New Generation Computing
https://doi.org/10.1007/s00354-024-00242-9

The Impact of Sentiment Scores Extracted from Product Descriptions on Customer Purchase Intention

Yi Sun[1] · Kaira Sekiguchi[1] · Yukio Ohsawa[1]

Received: 20 August 2023 / Accepted: 22 November 2023

Abstract

This study investigates whether and how the textual content of product descriptions, especially the sentiment element, influences buyers' purchase intentions. Using year-round digital transaction data from Mercari, a leading e-Commerce platform in Japan, we examine the interplay of hard and soft information signals exchanged between sellers and buyers. The study addresses two crucial questions: (1) Do the descriptions that sellers provide on product sales pages impact the buyer's intent to purchase? and (2) In what way does the description influence the buyer's purchase intention? Quantitative analysis is used to understand the relationship between product descriptions, sentiment elements, and purchase intentions. The results show that sentiment factors in product descriptions can serve as high-quality "signals" that can help buyers make informed purchasing decisions and reduce information asymmetry between buyers and sellers. This research contributes to understanding decision-making in online markets, particularly the role of soft information and sentiment analysis.

Keywords Purchase intentions · Sentiment analysis · Product descriptions · e-Commerce

✉ Yi Sun
sun-yi650@g.ecc.u-tokyo.ac.jp

Kaira Sekiguchi
kaira@sys.t.u-tokyo.ac.jp

Yukio Ohsawa
ohsawa@sys.t.u-tokyo.ac.jp

[1] Department of Systems Innovation, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 1138654, Japan

1 Introduction

Currently, sustainable development has become a pressing need, reflecting not only economic development but also the social and environmental dimensions of development. The Sustainable Development Goals (SDG) of the United Nations emphasize the need for a more balanced approach to address various social, economic, and environmental issues [1].

A common thread that runs through these dimensions is the promotion of trust. We believe that this is the key to sustainable development. Business transactions, especially in the context of customer-to-customer (C2C) transactions, have been the subject of various studies with a wide variety of perspectives [2]. However, the role of trust in business has not been fully elucidated even in the C2C domain [3, 4].

In recent years, C2C platforms, especially e-Commerce marketplaces selling used products, have received more attention. Under the influence of the COVID-19 pandemic, more and more people are choosing to buy products on used product trading platforms [5]. Furthermore, e-Commerce platforms are now more than just a place to sell second-hand products; more users are becoming accustomed to selling items on this platform that are difficult to purchase offline, such as masks that have not been seen in offline stores for some time [6].

Therefore, our research aims to mathematically reveal the role of trust in commerce and explore how machine learning can help build trust, facilitating more sustainable commerce in the process. In this study, we specifically explore a rarely discussed aspect of C2C transactions, the sentiment elements conveyed by sellers in product descriptions, and their impact on buyer intent [7, 8].

To facilitate informed purchasing decisions and successful transactions on e-Commerce platforms, it is crucial to reduce the information asymmetry between buyers and sellers [9, 10]. In the context of e-Commerce platforms, this usually means that the seller has complete information on the product being sold, while the buyer has to decide whether or not to buy the product, knowing nothing about the product except some information provided by the seller and other buyers [11]. This information is called "signaling" [12]. One way to reduce information asymmetry is to use high-quality ’signals’ to help buyers make a purchase decision. For example, online reviews provided by other buyers are an often studied signal that provides information about the product and reflects the customer’s experience with the product [13, 14]. Research has shown that the number and sentiment of online reviews can influence sales [15, 16]. In the context of C2C marketplaces, product descriptions created by sellers are another signal that has received less attention in previous research [17, 18]. Chang et al. [19] examined the impact of the quality of the product posts on consumers on the second-hand marketplace on Facebook through a questionnaire survey. Their study found, for example, that the completeness of product information, the accuracy of information, the aesthetics of the posts, and their popularity had significant effects on customers’ purchase intentions to varying degrees. However, their study did not analyze the text of the posts themselves.

2 Literature Review and Conceptual Framework

This study draws on research in three areas: (1) signaling theory and its impact on product sales; (2) the impact of text information on product sales; and (3) sentiment analysis in text.

2.1 Signaling Theory and Its Impact on Product Sales

In e-Commerce transactions, three parties are typically involved: the e-Commerce platform, the seller, and the buyer. A prevalent feature of these interactions is that these parties do not always share the same level of information about products and the market. The variation in access and interpretation of product-related information by these parties leads to an imbalance known as information asymmetry. One main approach to mitigate this imbalance is the application of signaling theory [9, 12, 20]. This theory focuses on various types of signal that function as tools to mitigate information asymmetry, enabling buyers to discern the quality of the product more accurately. In particular, quality signals can significantly shape a buyer's perception, foster trust, and improve purchase intention, thus influencing product sales [9, 10].

