,2018; Abighail et al.,2023;

Weng & Zhao, 2020; Ullal et al., 2021).The extant studies have reported the significant impact of online product review sentiments on product sales (Wu et al., 2018; Fan et al., 2017; Yu et al., 2012).The ability to classify reviews by sentiment enables automated monitoring of brand perception and competitive analysis, providing a dynamic overview of market positioning.

E-commerce and qualitative content analysis in Emerging Markets: The Indian Context: The rapid growth of e-commerce in emerging markets, particularly India, presents unique opportunities and challenges for businesses. The Indian online retail landscape is characterised by a diverse consumer base, varying digital literacy levels, and distinct cultural nuances in communication and consumption patterns. Understanding consumer sentiment in this context requires not only advanced analytical tools but also an appreciation for local specificities. Recent research has begun to explore sentiment analysis in the Indian e-commerce sector, with studies focusing on specific categories like fashion, utilising advanced NLP techniques to unravel customer sentiment (Lopez & Garza, 2022; Westerlund etal., 2019). These studies highlight the importance of sentiment models for cultural and regional specifications and the potential for deep insights into consumer preferences and market dynamics within these rapidly evolving economies.

This review of literature establishes the theoretical and empirical foundations for our study, positioning our analysis of Amazon India reviews within the broader academic literature on sentiment analysis and e-commerce. It highlights the critical role of OPRs, the power of sentiment analysis, the asymmetrical impact of positive and negative opinion and the specific considerations for emerging markets like India.

3. METHODOLOGY

The current research aims to investigate and comprehend the implicit information conveyed by consumers in the unstructured and heartfelt reviews on amazon.in. The study seeks to identify key semantic aspects affecting OPRs' assessment by highlighting latent topics, sentiments and intricacies within the textual content of the reviews.

3.1. DATA COLLECTION AND SOURCING

The current research has taken into consideration the leading marketplace in India: Amazon India.The number of annual online shoppers in India was estimated to be approximately 345 million in 2023 (Statista, 2023). As of 2022, Amazon India is a market leader in the Indian e-commerce market with more than 3.2 billion visitors per month (Statista, 2022). The rapidly growing e-commerce sector in the country attests to the promising potential of the Indian e-commerce market. Thus, for the current study, information related to the various products, categories and online product reviews were extractedfromwww. amazon.in. A fair, transparent and trusted online product reviews system makes Amazon India a very lucrative and promising choice for the study. Further, a multistage sampling technique was used to draw samples from amazon.in.

The study employed a meticulously designed multistage sampling technique to ensure a robust and meaningful dataset. For thesampling purpose, identification of Amazon best-selling categories (Beauty, Electronics, Fashion), followed by risk classification based on return policies (Shiprocket, 2020), was done. From each category, the top 100 products were extracted using Best Seller Rank (BSR), refined by subcategories. Further classification basedon lifecycle stage (new or mature) and price (low or high) was done. The final sample comprised 59 products, from which 5,900 reviews were extractedfromamazon.in.

3.2. METRICS, TOOLS AND METHODS USED

For sampling purposes, Keepa Price Tracker has been used to classify products as low or high price. To extract the reviews, web scraping was done using Python as the programming language.

The current study has used the review text statements and star ratings as the key determinants of online product reviews. The review text statement is the detailed textual content of the review, where consumers express their opinions, experiences and sentiments. The star ratings are the numerical ratings assigned by experienced customers (reviewers), typically on a scale of 1 to 5. The star ratingis a reflection of the overall satisfaction or dissatisfaction of the reviewer with the product.

3.3. QUALITATIVE TEXT ANALYSIS USING R STUDIO

A qualitative text analysis was done using R Studio. To understand the qualitative aspect of online product reviews, content analysis, topic modelling, predictive analysis and sentiment analysis were done (Welbers et al.,2017). A text file of the top 5900 reviews from the best-selling products was extracted from amazon.in. These products belong to the beauty, electronics and fashion categories, which are the top bestselling categories on the marketplace. The bestselling skin and hair care products were taken from the beauty category, whereas for the fashion category, clothing and accessories were taken, and bestselling electronics products were considered for the analysis.

To uncover the latent semantic dimensions of Online Product Reviews (OPRs), the text data underwent a thorough cleaning and was organised into a corpus (a collection of texts serving as a dataset for linguistic analysis). Subsequently, through tokenisation, a 'bag of words' was assembled for further analysis. Recognising that OPRs extend beyond mere words, a comprehensive methodology was adopted to unveil the underlying unstructured and implicit meanings within these reviews. A content analysis, topic modelling, and sentiment analysis were conducted across all reviews collectively. Furthermore, a comparative examination was conducted to elucidate qualitative distinctions embedded in reviews across the three categories.

Figure 1. Order of text analysis operations for data preparation and analysis.

