International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 5,Sep-Oct 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Sentiment Analysis of Product Reviews and Ratings on *Daraz*: A Study of Customer Perceptions in the Nepalese E-commerce Market

Bibek Bhandari*, Aashish Panthi**

*(Department of STEM, Himalayan White House International Secondary School, Kathmandu
Email: bhandareebivek@gmail.com)

** (Department of Computer Engineering, Kalika Manavgyan Secondary School, Butwal
Email :aashishpanthi11@gmail.com)

----------**********----------

Abstract:

This research paper presents a comprehensive analysis of customer perceptions within the Nepalese e-commerce market, focusing on the popular platform, *Daraz*. The study employs sentiment analysis, a computational linguistic approach of products over 10000, to analyze into the sentiments and opinions expressed in product reviews and ratings. Through the examination of a vast dataset, this study aims to uncover insights into the preferences, satisfaction levels, and concerns of consumers in Nepal's evolving e-commerce landscape. The findings shed light on the distinctive dynamics of the Nepalese e-commerce market and contribute to the broader understanding of consumer behavior in an increasingly digital age. This research not only advances the field of sentiment analysis but also offers valuable implications for businesses operating within this unique market, helping them adapt and cater to the preferences of Nepalese consumers.

Keywords **—Sentiment Analysis, Product Reviews, Ratings, E-commerce,Market Analysis, Computational Linguistics.**

----------**********----------

I. INTRODUCTION

In the digital age, e-commerce platforms have become integral channels for consumers to explore, evaluate, and acquire products. Within this dynamic landscape, customer feedback in the form of product reviews plays a pivotal role in shaping purchasing decisions. The opinions, sentiments, and evaluations expressed by consumers not only influence individual buying choices but also hold the power to sway broader market trends. Understanding and interpreting this wealth of information is paramount for businesses striving to thrive in the competitive e-commerce environment.

Sentiment analysis, a field within computational linguistics, provides the means to systematically extract and quantify subjective information from text data. It involves employing natural language processing techniques to discern the emotional tone, opinions, and attitudes expressed by individuals. By applying sentiment analysis to product reviews, we can uncover underlying sentiment patterns, providing a nuanced understanding of consumer preferences and perceptions.

This study embarks on a comprehensive exploration of consumer sentiments within the context of the Nepalese e-commerce market, focusing on the prominent platform, Daraz. Through the lens of sentiment analysis, a computational linguistic approach, we examine over 10,000 product reviews. This extensive dataset encompasses an array of product categories,

providing a diverse cross-section of consumer preferences and perceptions. The aim of this research is to unravel the intricate relationship between customer sentiments and product ratings, ultimately offering invaluable insights into the factors that influence consumer behaviour and preferences.

In the ensuing sections, we detail the methodology employed for data collection, pre-processing, and sentiment analysis. Subsequently, we present a thorough analysis of the results, which encompasses category-specific sentiment trends, word cloud visualizations, and more. Through this multidimensional exploration, we endeavour to shed light on the nuances of consumer sentiment, providing a foundation for data-driven strategies aimed at enhancing product development, marketing initiatives, and overall customer satisfaction on the *Daraz* platform.

II. METHODOLOGY

This section provides an overview of the proposed methodology of sentiment analysis for Daraz.com.np product reviews. Figure.1 depicts the phases of the current work starting with the data collection until results.

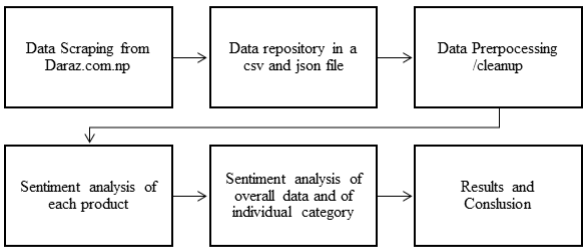

Fig. 1Research Flow Diagram

A. Data collection and pre-processing

Using the JavaScript framework called puppeteer and chromium browser, the reviews were scraped. Initially, a product link was given to the bot to scrape the data from. The link was opened in the chromium browser. Then, the bot scrolled down to the bottom of the page to load all the reviews. The bot scraped the reviews of that product and picked one of the random products from the recommendation section (eliminating the one already scraped) and the bot visited that product. This chain of continuously scraping the reviews and navigating to the next one continued.

The scraped review along with other details of the product was stored on both CSV and JSON files. The CSV file was used for human review and JSON file was used for the further analysis.
The duplicate items were removed from the data repository. The repository was closely observed to know if there are enough products in each category for the generalization of the data. Data with not enough information were removed programmatically.

B. Sentiment Analysis

The sentiment of the reviews was analyzed using sentiment node library. The sentiment of each review was analyzed and the average sentiment of all the reviews was calculated. The average sentiment was used to determine the sentiment of each product. The sentiment and the average score are related in the following ways:

- Positive: The average sentiment greater than 0.
- Neutral: The average sentiment equals to 0.
- Negative: The average sentiment less than 0.

The overall sentiment of the data was calculated in percentage.