Previous research indicates that a signal possesses two essential characteristics: its observability and its cost [12]. In an e-Commerce environment, where buyers cannot physically evaluate the product, they often depend on quality signals to assess the quality of the product [20–22]. If sellers can provide reliable quality signals, it helps buyers reduce information asymmetry, facilitates trust-building, and improves purchase intention. For example, [20] illustrate that a higher quality website, which acts as a signal, increases the probability of earning trust, stimulating purchase intent, and securing a premium price from potential buyers.

Signals from e-Commerce can generally be classified into hard information and soft information. Hard information is information that can be easily quantified and represented numerically, such as the seller's reputation system, which includes reputation scores, star ratings, or platform recommendations [11]. Metrics such as the number of likes and comments a product receives also constitute hard information, with several studies affirming the substantial influence of these factors on a product's popularity and subsequent sales [23]. For example, [13] found that the higher the number of likes a post receives on platforms such as Facebook, the greater the impact it has on shaping potential consumers' attitudes towards the product positively. Additionally, [19] propose that sellers aim to increase the number of likes, comments, and shares on platforms like Facebook's second-hand marketplace. The higher these figures, the greater the popularity of the post, thereby convincing potential buyers that the product offered is a worthwhile purchase."

On the other hand, soft information encompasses data that are not readily quantifiable and often require contextual understanding [11]. Common examples of soft information include online reviews and product descriptions. Previous research indicates that positive reviews from sellers can promote increased product sales [24–26]. Essentially, the more accurate and detailed the signals a seller can convey, the stronger their influence on trust-building. This, in turn, amplifies the buyer's willingness to make a purchase [3, 4, 27].

2.2 The Impact of Soft Information Signals on Product Sales

In e-Commerce research, previous studies have shown mixed results on the impact of online reviews on product sales. Kim et al. [6] suggest that the influence of online reviews on customer purchase decisions is altered by other signals, indicating a need to study additional soft information signals. Although research has focused on product descriptions in crowdfunding, there is a lack of attention paid to product descriptions in the context of second-hand e-Commerce. Chang et al. [19] found that more complete and accurate information in product sales posts leads to a higher probability of purchasing. However, they did not analyze the text of the posts themselves. To address this gap, researchers are exploring the use of machine-learning approaches, such as topic modeling and sentiment analysis.

Traditionally, most research on soft information signaling has focused on online reviews. However, the results of these studies are inconsistent and even vary within the same product category. For example, the impact of online reviews on movie ticket box office sales varies from study to study. Some findings suggest a positive correlation, while other studies conclude that there is no significant effect [14, 15]. Kim [6] argues that while the content of online reviews does influence customers' purchasing decisions, the extent of the influence is moderated by other signals. Specifically, when customers have access to other signals that reduce uncertainty about product quality, they rely less on online reviews. This, in turn, decreases the influence of online reviews on customers, helping to explain the conflicting findings in previous research on the impact of online reviews [6].

In contrast, textual information about product descriptions, another form of soft information signaling, has received relatively little attention, with most existing research focusing on the crowdfunding domain. In the context of crowdfunding websites, [17] explored the importance of project descriptions for project success. They identified three new signals from the project descriptions: length, readability, and tone. They found that these signals improved the accuracy of project success prediction models. In addition, they found that this new information had a significant impact on the success rate [17]. Jiang et al. [18] also noted that themes extracted from project descriptions significantly affect performance.

In the field of second-hand e-Commerce, [19] investigated the correlation between buyer purchasing decisions and the quality of product sales posts on Facebook. Their findings suggest that buyers are more inclined to pay attention to articles with comprehensive and accurate information, and therefore regard these articles as a reliable source for purchasing products [19]. However, their study did not include an analysis of the text of the posts themselves.

2.3 Methods of Text Analysis

This study offers insights into the application of machine-learning techniques in marketing research, focusing on the extraction and analysis of sentiment from textual data. With the advancement of digital technologies, researchers are using large amounts of unstructured data, primarily textual data, to gain actionable insights. In this context, topic modeling, widely recognized in the machine learning community, has been conceptualized as the grouping of large amounts of discrete textual data to reveal underlying patterns by aggregating semantically related words [28].

Blei et al. [28] further refined the premise of topic modeling by proposing the Latent Dreichler Assignment (LDA) model. The model posits that each topic is represented as a distribution of words in a document, and, conversely, each document is represented as a distribution of different topics [29]. However, as with most approaches, topic modeling faces a number of challenges. A prominent issue is the number of topics manually specified, which has a critical impact on the validity of the model [30]. Several studies confirm this view [18, 31, 32].

In 2018, [33] took an innovative perspective by comparing thematic modeling with physical systems. Koltkov advocates the inclusion of physical metrics such as entropy calculations, arguing that they can enhance topic models. By integrating benevolent entropy, a method was developed to autonomously determine the optimal number of topics. The underlying principle of this approach is that the maximum actionable information is obtained when the entropy associated with the number of topics (similar to the temperature in a physical system) is minimized [34].