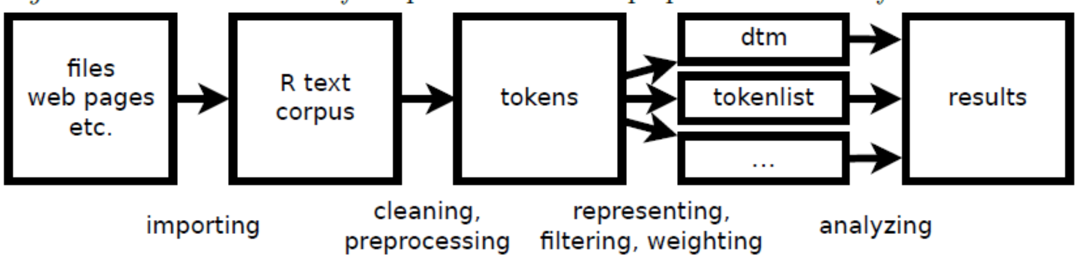

Figure 1 Source: Welbers etal., (2017) plan of quantitative text data analysis

Qualitative Content Analysis: Aunivariate text analysis ofa dataset of 5,900 reviews of top-rated products belonging to the best seller categories: beauty, fashion, and electronics of Amazon India, was carried out. Further, tokenisation and review processing were carried out using the quanteda package of R Studio. Word count determining the most common terms wasobtained as approximately 7,000 unique features, with descending order of their occurrence. The top 52 terms with at least 100 occurrences were shortlisted for further study. Visualisation was carried out using quanteda'stextplotfunction to produce both a common word cloud for the overall corpus and comparison clouds across the three product categories. A feature co-occurrence matrix using textstat functions with at least 0.5 frequency cut-off for determining the top 30 co-occurring terms also emerged.

Topic modelling with Latent Dirichlet Allocation (LDA) has been utilised for analysing the latent thematic structure of the input text from the product reviews. It follows an inductive approach, wherein rather than using an a priori system of coding, the computer programme itself discovers informative codes from a text, seeking patterns of word co-occurrence and establishing latent factors (e.g., topics, frames, writers) that explain these patterns on a quantitative level (Welbers et al, 2017). To determine the optimal number of topics, the perplexity score method was utilised, which is advisable for the computation in R (Blei et al.,2003)Thestm package, quantedapackage, and the textmodel function were utilized to generate summaries of the top 10 words for each topic, displaying words in proportion to their prevalence within each topic.The sentiment analysis method is an unsupervised method, a lexicon-based approach, that can classify online consumer reviews based on the overall semantic orientation of their content. Therefore, for this study, the sentiment analysis was conducted at the consumer review level. For conducting sentiment analysis syuzhet package in R Studio to get the word count assigned to each emotion. Theggplot2 package has been used to visualise the result of emotions word count. NRC word emotion association lexicon (EmoLex), which classifies the words with the 8 basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and 2 sentiments (positive and negative) (Mohammad & Turney,2010).

Further, a sentiment score analysis has been conducted on thetext file of products, using linear regression models forestablishing the relation between sentiment scores (negative and positive) inferred from the product text and their corresponding productrating. Understanding these relationships is crucial for identifying the influence of linguistic elements in product descriptions on customer perception and ultimately leading to product success.

4. DATA ANALYSIS AND INTERPRETATION

Content Analysis: Qualitative content analysis of the dataset, having 5,900 reviews, was conducted by performing word count analysis, word cloud analysis (common and comparative), andfeature co-occurrence matrix analysis.

Table 1 Word Count (Reviews Text Data)

Word	Count	Word	Count	Word	Count	Word	Count
Good	2882	Skin	326	Material	241	Working	175
Product	1723	Just	318	Sound	236	Amazon	164
Quality	1127	Phone	296	Fragrance	233	Amazing	162
Nice	730	Money	292	Value	228	Little	162
Price	619	Buy	288	Overall	227	Get	160
Use	494	Comfortable	276	Perfect	222	Love	10
Like	489	Great	267	Battery	209	Now	155
One	470	Time	256	Watch	201	Better	155
Also	449	Long	253	Colour	193	Color	155
Using	388	Size	253	Much	187	Wash	153
Hair	357	Fit	247	Fabric	181	Range	151
Really	330	Well	246	Used	179	Perfume	145
Best	330	Worth	246	Got	178	Looks	145

