



International Journal of Research in Management

ISSN Print: 2664-8792
ISSN Online: 2664-8806
Impact Factor: RJIF 8
IJRM 2025; 7(1): 780-784
www.managementpaper.net
Received: 15-02-2025
Accepted: 19-03-2025

Dr. Ravi Sidhu
Principal, Dips Institute of
Management and Technology,
Jalandhar, Punjab, India

Sentiment analysis on customer reviews of e-commerce site

Ravi Sidhu

DOI: <https://www.doi.org/10.33545/26648792.2025.v7.i1i.345>

Abstract

Sentiment analysis has emerged as one of the fastest-growing areas in computer science, driven by the increasing importance of understanding consumer opinions in the digital era. As advancements in Natural Language Processing (NLP) and machine learning continue to evolve, sentiment analysis plays a pivotal role in deciphering emotions, opinions, and attitudes expressed through textual data. This research leverages a dataset of product reviews collected from Flipkart.com, one of the largest e-commerce platforms. By analyzing the ratings provided by customers, the study performs a comprehensive sentiment analysis to categorize reviews as positive, negative, or neutral, offering insights into the general perception of various products.

The methodology includes preprocessing the dataset to clean and normalize the text, followed by sentiment classification using state-of-the-art machine learning models. Additionally, the research identifies the most frequently used words and word pairs (bigrams) to reveal significant patterns in customer feedback. A word cloud visualization is used to highlight the most prominent terms, while count plots are employed to examine the distribution of sentiments across different product categories. The study also explores the relationship between sentiment polarity and product ratings to understand how specific words correlate with customer satisfaction.

The findings suggest that identifying the most critical negative and positive words within the dataset can offer valuable insights into the psychological state of consumers. This understanding can not only help businesses enhance their customer service and marketing strategies but also provide a deeper look at consumer behavior and decision-making processes. Moreover, the study emphasizes the potential of sentiment analysis in the e-commerce domain, where understanding customer sentiment can significantly influence product development, sales strategies, and brand loyalty.

Keywords: Countplot, sentiment analysis, product reviews, word cloud, e-commerce, machine learning, natural language processing, customer feedback, consumer behavior, text classification

1. Introduction

Sentiment analysis, a subfield of text analysis, Natural Language Processing (NLP), and computational linguistics, focuses on the scientific identification, extraction, and study of subjective information from textual data ^[1, 2]. The rapid growth of e-commerce has significantly transformed consumer behavior, with online reviews replacing traditional "word of mouth." In the digital age, reviews play a pivotal role in shaping purchasing decisions and influencing the success of products in the marketplace. These reviews act as trust-building platforms, allowing consumers to assess the experiences of previous buyers, ultimately empowering them to make informed decisions about the products they intend to purchase ^[3].

Given the vast number of products manufactured by numerous brands, providing relevant and insightful reviews to consumers has become an essential aspect of the modern shopping experience ^[4]. In the past, individuals would rely primarily on their immediate social circle family and friends for opinions when considering a product purchase. However, in today's connected world, consumers have access to a multitude of user-generated reviews and discussions across public forums, social media platforms, and dedicated review sites on the web ^[5, 6]. This democratization of information allows potential buyers to evaluate a product based on the shared experiences of a much wider community.

For businesses and product owners, gathering relevant online reviews has become an invaluable tool for meeting the needs of their target market. By analyzing these reviews, businesses can adapt their products and services to better serve customers' preferences, improving customer satisfaction and loyalty.

Corresponding Author:
Dr. Ravi Sidhu
Principal, Dips Institute of
Management and Technology,
Jalandhar, Punjab, India

From a managerial perspective, insightful online reviews also offer critical intelligence regarding market competition and the prevailing trends that influence marketing strategies. More importantly, they help identify customer pain points and desires, providing actionable feedback to enhance existing products or guide the development of new offerings. As a result, the role of sentiment analysis in interpreting online reviews has become indispensable in guiding both product development and marketing decisions in today's competitive business landscape.