Although topic modeling offers a significant number of possibilities for textual data review, it remains critical to discriminate methods that are appropriate for specific research needs. While recognizing the potential of topic modeling, this study relies heavily on sentiment analysis to achieve its unique goals. Through the use of a sentiment dictionary, this study reveals the subtle sentiment tone that permeates product descriptions and other textual content on e-Commerce platforms. This approach provides insight into recognizable emotions, both positive and negative, that can significantly impact purchase decisions [35].

Sentiment is a subjective emotion that can influence an individual's thoughts and judgments [7, 8]. In the previous studies, textual data have often been used to extract sentiments such as praise or criticism; these insights can be used by others as a reference for decision-making [16]. As a result, sentiment analysis, especially for product or service reviews, has gained considerable popularity [8, 16, 36]. The main method of calculating sentiment scores is to utilize the word frequency of words in a text that have been assigned a sentiment value in a predefined dictionary to derive a sentiment score for the text. Some studies have aggregated positive and negative sentiments to arrive at a final score [8], while others have calculated positive and negative sentiments separately and examined their impact on the audience [18]. Some researchers have even hypothesized that customer attitudes towards sentiments in texts change with product quality [6].

Despite these advances, there is a clear research gap in analyzing the sentiment of descriptions made by sellers in business settings. Product descriptions are one of

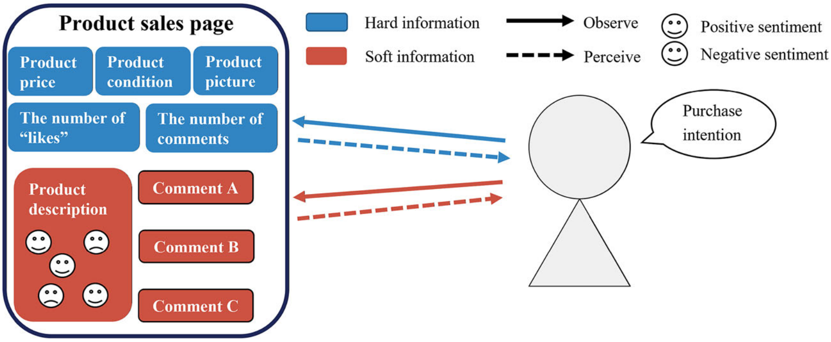

Fig. 1 Signal exchanges between sellers and buyers, adapted from Sun [5]

the critical components of online product information. These pieces of text, carefully crafted by sellers, provide crucial insights into a product's features, usability, and other salient details that are often the deciding factor for customers contemplating a purchase. This research embarks on a journey to uncover the hidden signals within these descriptions and their influence on consumers' purchase decisions. Based on the results of previous studies, this paper also uses sentiment scores computed through word frequency to observe whether this variable will actually affect buyers' purchasing decisions, in which aims to shed light on an integral part of online consumer behavior.

2.4 Research Framework

The primary objective of this study is to explore the influence of product descriptions and sentiment components, calculated based on word frequency, on customers' purchase intentions. To accomplish this, the research scrutinizes product description texts drawn from sales data collated from Mercari, one of the largest online shopping platforms in Japan. Figure 1 provides a visual representation of the prospective signals exchanged between sellers and buyers on the Japanese second-hand e-Commerce site Mercari, a conceptual framework adapted from previous research [5].

The selling process begins with sellers listing their products on the platform, providing crucial details such as the product's price, condition, accompanying photos, and a written description. Customers receive these signals and, in response, provide feedback in the form of 'likes' and comments. This customer feedback, directed to the seller, can serve as a reference point for other potential buyers considering the same product.

In this setup, certain elements, such as the price of the product, its condition, the number of 'likes', and the number of comments, can be classified as hard information, as they are quantifiable in numerical terms. On the contrary, the product description and user comments are considered soft information due to their textual nature. These soft signals often harbor the unspoken sentiments of both sellers and buyers. Furthermore, product photos can encapsulate both hard and soft information; they carry explicit digital information, but can also showcase contextual information that requires interpretation. Buyers, upon receiving these signals, make a purchase decision based on their interpretations.

The focus of this study lies in uncovering the hidden signals in the product description texts written by sellers and their subsequent influence on buyer purchase intentions, a subject scarcely discussed in previous research. To illustrate the impact of product descriptions on customer behavior, this study aims to answer the following questions: (1) Do the descriptions that sellers provide on product sales pages impact the buyer's intent to purchase? (2) If so, in what way does the description influence the buyer's purchase intention? To address these research objectives, the study employs a quantitative methodology to dissect the relationship between product descriptions, sentiment elements derived from word frequency, and purchase intentions. The findings of this study have the potential to provide invaluable insights for online sellers and marketers, allowing them to optimize product descriptions and improve customer satisfaction.