Figure 2 Common Word Cloud

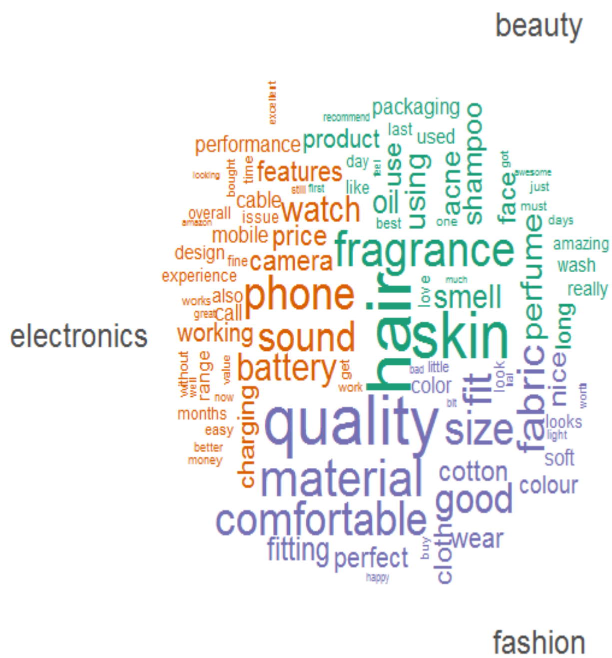

Figure 3 Comparison of Word Cloud Analysis

The results of word count analysis (Table 1)showed that the words such as ‘good’ (2882), ‘product’ (1723), ‘quality’ (1127), ‘nice’ (730), and ‘price’ (619) were found to be used the maximum numberof times in the text data. These findings indicate that consumers are more inclined toexpress their satisfaction by using positive words like good, nice, best, perfect, better,great, and amazing, signifying the importance of “experiential value” sought by theconsumers. To elaborate on their experiences, they prioritise words like ‘money,price, worth, range’, emphasising the significance of “economic value”. Certainwords, such as material, fragrance, fabric, smell, long, and colour, highlight the productattributes, indicative of the pertinence of the “functional values”.Figure 2 illustrates the univariate analysis using the Quantedatextplot function to visualise the results in the form of a word cloud, highlighting the essence of thecorpus. The size of words like ‘good’, ‘product’, ‘quality’, ‘nice’, and‘price’ corresponds to their frequency of occurrence in the corpus. The resultssuggest that words like ‘good’ indicate that most reviewers want to express theirpositive experience with the product.

Figure 3 presents a comparison cloud of the three categories (Beauty, Fashion, andElectronics), visualised using the same quanteda package and function. Features withsimilar colours indicate their belonging to one category. In the visual plot, green-colored words belong to the beauty category, while blue and red belong to the fashion andelectronics categories, respectively. The size of the words indicates their relativefrequency in the respective categories.

In the beauty category, the size of the word ‘hair’ indicatesits highest frequency of occurrence, followed by ‘skin’, ‘fragrance’,‘perfume’, ‘smell’, ‘acne’, and ‘shampoo’. The results indicate thatreviewers are inclined to mention products, their benefits, and sensory appealsexperienced in the reviews they share.In the fashion category, words like ‘quality’, ‘comfortable’, ‘material’,‘fabric, ‘size’, ‘fit’, ‘good’, and ‘cotton’ stand out in the opinionsposted by consumers. The results indicate that consumers share insights aboutthe product’s quality, fabric, or material, highlighting the importance offunctional values for customers buying fashion products. Words like ‘good’,‘perfect’, ‘nice’, and ‘happy’ indicate a positive experience with theproducts, suggesting that happy consumers express their satisfaction usingwords indicative of their experience, thus driven by experiential values.

In the electronics category, words highlighting theproduct type, such as ‘phone’, ‘watch’, ‘camera’, and ‘mobile’, arefrequently used. Words highlighting product attributes such as ‘features’,‘design’, ‘sound’, ‘battery’, and ‘charging’ are also quite common.Terms like ‘price’, ‘performance’, and ‘working’ suggest that consumers emphasise the functional characteristics of electronic products. This indicatesthat the locus of interest lies in “functional values” when purchasingelectronics.These results suggest that consumers are willing to express their opinions andsatisfaction with beauty and fashion category goods, rooted in psychological,economic, and functional values. In contrast, for electronic products, functional valuespredominantly drive buying decisions and post-purchase judgments.

5. FEATURE CO-OCCURRENCE MATRIX

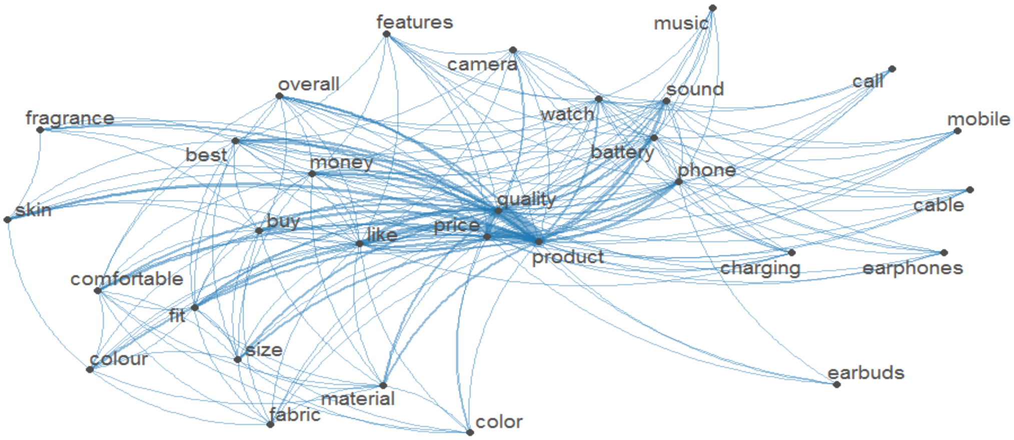

Figure 4 Feature Co-Occurrence Matrix

Visualisation was done to plot the co-occurrence matrix. Figure 4 displays the result of the feature co-occurrence matrix. A text plot analysis with a minimum frequency of 0.5 was generated with the top 30 featured words using quanteda text.stats package. It is clear from the plot that the word 'earbud' is connected to 'phone', 'price', 'product', and 'quality', suggesting that 'earbud' is present with these fourwords in the same document. This does not imply direct proximity but rather cooccurrencewithin the same text. The frequency of co-occurrence is indicated by thethickness of the lines; higher frequency results in thicker lines. The plot indicates anoticeable thickness in the lines connecting terms such as 'product' and 'quality,' 'product' and 'price,' as well as 'product' and 'skin.Thus, the network plot indicates co-occurrence and frequency of co-occurrence offeatures, suggesting that words reflecting the characteristics of the products are co-occurringin the comments.