2. Literature Review

^[8] One of the foundational efforts in sentiment analysis involves classifying products into categories such as positive, negative, and neutral to represent the sentiment expressed in product reviews. Negative sentiment products can be particularly useful in fields such as cybercrime investigation, where they can serve as evidence. This study employs the content analysis method for product classification, which involves feature extraction followed by a classification step. However, many traditional applications of opinion mining rely on the "bag-of-words" model, which does not account for context an essential factor for accurate sentiment analysis. The paper also discusses recent advancements in sentiment analysis and its related subfields, such as feature extraction and sentiment classification using various machine learning techniques.

^[9] The emotional analysis of various operations, such as quality control, market research, competitor analysis, product reviews, and customer feedback, is integral to sentiment analysis. Sentiment analysis, often paired with natural language processing (NLP), faces several challenges due to the complex nature of language. Many words in different languages have meanings that shift depending on the context in which they are used. As a result, tools for sentiment analysis are not universally applicable across all dialects. Recent studies have highlighted specific issues in domains like sports and politics, where sentiment analysis plays a critical role. This study provides an in-depth analysis of sentiment analysis by exploring its various aspects, including operations, tools, and methodologies.

^[10] This paper presents basic data extraction techniques, followed by data cleaning and preprocessing methods that are crucial in sentiment analysis. Specifically, it discusses three common preprocessing techniques tokenization, stop word removal, and stemming applied to a Twitter dataset. The results indicate that preprocessing significantly reduces the dimensionality of data, leading to higher-performing and more accurate sentiment analysis techniques. In particular, the stop word removal technique is shown to enhance accuracy by eliminating irrelevant words from the dataset. The paper suggests that these preprocessing techniques can be applied to other datasets in different domains, with the stop word list customized according to the specific domain for better results.

^[11] In this research, unstructured customer data is filtered to remove noisy content and preprocessed to assess sentiment. The proposed methodology aims to help future buyers make better purchasing decisions by analyzing customer feedback about specific brands. This work emphasizes the importance of effectively handling unstructured data and transforming it into actionable insights for both businesses and consumers.

^[12] Sentiment analysis encompasses various steps, including data collection, data preprocessing, feature extraction,

sentiment classification, and evaluation. This paper reviews the sentiment analysis process and its application to reviews of mobile applications available on platforms like the Google Play Store, Amazon, Coursera, and Twitter. Machine learning algorithms, including Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest, were employed to classify user reviews into positive, negative, or neutral categories. The performance of these algorithms was evaluated based on precision, recall, accuracy, and F1 score. The study highlights the significance of sentiment analysis in identifying customer concerns and improving service quality for businesses.

^[13] This study addresses both binary sentiment classification (positive vs. negative) and multi-class sentiment analysis (positive, negative, and neutral) for the English language. Using three different datasets for evaluation, the authors propose a novel two-stage classification approach. The first stage involves emotion detection, and based on the identified emotion, the second stage classifies the sentiment. A zero-shot transformer model is used for the emotion detection stage, which does not require prior training and can extract emotion probabilities from the text. This method offers an innovative approach to sentiment analysis, distinguishing it from traditional sentiment classification techniques.

^[14] Machine learning-based sentiment analysis techniques are applied to reviews from various domains, including products, movies, and restaurants, sourced from platforms like Amazon, IMDb, and Yelp. One of the key challenges in sentiment analysis is detecting emotions from informal reviews that contain slang, hashtags, and other non-standard language. To address this issue, the authors introduce the concept of feature vectors. Preprocessing is first performed on each review to clean the text, followed by two phases of feature extraction: the first phase extracts review-specific words, while the second phase removes them from the text. The extracted feature vector is then converted back to normal text for analysis. This approach improves the effectiveness of sentiment analysis by capturing the most relevant features of the review.

3. Data set and its characteristics

The data set used in this study comprises detailed information collected from Flipkart.com, one of the leading e-commerce platforms. It includes various attributes related to the products, such as product names, prices, ratings, reviews, and summaries. The dataset spans across a wide range of product categories, with a total of 104 distinct product types. These categories include electronics, clothing for men, women, and children, home décor items, automated systems, and more.