3 Data and Methodology

3.1 Data

The main objective of this study is to explore the relationship between product descriptions and customers' purchase intentions. More specifically, our objective is to figure out whether and to what extent the sentiments embedded in product descriptions influence customer behavior. For this purpose, we use product sales data from Mercari, one of the largest online shopping platforms in Japan, known for its extensive marketplace for second-hand goods.

We chose to focus on the sales of all figure-based products for several reasons. In particular, the wide variety of figures available on Mercari includes limited-edition items that are usually sold at high prices. For example, an item initially priced at 100,000 yen can subsequently sell for 300,000 yen or even 500,000 yen due to its rarity and collector demand. In such a market environment, both buyers and sellers are likely to invest a great deal of effort in shaping the transnational relationship between them, making the figure category a rich field for studying the impact of product descriptions and their sentiment elements.

Our data set contains more than 4.4 million transactions recorded over a 1-year period, covering product sales from January through December of a given year over the past 5 years. This robust data set encompasses all aspects of each sale, including the product description, the price listed, the number of product reviews received, the number of "likes" collected, and the condition of the product at the time of sale.

Using text-analytic techniques and econometric models, this study seeks to reveal hidden patterns and insights that can help sellers maximize profits and build stronger relationships with customers. Additionally, these findings are valuable to Mercari, which is working to optimize its platform and improve the user experience for both sellers and buyers.

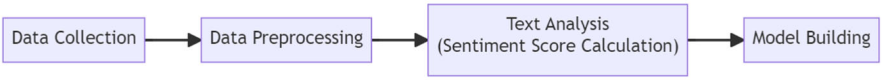

Fig. 2 Methodology of this research

3.2 Methodology

The research methodology used in this study is a combination of quantitative analysis techniques and a comprehensive study of text data. Our approach can be depicted in Fig. 2, which we will detail later.

1. Data Collection: This stage involves collecting sales data from the Mercari platform, including product descriptions, product prices, number of 'likes', and the number of reviews. This data set serves as a basis for subsequent stages.
2. Data Pre-processing: We pre-processed the collected data to eliminate inconsistencies and prepare them for further analysis. This includes segmenting product descriptions into individual words and associating them with their respective sentiment scores, using the Japanese sentiment dictionary [35].
3. Text Analysis (Sentiment Score Calculation): Using the pre-processed data, we conduct a textual analysis of the product descriptions. Each word in the product description is assigned a sentiment score from 1 to 5, indicating its degree of negative to positive. These scores are then used to calculate an overall sentiment score for each product description.
4. Model Building: We build a logistic regression model using the calculated sentiment scores along with other hard information such as the number of 'likes', the price of the product, and the number of reviews. These variables serve to predict the sale outcome (sold or canceled) for each product.

With our data set and the process ready, we proceed as follows.

Firstly, we employ a Japanese sentiment dictionary [35] produced by psychologists to translate the sentiments encapsulated in the product descriptions. These descriptions, crafted by the sellers of Mercari, offer valuable information on their attitudes, positive or negative, towards the products. We use this dictionary to assign sentiment scores to product descriptions, treating these scores as practical representations of the seller's attitude. This process can also be expressed as follows:

$$\text{Sentiment Score} = \sum_{w} \text{score}_w, \tag{1}$$

where score_w is the sentiment score for the wth word in the product.

In the next step, we use the sentiment scores obtained in Step 1 as variables in the logistic regression, expressed in Equation 2

$$\begin{aligned} \ln\left(\frac{P(\text{Sold}=1|x)}{1-P(\text{Sold}=1|x)}\right) &= \alpha + \beta_1 * Count_{\text{like}} + \beta_2 * \text{Count}_{\text{comment}} \\ &\quad + \beta_3 * \text{Price}_{\text{product}} + \beta_4 * \text{Sentiment Score}. \end{aligned} \tag{2}$$

In this model, other signals with hard information, such as the number of ’likes’, the price of the product, and the number of reviews, have also been used. We list the variables’ definitions used in the model in Table 1.

The dependent variable “sold” in our study is a binary variable that represents the status of the sale of a product. In this representation, a value of 1 means that a product is sold out, while a value of 0 means that the sale has been canceled. We have intentionally excluded products that are still available for sale. The rationale behind this choice is that the outcome of these sales has yet to be determined, creating an element of uncertainty. As such, we cannot conclusively determine whether the product description has any influence on these active sales or not.

In this investigation, we have chosen to use sentiment scores as an independent variable within our model. This decision has been driven by our aim of investigating whether hidden information within product descriptions, as revealed by sentiment scores, has any bearing on product sales. Accordingly, to verify whether sentiment scores have a more pronounced effect on sales, we also utilize as independent variables certain signals that contain ’hard information’ and are displayed on the product sales page. These signals include, for example, the total number of ’likes’ given by potential buyers, the total number of comments left by these buyers, and the price listed of the product.