6. TOPIC MODELLING

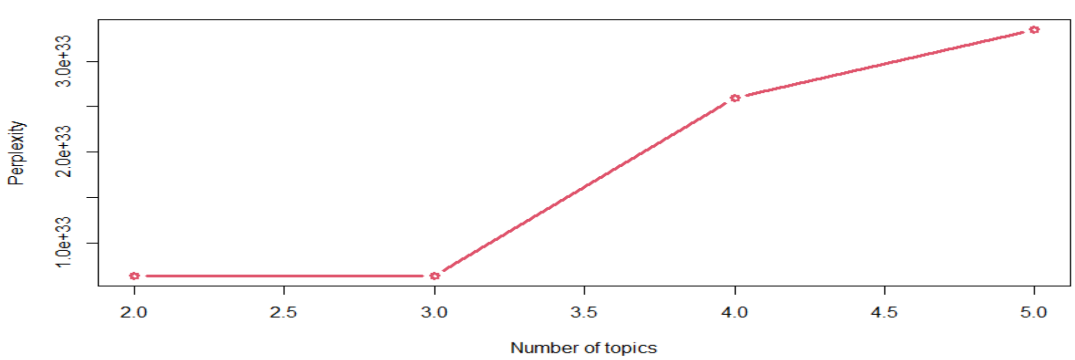

Figure 5 Predicting the Number of Topics in Text Data

A scree plot (Figure 5) was plotted with the probable number of topics on the x-axis and the perplexity score on the y-axis. A lower perplexity score is taken for the probable number of topics. As a result, 3 topics were found to emerge in the dtm(document term matrix).

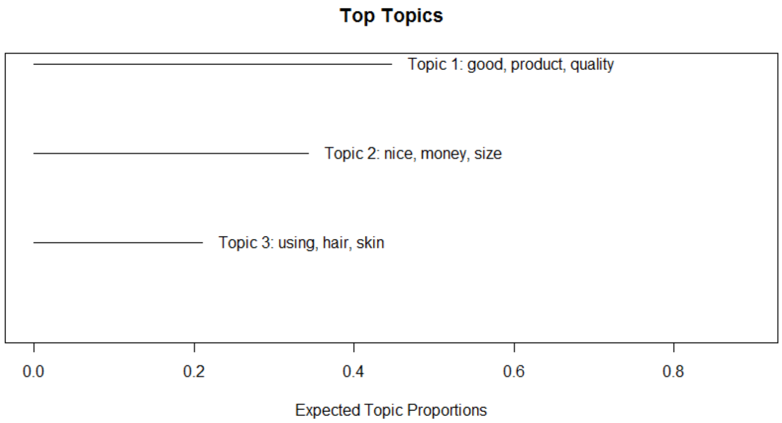

Figure 6 Expected Topic Proportion: Topic modelling

Topic modelling was performed on three topics using the top features extracted from 5,740 documents, with a dictionary comprising 6,995 words for further labelling and interpretation. For analysis purposes, the stm package, quantedapackage, and the textmodel function were utilised to generate summaries of the top 10 words. Figure 6 displays the words in proportion to the three topics. The results can be interpreted in four different ways:

Topic 1 Top Words:

Highest Prob: good, product, quality, price, best, also, one

FREX: good, mobile, delivery, mouse, superb, item, original

Lift: accha, alexa, allen, amazon, good, ambrane, amp, apple

Score: good, product, quality, price, received, best, sound

Topic 2 Top Words:

Highest Prob: nice, money, size, comfortable, material, fragrance, worth

FREX: fragrance, fabric, perfume, smell, soft, hai, cotton

Lift: aaya, agar, amber, antiperspirant, aromatic, bahot, baki

Score: nice, material, fabric, fragrance, perfume, cotton, jeans

Topic 3 Top Words:

Highest Prob: using, hair, skin, use, battery, watch, phone

FREX: hair, skin, watch, features, oil, shampoo, acne

Lift: इसके, कोई, दिया, देख, फोन, बात, मैं

The current analysis has used the Frex method for interpretation, which is a widely recommended method.

Table 2 Topic labelling

Topic	Label	Top Features	INDICATION
Topic 1	**Functional Values**	*Frex: good, mobile, delivery, mouse, superb, item, original.*	*Functional aspects of the products are the locus of interest for the customers*
Topic2	**Experiential Value**	*Frex: fragrance, fabric, perfume, smell, soft, cotton, smells*	*Reflects that consumers like to share their sensory experiences in their comments*
Topic 3	**Product Description**	*Frex:* hair, skin, watch, features, oil, shampoo, acne	*Describing the product features is important for customers*

Sentiment Analysis: By using the syzuhet package in R, each word from the text data was assigned an emotion or sentiment. Further, a sentiment score analysis was conducted on a text file of products using the sentimentr function in R Studio