The dataset consists of 205,053 rows and five columns, each representing different product-related features. The five primary features are outlined in Table 1 below

Table 1: Features included in the data set

Feature	Description
Product Name	Name of the Product
Product Price	Price of the Product
Rating	UserRatingbetween 1to5
Review	User reviews are provided for every Product

4. Methodology

Customer feedback can generally be categorized into three main sentiment classes: positive, negative, or neutral [6], [15]. To analyze sentiment, we first calculate two key metrics for each review:

Subjectivity and Polarity

- **Subjectivity:** This value ranges from 0 to 1, where 0 indicates objective statements (factual) and 1 indicates subjective statements (opinion-based). Subjectivity helps determine how much of the review expresses personal feelings or opinions as opposed to factual content.
- **Polarity:** This value ranges from -1 to +1, with -1 indicating a strongly negative sentiment, +1 indicating a strongly positive sentiment, and 0 indicating neutral

sentiment.

The overall sentiment for each review is computed as the summation of the product of the subjectivity and polarity values. This combined measure captures both the intensity and the subjective nature of the sentiment expressed in the review. Specifically, the sentiment value is calculated by multiplying the subjectivity and polarity scores for each individual review, and then summing these values across all reviews in the dataset. This approach allows for a nuanced understanding of the sentiment present in the entire set of customer feedback.

Table 2 provides a detailed snapshot of the sentiment calculation process, illustrating how subjectivity and polarity values are determined for each review and how the overall sentiment score is derived.

Table 2: Sentiment Calculation

Review	Subjectivity (s)	Polarity (p)	Sentiment Score (s*p)
Great cooler and excellent airflow and for this price it is so amazing and unbelievable just love it	0.812500	0.725000	0.5890625
Bestbudget2fitcoolernicecooling	0.566667	0.666667	0.37777819
The quality is good but the power of the air is decent	0.633333	0.433333	0.27444409
Very bad product it's only a fan	0.933333	-0.455000	-0.42466652
Ok ok product	0.500000	0.500000	0.25

The sentiment of a review is determined based on the polarity value. If the polarity is greater than 0, the text is classified as positive. If the polarity is less than 0, the text is classified as negative. When the polarity equals 0, the review is considered neutral, indicating neither a strongly positive nor a strongly negative sentiment.

Additionally, the subjectivity score ranges from 0.0 to 1.0. A higher subjectivity score signifies that the text is more opinionated or subjective, while a lower score suggests that the text is more objective or factual.

To visualize the distribution of sentiments across the dataset, Figure 1 provides a graphical representation of the sentiment analysis results. It offers an overview of the overall sentiment distribution among the reviews. Furthermore, Figure 2 presents a Count Plot of the sentiments, categorizing them into positive, negative, and neutral classes to display the frequency of each sentiment type across the dataset.