To comprehensively examine the effect of varying price points on our findings, we segment the data and apply a logistic regression model at multiple price criteria. These criteria involve systematically removing data below a set price threshold, starting at intervals of 1,000 yen between 1,000 yen and 100,000 yen, and then shifting to intervals of 10,000 yen between 100,000 yen and 200,000 yen. We have chosen not to segment further beyond the 200,000 yen mark. The motivation behind this methodology is to ensure an adequately sized sample for each segment, thus ensuring the statistical robustness of our findings.

To evaluate our hypothesis in a more rigorous manner, we employ the technique of K-Fold cross-validation. This approach aids in mitigating the risk of over-fitting our model to the data. For the purposes of this study, we set the value of K at 5. This choice balances the need for a thorough exploration of the data, while avoiding over-fitting, thereby ensuring the reliability of our conclusions.

4 Empirical Result

Table 2 presents the results of the descriptive statistical analysis of the independent and dependent variables in this study. The median sentiment score derived from the text analysis was 18.88 and the mean was 26.85. Higher scores indicate more positive sentiments embedded in the text. Understanding the sentiment score is critical, because it provides an empirical way to understand customer sentiment. Higher sentiment scores indicate greater acceptance of a product or service, which can inform marketing strategies and potentially increase customer engagement.

When examining data related to product pricing, the median price was determined to be 2280 yen, indicating that half of the products fall into the low price range. However, the average price was as high as 21,142 yen, suggesting that the data set

Table 1 Definition of variables

Type	Variable	Definition
Dependent variable	Sold	A dummy variable 1 if the product d is in the state of 'sold' and 0 if the product d is in the state of 'cancel'
Independent Variable	$\mathrm{Count}_{\mathrm{like}}$	The total number of 'likes' that potential buyers clicked
	$\mathrm{Count}_{\mathrm{comment}}$	The total number of comments that potential buyers made
	$\mathrm{Price}_{\mathrm{product}}$	Product price
	Sentiment score	The sentiment score extracted from the product description

Table 2 Descriptive statistics

Variables	Min	Median	Mean	Max	Std
Sentiment score	0	18.88	26.85	385.42	28.81
Price	300	2280	21,142	9,999,999	380671.44
Num of likes	0	1	2.30	357	3.29
Num of comments	0	0	1.16	479	2.61
Sold	0	1	0.61	1	0.49

contained high-priced products that deviated from the average price. These findings emphasize the heterogeneity of product price ranges and the existence of high-end products. Sellers can use this information to cater to different customer segments and refine pricing strategies to maximize revenue potential.

In terms of buyer-seller interaction metrics, the median and mean number of "likes" were 1 and 2.3, respectively. On the contrary, the median and mean number of reviews were 0 and 1.16, respectively. These data suggest that a large portion of product buyers are not actively interacting with sellers. From a business perspective, this finding suggests that there is an opportunity to encourage customers to be more actively engaged, making it possible to develop strategies that promote positive interactions between buyers and sellers, thereby increasing customer satisfaction and loyalty.

To better understand sentiment scores and their inherent attributes, we logarithmically transformed the data and integrated them into a logistic regression model, as shown in Fig. 3.

The variable coefficients in Fig. 3 provide important information from both a statistical and a business perspective. The blue line represents the constant coefficient, which indicates that the customer's initial purchase intention is negatively related to the product price. In other words, an increase in the price of the product reduces the likelihood that customers will make a purchase. This finding is a key to the pricing strategy and implies that sellers should be careful with their pricing policies, because too high a price will discourage potential buyers, which may reduce sales.

The orange line indicates the review coefficient. Initially, this line shows an upward trend, indicating that an increase in the number of reviews positively influences customers' purchasing decisions, especially for low-priced items. However, this effect decreases as the price of the product increases and eventually levels off. Therefore, for low-priced items, sellers should stimulate and promote customer reviews to increase sales. For higher priced items, a large number of reviews may be less persuasive, suggesting the need for a different marketing strategy.

In contrast, the red line indicates the "like" factor. The impact of this factor may be negligible initially, but increases as the price of the product rises. The number of comments and likes are both important indicators of a product's popularity. The unique impact of these variables can be attributed to the unique dynamics of Mercari, where reviews often go beyond simple product reviews to include aspects such as price negotiation discussions. This observation suggests that sellers of high-priced items should focus on accumulating "likes," as this appears to enhance the appeal and