Table 3 Word Count for the Emotions

Emotion	Count
"Positive"	953
"Negative"	582
"Trust"	498
"Anticipation"	328
"Joy"	326
"Sadness"	263
"Fear"	231
Anger"	204
"Surprise"	183
"Disgust"	164

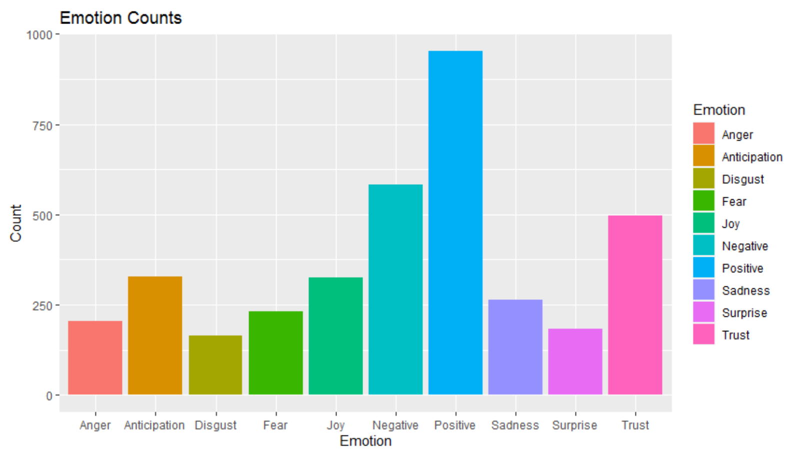

Figure 7 Emotion Counts

Table 3 and Figure 7 illustrate the results from the sentiment analysis on the dataset of 5,900 review texts from the three categories, exhibiting very clearly that 'positive' sentiment was found to be more frequently expressed, followed by words suggesting 'negative' sentiment. It was found that emotions like 'trust', 'anticipation' and 'joy' are the three top emotions embedded in the text dataset. The words expressing an extremely negative emotion, such as 'disgust', are the least frequently used by consumers in their reviews. The findings strongly suggest that consumers exhibit more inclination to express their positive experience through sentiments and emotions within their reviews. The expression of negative emotions, such as disgust and anger, is less prevalent in the top reviews. Interestingly, the results indicate that customers tend to project their trust in the product in their comments, which is a notable revelation. The use of words suggestive of trust in the product can have a remarkable influence on prospective customers, as it mitigates the perceived risk arising from uncertainties.The relatively lower presence of negative emotions in the dataset can be attributed to the fact that the reviews were collected exclusively from best-selling products in Amazon India's beauty, electronics, and clothing & accessories categories. Best sellers generally reflect market acceptance and consumer satisfaction, meaning they are more likely to attract favourable feedback.

Further, a sentiment score analysis was conducted on a text file of products.The analysis utilises linear regression models to assess the relationship between sentiment scores (negative and positive) derived from product text and their corresponding product ratings. Understanding these relationships is crucial for identifying how linguistic elements in product descriptions influence customer perception and ultimately, product success.

The statistical output from R Studio, including coefficients, standard errors, t-values, p-values, R-squared, and F-statistics, provides the basis for interpreting the model's performance and the significance of the relationships observed.

```
Call:
lm(formula = x$RATING ~ s$negative + s$positive)

Residuals:
   Min      1Q  Median      3Q     Max
-3.5121 -0.3863  0.2811  0.7395  1.6970

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.26055    0.02583 164.970  < 2e-16 ***
s$negative -0.66742   0.06675  -9.999  < 2e-16 ***
s$positive0.12577    0.03305   3.806  0.000146 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8511 on 1997 degrees of freedom
Multiple R-squared:  0.05542,           Adjusted R-squared:  0.05448
F-statistic: 58.59 on 2 and 1997 DF,  p-value: < 2.2e-16
```

Figure 8 Sentiment Score Analysis

Figure8 displays the results of sentiment score analysis. The coefficient for the negative sentiment score is -0.66742, which is statisticallysignificant ($p < 2e-16$, i.e $p < .001$). This indicates that for each one-unit increase in the negativesentiment score, the product rating is predicted to decrease by approximately 0.67points, holding the positive sentiment score constant. Conversely, the coefficient forthe positive sentiment score is 0.12577, also statistically significant ($p = 0.000146$, i.e $p < .001$). Thissuggests that for each one-unit increase in the positive sentiment score, the productrating is predicted to increase by approximately 0.13 points, holding the negativesentiment score constant. The larger absolute value of the negative coefficientcompared to the positive coefficient suggests that negative sentiment has a strongerimpact on product ratings than positive sentiment.

Asymmetrical Impact of Sentiment: Negative sentiment consistently showed a stronger detrimental effect on product ratings compared to the positive influence of positive sentiment. This asymmetry highlights the critical importance for e-commerce businesses to proactively manage and address negative customer feedback, as it can have a disproportionately larger impact on product perception.

Limited Explanatory Power of Sentiment: The analysis yielded an Adjusted R2 value of .054, indicating that approximately 5.4% of variance in product ratings can be explained by the sentiment scores (positive and negative) extracted from review texts. This suggests that product ratings are a complex outcome influenced by a multitude of factors beyond just thesentiment expressed in reviews, including objective product quality, brand reputation, pricing, and customer service.