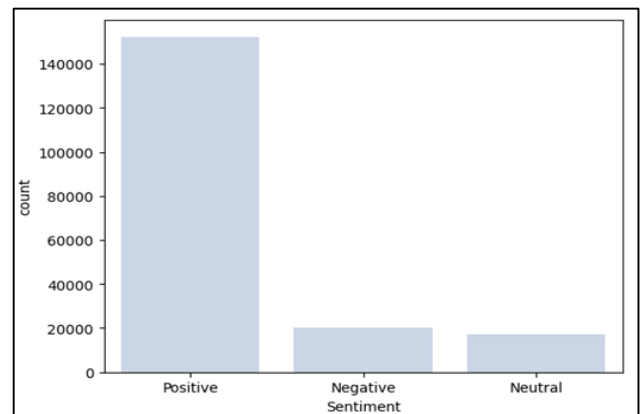


Fig 2: Count Plot of the sentiments

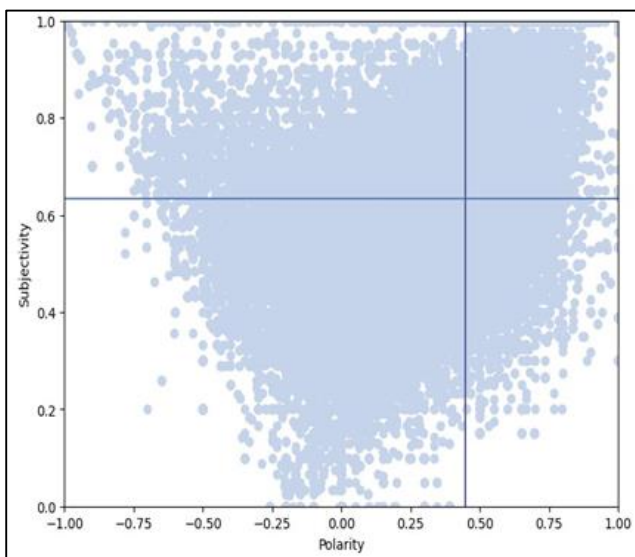


Fig 1: Visualization

5. Results

The sentiment analysis of customer reviews in the e-commerce dataset reveals the most frequently used terms associated with both positive and negative sentiments. These terms provide valuable insights into consumer opinions and preferences, helping to distinguish between favorable and unfavorable evaluations of products.

To effectively separate positive and negative reviews, the dataset was sorted according to the sentiment values, categorizing them based on the calculated sentiment scores. This allowed us to identify key terms that drive either positive or negative sentiment in the reviews.

- **Positive Word Cloud (Figure 3):** This word cloud visualizes the most frequently occurring words associated with positive sentiments in the dataset. Words like "excellent," "great," "good," and "happy" appear prominently, indicating the aspects of the products that customers tend to appreciate the most.
- **Negative Word Cloud (Figure 4):** In contrast, Figure 4 showcases the word cloud for negative sentiments, highlighting terms such as "poor," "bad," "disappointing," and "unreliable". These words reflect

common concerns and issues voiced by customers who were dissatisfied with the product or service.

- **Most Frequently Used Terms (Figure 5):** Figure 5 presents a detailed list of the top terms that appeared most often across the dataset, regardless of sentiment polarity. This list offers a comprehensive view of the key product features, attributes, and customer feedback trends that are most discussed in the reviews.

The visualizations and word clouds provide an intuitive way to understand the language used by consumers, showcasing the elements that significantly influence their overall sentiment toward products. This analysis can serve as a guide for businesses to improve product offerings and customer satisfaction based on recurring themes in the reviews.



Fig 3: Positive word Cloud



Fig 4: Negative Word cloud

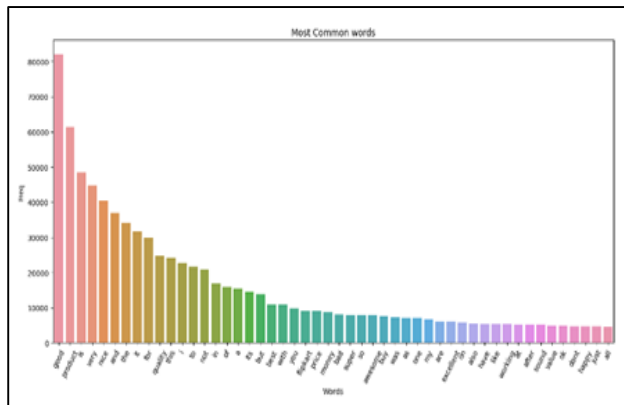


Fig 5: Most Common Words

6. Conclusions

This research provides valuable insights into how various attributes of e-commerce product reviews relate to one another and how sentiment analysis can be effectively applied to understand consumer perceptions. By analyzing a large dataset of product reviews, this study offers a deeper understanding of customer opinions, highlighting the factors that influence their satisfaction or dissatisfaction with the products they purchase.

The results of sentiment analysis can serve as a powerful tool for both consumers and product designers. For consumers, sentiment analysis enables them to make informed purchasing decisions by evaluating the experiences of previous buyers. By gaining an overview of the general sentiment surrounding a product, consumers can better assess its quality and suitability to their needs.