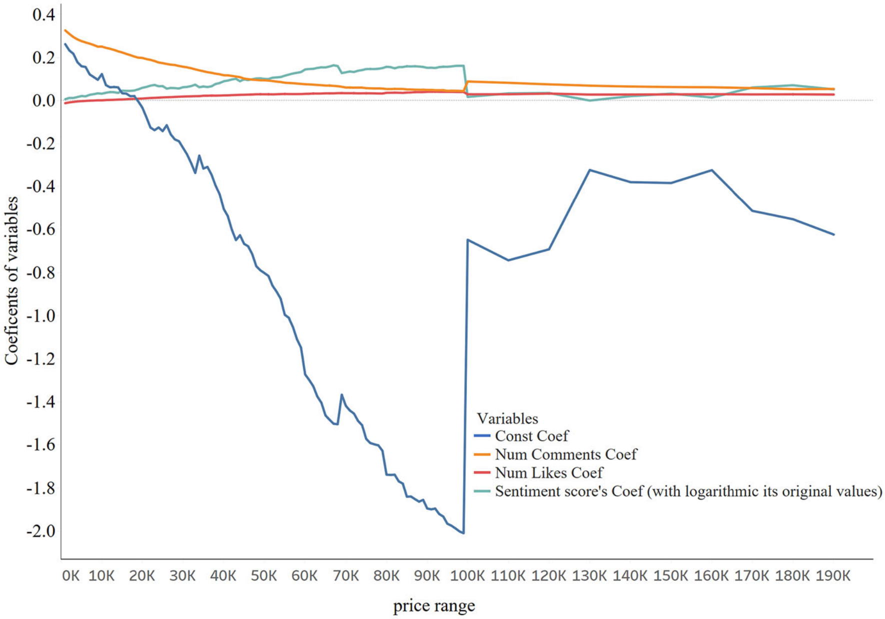

Fig. 3 Sentiment scores' coefficient in different price range

potential sales of their products. In addition, sellers should understand and adapt to the different characteristics of their respective platforms.

The green line corresponds to the sentiment score factor. The impact of the sentiment score increases with the price of the product. Customers may pay more attention to the sentiments conveyed in product descriptions when considering higher priced items. However, it should be noted that this trend reverses for all variables above the 100,000 yen threshold. For sellers, this emphasizes the need to create emotionally rich product descriptions, especially for higher priced items. For ultra-high-priced items (over 100,000 yen), a different strategy is required as the influence of sentiment scores and "likes" diminishes.

Figure 4 presents the *p* values associated with each coefficient. For most price ranges, the *p* values hover around 0, indicating that the coefficients are statistically significant. Thus, the correlation between customer behavior and the evaluation variables (product price, number of reviews, number of "likes", and sentiment scores) is not coincidental. From a business perspective, this highlights the importance of these variables in influencing customer buying behavior, suggesting that they should be central to the development of business strategies.

Beyond the 100,000 yen range, the p value associated with sentiment scores deviates from the nonsignificant region, suggesting that there are limitations to sentiment scores calculated using word frequency, especially for high-priced products. This insight is crucial for sellers of high-priced products, suggesting that relying solely on sentiment scores calculated using traditional methods may not provide accurate insight into customer sentiment, which may lead to ineffective business strategies.

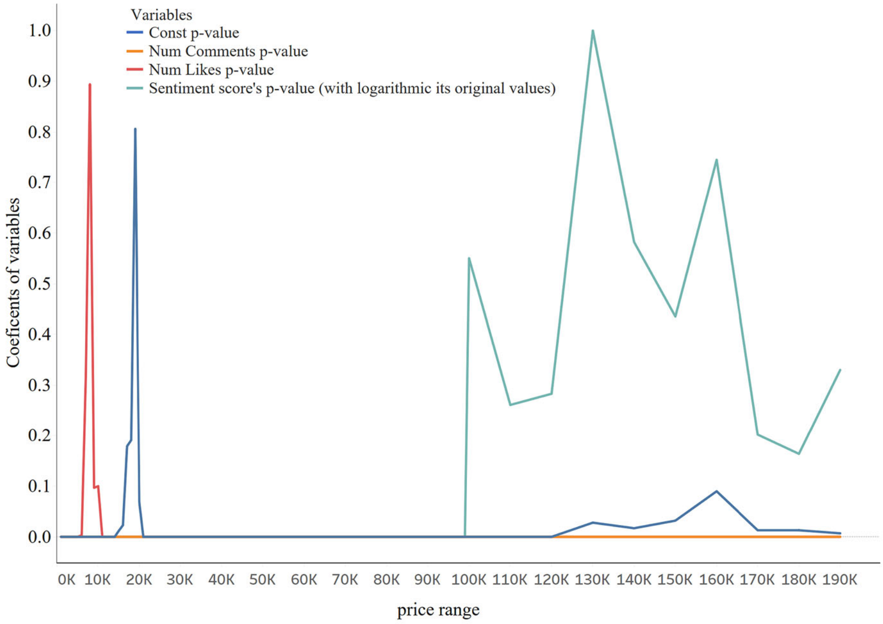

Fig. 4 Sentiment scores' *p* value in different price range

The results of the study reveal key insights related to two questions posed in Chapter 2.5. (1) The study found that product descriptions significantly influence buyers' purchase intentions. Specifically, the sentiment score in product descriptions (shown as the green line in Fig. 3) increases with product price, suggesting that customers may place more importance on the sentiment expressed in product descriptions when considering higher priced products. (2) We note that these descriptions moderated purchase decisions in a way that was related to price. For example, sentiment scores derived from product descriptions had a significant effect on purchase decisions for medium-priced products, but the effect appeared to diminish for ultra-high-priced products (100,000 yen or more). This emphasizes that the impact of product descriptions on purchase intention varies by product price point and suggests that sellers should adjust product descriptions according to product price point.