High Baseline Rating: The statistically significant intercept of 4.26055 for the combined model indicates a relatively high baseline product rating even without explicit positive or negative sentiment. This may be due to factors such as a general tendency for satisfied customers to leave reviews, Amazon's product curation, or the intrinsic quality of best-selling products. It may also reflect a positive bias in review submission, where only highly satisfied or highly dissatisfied customers are motivated to leave feedback.

These findings highlight the significant, though asymmetrical, influence of sentiment on product ratings, with negative sentiment having a more potent detrimental effect. The qualitative analysis further enriches this understanding by revealing the specific linguistic patterns, thematic discussions, and emotional expressions that characterise online reviews across different product categories on Amazon India.

7. FINDINGS AND DISCUSSION

This study provides a comprehensive analysis of online product reviews in the Indian e-commerce market, combining qualitative text analysis with quantitative regression modelling to understand the impact of sentiment on product ratings. The findings offer several key insights that have significant implications for e-commerce businesses, marketers and researchers.

The qualitative analysis revealed that consumers use a rich and varied lexicon to express their experiences, with distinct patterns emerging across different product categories. The predominance of positive words like 'good', 'quality', and 'nice' suggests a general inclination towards sharing positive feedback. However, the category-specific word usage highlights the importance of understanding the unique value propositions that consumers seek in different product types. While functional attributes are paramount for electronics, sensory and experiential aspects are more critical for beauty and fashion products. The emphasis on 'battery life' in Electronics versus 'fit' in Fashion reflects distinct consumer needs and evaluation criteria, suggesting that businesses should tailor their product development and communication strategies accordingly. This underscores the need for businesses to tailor their product descriptions, marketing messages, and customer service to align with these category-specific expectations.

The thematic analysis through topic modelling further reinforces this point, identifying three core themes: Functional Values, Experiential Values, and Product Description. These topics resonate with established consumer behaviour theories, where purchasing decisions are influenced by both utilitarian (functional) and hedonic (experiential) aspects.

Topic 1 highlights that consumers frequently discuss the functional aspects andperformance of products, particularly in the electronics category, emphasising efficiency, reliability, and utility.Topic 2 (Experiential Values) had words such as ‘fragrance’, ‘fabric’,‘perfume’, ‘smell’, ‘soft’, and ‘cotton’, emphasising the significance of sensory and emotional experiences in consumerreviews, especially for beauty and fashion products.Consumers articulate theirsatisfaction or dissatisfaction based on how products feel, smell, or enhance theirpersonal experience (Lopez &Garza, 2022; Krishna, 2012). The 'Product Description' topicreflects reviews that focus on the physicalcharacteristics, attributes, and specific use-cases of products, often detailing howa product addresses a particular need or problem (Yuan, 2018; Camilleri,2018; Rai, 2012). It emphasises the critical role of accurate and comprehensive product information, as discrepancies can lead to negative sentiment and lower ratings, aligning with information asymmetry theories in e-commerce. Thisthematic structure provides a useful framework for businesses to understand the key drivers of consumer satisfaction and dissatisfaction. By focusing on these themes, businesses can identify areas for product improvement, enhance their marketing communications, and better meet the needs of their target audience. Topic 1 highlights that consumers frequently discuss the functional aspects andperformance of products, particularly in the electronics category, emphasising efficiency, reliability, and utility.

The sentiment analysis, revealing a higher frequency of positive sentiment and emotions like 'trust', 'anticipation', and 'joy', reinforces the notion that online reviews serve as a powerful mechanism for social proof. The prominence of 'trust' is particularly significant in the Indian context, where perceived risk in online transactions can be higher due to factors like payment security and product authenticity concerns. Positive reviews, especially those conveying trust, can significantly mitigate these risks, thereby facilitating purchase decisions. Negative reviews are considered helpful than positive reviews (Lee et al., 2017; Ghasemaghaei et al., 2018). The lower frequency of intense negative emotions like 'disgust' and 'anger' suggests that while negative experiences occur, they might be less frequently or less intensely expressed for best-selling products, or perhaps consumers are more likely to voice extreme dissatisfaction through direct channels rather than public reviews (Malik & Hussain, 2017; Weng & Zhao,2020; Ullah et al., 2016).Since these best-selling products already enjoy strong popularity and market acceptance, the reviews tend to emphasise positive experiences, trust, anticipation, and joy, while strongly negative emotions such as disgust appear far less frequently.

Despite being statistically significant, the sentiment score explained only 5.4% of the variance in product ratings, suggesting that ratings are shaped by other factors also. Collectively, asymmetrical effects with product ratings and significant but low adjusted R2 suggest that sentiment is a meaningful but partial predictor of consumer evaluation (Geetha et al.,2017). Therefore, a holistic approach to product management and customer satisfaction is essential for businesses aiming to achieve high product ratings and sustained success in the competitive e-commerce landscape of India.

Future research could expand on this study by incorporating additional variables such as product price, brand recognition, and specific product features to develop a more comprehensive model for predicting product ratings. Furthermore, exploring the application of more advanced sentiment analysis techniques, such as aspect-based sentiment analysis, could provide granular insights into specific product attributes driving positive or negative feedback. Comparative studies across different emerging markets would also offer valuable cross-cultural perspectives on consumer review behaviour and its impact on e-commerce.