To calculate how much percentage of reviews are positive, negative or neutral, a simple algebraic calculation was done. (The number of positive reviews/the total number of reviews) * 100 gave the percentage of positive reviews.

Similarly, the percentage of neutral and negative reviews were collected. Then the result was

presented in a pie chart using the JavaScript libraries: chart js and the reactjs.

Using the same equation, the sentiment of each category was done and presented in a bar graph using the chart js library and the react library.

III. ANALYSIS AND RESULTS

After processing and cleaning up 23,587 scrapped reviews from 10,938 different products we categorized in a table with number of reviews.We found 6,761 products were reviewed and 4,177 were non reviewed, meaning 38.19% of the products did not have any review. With total of 23,587 reviews, each product had average of 2.16 reviews as shown in Figure 2.

Category	No. of Products	No. of Reviewed Products	No. of Non reviewed Products	No. of Reviews
Computers & Laptops	440	263	177	1081
Home Appliances	490	302	188	1081
Sports & Outdoors	548	329	219	1061
Kitchen & Dining	281	178	103	541
Health & Beauty	1943	1128	815	3986
Toys & Games	248	149	99	439
Furniture & Decor	375	250	125	1035
Watches Sunglasses Jewellery	777	507	270	1783
TV, Audio / Video, Gaming & Wearables	560	352	208	1430
Cameras	231	121	110	413
Fashion	1470	947	523	3139
Stationery & Craft	264	167	97	581
Mother & Baby	348	214	134	625
Mobiles & Tablets	688	417	271	1337
Motors	160	98	62	402
Media, Music & Books	318	184	134	635
Tools, DIY & Outdoor	612	379	233	1309
Laundry & Cleaning	142	84	58	231
Groceries	692	440	252	1639
Bedding & Bath	102	72	30	248
Pet Supplies	51	34	17	110
Bags and Travel	196	145	51	475
Digital Goods	2	1	1	6

Fig. 2Categorization of products after data processing/clean-up

A. Overall Sentiment and ratings analysis

The analysis of overall sentiment in the product reviews revealed significant insights. Among the collected reviews, 52.505% exhibited a positive sentiment, indicated by a distinct green coloration. Conversely, 4.224% conveyed a negative sentiment, visually represented in red. A substantial portion, 47.271%, of reviews fell within the neutral sentiment category. These findings are visually summarized in Figure 3, a pie chart illustrating the distribution of sentiments.

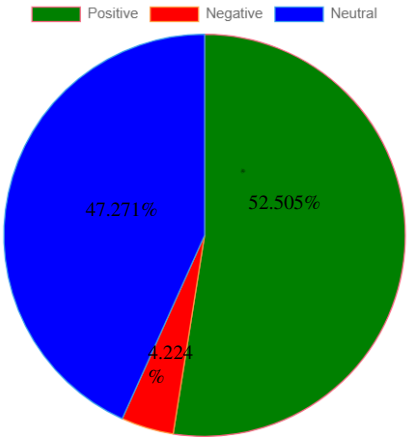

Fig. 3Overall Sentiment of all the reviews

In tandem with sentiment analysis, the distribution of overall product ratings was investigated. Figure 4 presents this information in the form of a pie chart, showcasing that a majority of products (77.81%) garnered ratings between 4 and 5. A smaller fraction (19.56%) fell within the 3 to 4 rating range. Products receiving ratings between 1 and 2 constituted a mere 1.48%, while those with ratings between 2 and 3 were even fewer at 1.15%.

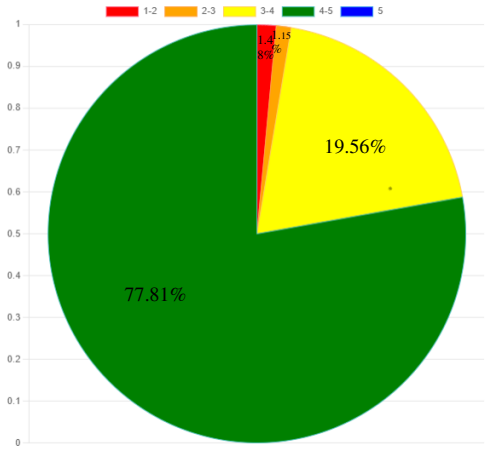

Fig. 4Overall ratings of the Products

To further contextualize the reviews, word clouds were generated for distinct sentiment categories. Figure 5 encapsulates the entirety of terms

employed in the reviews, offering a visual representation of frequently occurring keywords. Figures 6 and 7, on the other hand, focus specifically on positive and negative sentiments, respectively. In these visualizations, keyword size correlates with frequency of usage, providing valuable insights into consumer emphasis.