This finding highlights the need to investigate alternative methods for calculating sentiment scores for high-priced products. This avenue for future research will not only address existing limitations, but also provide organizations with a more accurate tool for measuring customer sentiment, thus enabling them to adjust their marketing strategies more effectively. From a business perspective, this will require sellers to continually refine their analytic processes and keep up with the development of sentiment analysis technology, especially for high-end products.

5 Conclusion

The main objective of this study is to experimentally investigate whether and how the textual content of product descriptions, especially the sentiment element, influences customers' purchase intentions. In an era where sustainability is becoming increasingly important, understanding trust relationships in business, especially on customer-to-customer (C2C) platforms, has become a vital, yet often overlooked aspect of sustainable business practices.

Based on year-round digital transaction data from Mercari, Japan's leading C2C second-hand trading platform, we found that product descriptions written by sellers do influence buyers' purchasing decisions to some extent. To realize this analysis, we informatively used a set of well-established Japanese sentiment dictionaries to extract sentiment scores from textual descriptions. This method assigns scores ranging from 1 (indicating negative sentiments) to 5 (indicating positive sentiments) to individual words in the descriptions.

A noteworthy trend emerged during the analysis: As the price of the product increased, the description had an increasingly positive effect on the purchase decision. This observation suggests a compelling relationship between positive sentiment elements in product descriptions and better sales performance. We measure this sales performance using a logistic regression model that incorporates various factors such as the number of product reviews and likes. However, the relationship becomes statistically insignificant when the price of the product exceeds the 100,000 yen threshold.

Our study demonstrates that on C2C platforms such as Mercari, customers tend to take the sentiment tone of product descriptions more seriously as the price of the product increases. For lower priced items, sentiment content seems to have less impact. This observation suggests that textual content does influence customers' purchase propensity to some extent, with far-reaching potential implications for C2C e-Commerce practices.

Our study has practical implications for e-Commerce managers. First, our results emphasize the impact of product descriptions on consumers' purchasing decisions, especially the fact that sentiment factors can serve as high-quality "signals" that can help buyers make informed purchasing decisions and reduce information asymmetry between buyers and sellers. To increase sales success, sellers should pay attention to the sentiment color of the language when creating product descriptions and use positive words as much as possible to enhance buyers' willingness to purchase. Second, our study also provides information on the design and management of e-Commerce platforms. Platforms can help sellers create more attractive product descriptions by providing better tools and guidance, such as setting up some writing guides or tools that automatically optimize product descriptions.

However, our study has limitations. We focus mainly on how hidden sentiment scores in product descriptions affect purchase intention, thus avoiding the inclusion of too many variables in the logistic regression model. Future studies can further enrich this approach by incorporating more variables (e.g., time from listing to selling of the item) to gain a more comprehensive understanding of the purchase decision mechanism. Furthermore, it is necessary to address anomalous data, such as products with unreasonably inflated prices, in subsequent studies.

In addition, the results of the current study are limited to specific product categories on the Mercari platform. To gain a more comprehensive understanding of the observed phenomena, future research should explore differences between product categories. The limited explanatory power of our analysis for high-priced products means that simple word frequency calculations may not be sufficient. This limitation suggests that more sophisticated methods are needed to explore the sentiment scores contained in the texts.

In conclusion, this study provides valuable information on the role of sentiment in product descriptions and its impact on consumer purchase decisions. Although there is room for improvement and further research, these findings contribute to an initial understanding of how consumers respond to sentiment signals in product descriptions. More importantly, they emphasize the potential importance of creating more enjoyable and trustworthy shopping environments, especially on C2C platforms, to improve sustainable business practices by strengthening the critical role of trust in such environments.

Acknowledgements This research is part of the results of Value Exchange Engineering, a joint research project between R4D, Mercari, Inc. and the RIISE, University of Tokyo. This study is also partly supported by JSPS KAKENHI 20K20482.

Funding Open Access funding provided by The University of Tokyo.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author declares that there is no conflict of interest or personal relationships that could have appeared to influence the work reported in this article.