8. IMPLICATIONS

For e-commerce companies and marketers, this study provides actionable findings that can guide their strategic decisions. Notably, evidence of the asymmetrical effect of negative sentiment makes clear the imperative for proactive reputation management. Companies should invest in strong systems for monitoring customer feedback on an ongoing basis and incorporate diligent systems for responding quickly and effectively to negative posts. Focusing on customers' concerns and resolving problems can not only limit the fallout from negative posts but also transform a negative experience for customers into a positive one and build customers' trust and loyalty.

The findings specific to each category emphasise the necessity of a customised approach to marketing and product presentation. Experiential values are paramount for the beauty and fashion category; companies should emphasise rich, experientially rich descriptions of products and user-generated content highlighting the product experience. For the electronics category, functional values are most imperative; thus, emphasis should be placed on communicating comprehensive technical details, performance levels, and clear details on product features. In each case, the customised strategies have the potential to serve the customers better and boost satisfaction.

The thematic analysis provides an informative structure for ideas on product development and improvement. By elucidating the prominent themes that appear and reappear in customers' reviews (Functional Values, Experiential Values, and Product Description), companies canidentify areas in which their products are doing well and areas for improvement. This data-driven strategy for developing products has the potential for greater product success and greater alignment with customers' needs.Methodologically, the combination of sentiment analysis, topic modelling, and thematic analysis establishes the wealth of mixed methods for fine-grained extraction of insights from unstructured text. Notably, the weak explanatory power of sentiment for review ratings indicates the necessity for broader encompassing conceptual models considering multiple drivers of consumer judgment. This creates rich possibilities for further understandingof the complex dynamics determining review influence on the Internet.

The e-commerce platforms, such as Amazon India, must consider the native language of the place (e.g Hindi) as it was observed that customers express their emotions, sentiments and opinions in their native language, which cannot be ignored. Overlooking the reviews in the native language may lead to sampling bias, as meaningful reviews in Hindi may be excluded from analysis. Even the translations may not preserve the sentiment, which dilutes the whole purpose of semantic analysis. Thus, the marketers and practitioners may devise some algorithms to accommodate the use of the native language as well.

8.1. THEORETICAL IMPLICATIONS

This paper contributesto the literature on online review content, affective use of text, and e-commerce. Firstly, it provides acomprehensive analysis of a large dataset from the Indian e-commerce domain, which has remained relatively less explored so far as compared with more mature markets. The findings on the specific linguistic features, thematic compositions, and affective uses of Indian customers complement our understanding of cross-cultural differences in online review content.Secondly, the paper provides empirical support for the hypothesis of negativity bias in writingof online reviews and demonstrates that negative content has a stronger influence on judgment as compared with positive content. The asymmetrical influence of sentiment on ratings of products provides further empirical support for this hypothesis and its prominence in the digital era.Thirdly, the paper provides inputs toward the literature on qualitative analysis of text by establishing the effectiveness ofseveral methods, including content analysis, topic modelling, and sentiment analysis, for building an exhaustive understanding of unstructured text data. The methodsadopted in this paper can serve as groundwork for future research.

Sentiment scores alone explain only a small portion of thevariance in product ratings, suggesting that product ratings are a complex outcomeinfluenced by multiple factors beyond expressed sentiment. The findings highlight thecritical importance for e-commerce businesses to proactively manage and addressnegative customer feedback and adopt a holistic approach to product managementand customer satisfaction. Sentiment analysis should be supplemented with further data sources, such as sales data, customer service interactions, and market research, so as to obtain a general view of the customer experience. Through the integral use of data, organisationscan make wiser decisions and reap long-term benefits.

In conclusion, online reviews are a powerful source of consumer insights. Qualitative content analysis, topic modelling and sentimentanalysis provide valuable tools for understanding their impact. Whilenegative sentiment demands special attention, a comprehensive strategy thatconsiders multiple factors influencing product perception is essential for businesses tothrive in the dynamic and competitive e-commerce environment.Strategically, e-commerce businesses should adopt a dual approach—proactively managing negative feedback while amplifying positive narratives—to enhance customer satisfaction and trust. Academically, the research contributes by evidencing the interplay of emotional, semantic, and many other factors in review analysis, offering pathways for future studies to incorporate richer variables and advanced text mining techniques.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Abighail, B. M. D., Firmanda, M. R., &Anggreainy, M. S. (2023). Sentiment Analysis E-commerce Review. Procedia Computer Science, 227, 1039-1045.

Alzate, M., Arce-Urriza, M., &Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. Journal of Retailing and Consumer Services, 67, 102989.

Bell, E., Harley, B., & Bryman, A. (2022). Business research methods. Oxford university press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Bizrate Insights. (2024). Consistency rules: Online shoppers are creatures of habit. https://bizrateinsights.com/consistency-rules-online-shoppers-are-creatures-of-habit/

BrightLocal. (2024). Google reviews usage statistics: Local Consumer Review Survey.BrightLocal.https://www.brightlocal.com/research/local-consumer-review-survey-2024/

BrightLocal. (2023). Local Consumer Review Survey 2023. BrightLocal.https://www.brightlocal.com/research/local-consumer-review-survey-2023/

Camilleri, A. R. (2020). The importance of online reviews depends on when they are presented. Decision Support Systems, 133, 113307.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3), 477-491.