Fig. 5Overall World Cloud of all the reviews

Fig. 6World Cloud of positive reviews

Fig. 7 World Cloud of positive reviews

B. Sentiment Analysis and Ratings across categories

The sentiment analysis extended to various product categories, shedding light on consumer preferences within specific domains. Figure 8 illustrates sentiment distribution across different categories. Notably, health and beauty products emerged as the category with the highest number of positive sentiments, spanning over 950 products. Additionally, this category exhibited a notable prevalence of neutral sentiments, encompassing over 900 products. Negative sentiments were relatively lower, representing just over 70 products

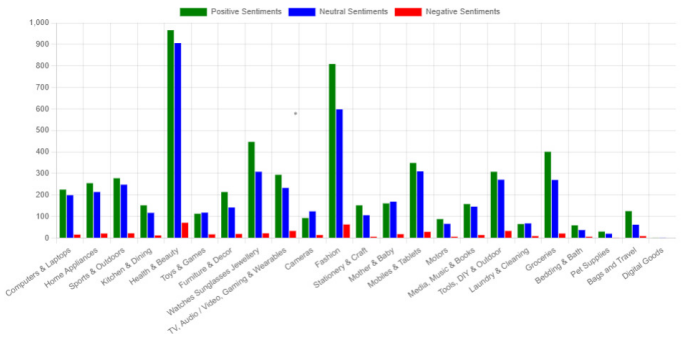

Fig. 8Sentiment Distribution by Categories

Average ratings per category were alsoscrutinized, as depicted in Figure 9. Stationery and craft products garnered the highest average rating of 4.325, followed closely by pet supplies with an average rating of 4.343. In contrast, digital goods garnered the lowest average rating, standing at 2.45. This was closely followed by TV/audio/video, gaming, and wearables, each with an average rating of 3.296.

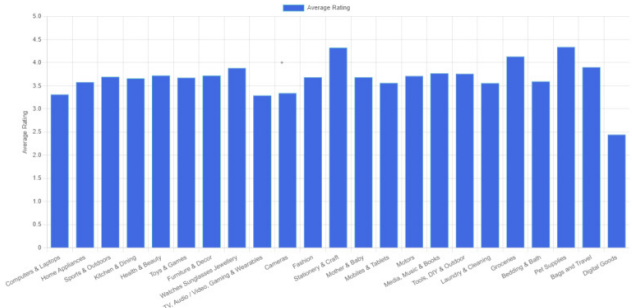

Fig. 9Average rating per Product Category

IV. CONCLUSIONS

This study delved into the sentiments and ratings of customers within the Nepalese e-commerce

domain, focusing on the Daraz platform. Through sentiment analysis, we gleaned valuable insights into consumer perceptions. Over half of the reviews expressed positive sentiments, reflecting a substantial level of customer satisfaction. Neutral sentiments also featured prominently, indicating a diverse range of consumer experiences.

Product ratings corroborated this positive outlook, with a significant proportion of products receiving high scores. Word cloud visualizations provided further depth, revealing the most resonant aspects for consumers. Additionally, sentiment trends across various categories unveiled areas of strength, particularly in health and beauty products.

The disparities in average ratings by category highlight opportunities for refinement. Stationery and craft products, along with pet supplies, garnered the highest ratings, suggesting areas of excellence. Conversely, electronic categories indicated room for improvement.

In conclusion, this research equips businesses with actionable insights to enhance product development and marketing strategies on the *Daraz* platform. By understanding and addressing consumer sentiments, businesses can bolster customer satisfaction and overall e-commerce experiences in Nepal.

LIMITATIONS & SUGGESTIONS FOR FUTURE WORK

While this study provides valuable insights, it is important to acknowledge certain limitations. Firstly, the use of English letters to represent Nepalese words in reviews posed a challenge to accurate sentiment analysis. Additionally, including unreviewed products in the neutral sentiment category may have skewed the overall sentiment distribution.

Furthermore, the data was collected in September 2023, which may not fully reflect current sentiments and trends. Additionally, products with zero ratings or reviews could not be analysed, potentially excluding valuable information. Lastly, the study's dataset comprised 10,938 products, which, while extensive, may not fully represent the platform's entire product range.

For future work, refining methods to handle transliterations, differentiating between genuinely neutral sentiments and those from unreviewed products, and exploring ways to include products with zero ratings or reviews could enhance the accuracy of sentiment analysis. Additionally, ensuring timeliness in data collection and considering methods for dataset expansion could improve generalization. Finally, applying various machine learning algorithms to predict customer behaviour within the e-commerce market holds promise for businesses seeking to tailor their offerings to meet specific customer needs and preferences.

.

REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.
[2] Rakshit, P., Srivastava, P. K., Afjal, M., & Srivastava, S. K. (2021). Sentimental Analytics on Indian Big Billion Day of Flip Kart and Amazon. SN Computer Science. https://doi.org/10.1007/s42979-020-00441-3
[3] Liu, B. (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.
[4] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., &Passonneau, R. (2011). Sentiment Analysis of Twitter Data. In Proceedings of the Workshop on Languages in social media (pp. 30-38).
[5] Halde, A., Uttekar, A., & Vishwakarma, A. (2022). Sentiment Analysis on Amazon Product Reviews. International Research Journal of Modernization in Engineering Technology and Science, 04(04), 1280.