From the perspective of product designers and businesses, sentiment analysis provides crucial feedback that can directly influence product development. By calculating the overall sentiment of reviews, designers can identify the strengths and weaknesses of their products, allowing them to focus on areas that require improvement. Positive feedback can highlight features that resonate well with customers, while negative sentiment can shed light on recurring issues or unmet needs. This feedback loop is essential for designing products that better align with customer expectations and improve overall satisfaction. Moreover, sentiment analysis can help businesses identify emerging trends and customer preferences, enabling them to introduce new products or refine existing ones to better meet market demands. By leveraging the insights gained from this research, designers and businesses can create products that not only fulfill functional requirements but also resonate emotionally with consumers, enhancing brand loyalty and driving success in a competitive market.

7. References

1. Jagdale RS, Shirsat VS, Deshmukh SN. Sentiment analysis on product reviews using machine learning techniques. In: *Advances in Intelligent Systems and Computing*. Springer Verlag, 2019, p. 639-47. DOI: 10.1007/978-981-13-0617-4_61.
2. Ranjan S, Singh I, Dua S, Sood S. Sentiment analysis of stock blog network communities for prediction of stock price trends. *Indian Journal of Finance*. 2018;12(12). DOI: 10.17010/ijf/2018/v12i12/139888.
3. Singla Z, Randhawa S, Jain S. Statistical and sentiment analysis of consumer product reviews. In: *Proceedings of the 8th International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2017, p. 1-6. DOI: 10.1109/ICCCNT.2017.8203960.
4. Singla Z, Randhawa S, Jain S. Sentiment analysis of customer product reviews using machine learning. In: *Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2)*, 2018, p. 1-6. DOI: 10.1109/I2C2.2017.8321910.
5. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018. DOI: 10.1002/widm.1253.
6. Kaur M, Verma R, Ranjan S. Political leaders' communication: A Twitter sentiment analysis during

- the COVID-19 pandemic. *Jurnal The Messenger*. 2021;13(1). DOI: 10.26623/themessenger.v13i1.2585.
7. Jin J, Jia D, Chen K. Mining online reviews with a Kansei-integrated Kano model for innovative product design. *International Journal of Production Research*, 2022. DOI: 10.1080/00207543.2021.1949641.
 8. Wase K, *et al.* Sentiment analysis of product reviews. *International Journal of Innovations in Engineering and Science*. 2018;3(5).
 9. Shah D, Singh A, Prasad SS. Sentimental analysis using supervised learning algorithms. In: *Proceedings of the 2022 3rd International Conference on Computation, Automation, and Knowledge Management (ICCAKM)*. IEEE, 2022, p. 1-5. DOI: 10.1109/ICCAKM54721.2022.9990320.
 10. Garg N, Sharma K. Text pre-processing for multilingual sentiment analysis based on social network data. *International Journal of Electrical and Computer Engineering*. 2022;12(1):776-784. DOI: 10.11591/ijece.v12i1.pp776-784.
 11. Biradar SH, Gorabal JV, Gupta G. Machine learning tool for exploring sentiment analysis on Twitter data. *Materials Today: Proceedings*, 2022. DOI: 10.1016/j.matpr.2021.11.199.
 12. Sankhe J, Batavia K, Borse H, Sharma S. Survey on sentiment analysis. *International Research Journal of Engineering and Technology*. 2022.
 13. Tesfagergish SG, Kapočiūtė-Dzikienė J, Damaševičius R. Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. *Applied Sciences (Switzerland)*. 2022;12(17). DOI: 10.3390/app12178662.
 14. Ghosh S, Hazra A, Raj A. A comparative study of different classification techniques for sentiment analysis. *International Journal of Synthetic Emotions*. 2020;11(1):49-57. DOI: 10.4018/ijse.20200101.oa.
 15. Kaur R, Ranjan S. Sentiment analysis of 21 days COVID-19 Indian lockdown tweets. *International Journal of Advance Research in Science and Engineering*, 2020.