Fan, Z. P., Che, Y. J., & Chen, Z. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. Journal of Business Research, 74, 90-100.

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43-54.

Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. Internet Research, 28(3), 544-563.

Ghose, A. and Ipeirotis, P.G. (2011), "Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics", IEEE Transactions on Knowledge and Data Engineering, Vol. 23 No. 10, pp. 1498-1512.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. Journal of statistical software, 40, 1-30.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. Journal of interactive marketing, 18(1), 38-52.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative health research, 15(9), 1277-1288.

Jaichandran, R., Bagath Basha, C., Shunmuganathan, K. L., Rajaprakash, S., &Kanagasuba Raja, S. (2019). Sentiment analysis of movies on social media using R studio. International Journal of Engineering and Advanced Technology, 8(6), 2171-2175.

Kim, H., Cho, I., & Park, M. (2022). Analyzing genderless fashion trends of consumers' perceptions on social media: using unstructured big data analysis through Latent Dirichlet Allocation-based topic modeling. Fashion and Textiles, 9(1), 6.

Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. Journal of consumer research, 18(2), 243-250.

Krippendorff, K. (2018). Content analysis: An introduction to its methodology. Sage publications.

Krishna, A. (2012), "An integrative review of sensory marketing: engaging the senses to affect perception, judgment and behavior", Journal of Consumer Psychology, Vol. 22 No. 3, doi: 10.1016/j.jcps.2011.08.003.

Lee, J., Park, D. H., & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: An advertising perspective. Internet research, 21(2), 187-206.

Lee, M., Jeong, M., & Lee, J. (2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. International Journal of Contemporary Hospitality Management, 29(2), 762-783.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. Information & Management, 56(2), 172-184.

Lopez, A., & Garza, R. (2022). Do sensory reviews make more sense? The mediation of objective perception in online review helpfulness. Journal of Research in Interactive Marketing, 16(3), 438-456.

Malik, M. S. I., & Hussain, A. (2017). Helpfulness of product reviews as a function of discrete positive and negative emotions. Computers in Human Behavior, 73, 290-302.

National Retail Federation. (2020). 2020 retail returns survey. National Retail Federation. https://nrf.com/media-center/press-releases/428-billion-merchandise-returned-2020

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J., 2007. The Development and Psychometric Properties of LIWC2007. Austin, Tx LIWC.Net 1. https://doi.org/10.1177/026377588300100203.

Rai, R. (2012, August). Identifying key product attributes and their importance levels from online customer reviews. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 45028, pp. 533-540). American Society of Mechanical Engineers.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems, 81, 30-40.106.

Shiprocket. (2020, December 19). Most demanded selling products online in India. Shiprocket. Authored by Puneet Bhalla. Retrieved from https://360.shiprocket.in/blog/most-demanded-selling-products-online-india/

Statista. (2022). Monthly visits on leading marketplace platforms in India. Retrieved from https://www.statista.com/statistics/1239038/india-monthly-visits-on-leading-marketplace-platforms/

Statista.(2023). Number of digital buyers in India. Statista.https://www.statista.com/statistics/251631/number-of-digital-buyers-in-india/

Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. Decision Support Systems, 81, 41-53.

Ullal, M. S., Spulbar, C., Hawaldar, I. T., Popescu, V., &Birau, R. (2021). The impact of online reviews on e-commerce sales in India: A case study. Economic Research-EkonomskaIstraživanja, 34(1), 2408-2422.

Vespestad, M. K., & Clancy, A. (2021). Exploring the use of content analysis methodology in consumer research. Journal of Retailing and Consumer Services, 59, 102427.

Wahpiyudin, C. A. B., Mahanani, R. K., Rahayu, I. L., Simanjuntak, M., Sumarwan, U., Yuliati, L. N., ... &Muflikhati, I. (2022). The credibility of consumer reviews on three e-commerce in Indonesia: Mixed method approach. JurnalIlmuKeluarga dan Konsumen, 15(3), 287-299.

Wąsowicz-Zaborek, E. (2023). Content analysis of hotel reviews as a quality management tool: Preliminary verification of the SERVQUAL scale. Turyzm, 33(1), 29-40.

Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. Communication methods and measures, 11(4), 245-265.

Weng, D., & Zhao, J. (2020). Positive emotions help rank negative reviews in e-commerce. arXiv preprint arXiv:2005.09837.

Westerlund, M., Mahmood, Z., Leminen, S., &Rajahonka, M. (2019). Topic modelling analysis of online reviews: Indian restaurants at Amazon. com. In ISPIM Conference Proceedings (pp. 1-14). The International Society for Professional Innovation Management (ISPIM).

Wu, J., Du, L., & Dang, Y. (2018, July). Research on the Impact of Consumer Review Sentiments from Different Websites on Product Sales. In 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C) (pp. 332-338). IEEE.

Yuan, H., Xu, W., Li, Q., & Lau, R. (2018). Topic sentiment mining for sales performance prediction in e-commerce. Annals of Operations Research, 270(1), 553-576.