References

1. Oláh, J., Kitukutha, N., Haddad, H., Pakurár, M., Máté, D., Popp, J.: Achieving sustainable e-commerce in environmental, social and economic dimensions by taking possible trade-offs. Sustainability **11**(1), 89 (2018)
2. Sun, K., Liu, Y., Chen, X., Szolnoki, A.: Evolution of trust in a hierarchical population with punishing investors. Chaos Solit. Fract. **162**, 112413 (2022). https://doi.org/10.1016/j.chaos.2022.112413
3. Chen, J., Zhang, C., Xu, Y.: The role of mutual trust in building members' loyalty to a c2c platform provider. Int. J. Electron. Commer. **14**(1), 147–171 (2009). https://doi.org/10.2753/JEC1086-4415140105
4. Chen, D., Lai, F., Lin, Z.: A trust model for online peer-to-peer lending: a lender's perspective. Inform. Technol. Manag. **15**, 239–254 (2014). https://doi.org/10.1007/s10799-014-0187-z
5. Sun, Y., Ohsawa, Y.: The impact of product descriptions on customers' trust in sellers and purchase intention, pp. 1930–1935 (2022). https://doi.org/10.1109/BigData55660.2022.10020527

6. Kim, R.Y.: When does online review matter to consumers? The effect of product quality information cues. Electron. Commer. Res. **21**(4), 1011–1030 (2021). https://doi.org/10.1007/s10660-020-09398-0
7. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data **2**(1), 1–14 (2015). https://doi.org/10.1186/s40537-015-0015-2
8. Situmeang, F., Boer, N., Zhang, A.: Looking beyond the stars: a description of text mining technique to extract latent dimensions from online product reviews. Int. J. Market Res. **62**(2), 195–215 (2020). https://doi.org/10.1177/1470785319863619
9. Mavlanova, T., Benbunan-Fich, R., Koufaris, M.: Signaling theory and information asymmetry in online commerce. Inform. Manag. **49**, 240–247 (2012). https://doi.org/10.1016/j.im.2012.05.004
10. Spence, A.: Signaling in retrospect and the informational structure of markets. Am. Econ. Rev. **92**, 434–459 (2002). https://doi.org/10.1257/00028280260136200
11. Liberti, J.M., Petersen, M.A.: Information: hard and soft. Rev. Corporate Finance Stud. **8**(1), 1–41 (2018). https://doi.org/10.1093/rcfs/cfy009
12. Connelly, B.L., Certo, S.T., Ireland, R.D., Reutzel, C.R.: Signaling theory: a review and assessment. J. Manag. **37**(1), 39–67 (2011). https://doi.org/10.1177/0149206310388419
13. Lee, K., Lee, B., Oh, W.: Thumbs up, sales up? The contingent effect of facebook likes on sales performance in social commerce. J. Manag. Inform. Syst. **32**, 109–143 (2015). https://doi.org/10.1080/07421222.2015.1138372
14. Liu, Y.: Word of mouth for movies: its dynamics and impact on box office revenue. J. Market. **70**, 74–89 (2006). https://doi.org/10.1509/jmkg.70.3.074
15. Chintagunta, P., Gopinath, S., Venkataraman, S.: The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. Market. Sci. **29**, 944–957 (2010). https://doi.org/10.2139/ssrn.1331124
16. Hu, N., Koh, N., Reddy, S.: Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. Decis. Support Syst. **57**, 42–53 (2014). https://doi.org/10.1016/j.dss.2013.07.009
17. Zhou, M., Lu, B., Fan, W., Wang, G.A.: Project description and crowdfunding success: an exploratory study. Inform. Syst. Front. **20**, 259–274 (2018). https://doi.org/10.1007/s10796-016-9723-1
18. Jiang, C., Han, R., Xu, Q., Liu, Y.: The impact of soft information extracted from descriptive text on crowdfunding performance. Electron. Commer, Res. Appl. 43, 101002 (2020) https://doi.org/10.1016/j.elerap.2020.101002
19. Chang, H., Lu, Y.-Y., Lin, S.: An elaboration likelihood model of consumer respond action to facebook second-hand marketplace: impulsiveness as a moderator. Inform. Manag. **57**, 103171 (2019). https://doi.org/10.1016/j.im.2019.103171
20. Gregg, D., Walczak, S.: The relationship between website quality, trust and price premiums at online auctions. Electron. Commerce Res. 10, 1–25 (2010). https://doi.org/10.1007/s10660-010-9044-2
21. Kardes, F.R., Posavac, S.S., Cronley, M.L.: Consumer inference: a review of processes, bases, and judgment contexts. J. Consum. Psychol. **14**(3), 230–256 (2004). https://doi.org/10.1207/s15327663jcp1403_6
22. Rindova, V.P., Williamson, I.O., Petkova, A.P., Sever, J.M.: Being good or being known: an empirical examination of the dimensions, antecedents, and consequences of organizational reputation. Acad. Manag. J. **48**(6), 1033–1049 (2005). https://doi.org/10.5465/amj.2005.19573108
23. Ter Huurne, M., Ronteltap, A., Corten, R., Buskens, V.: Antecedents of trust in the sharing economy: a systematic review. J. Consumer Behav. **16**(6), 485–498 (2017). https://doi.org/10.1002/cb.1667
24. Bente, G., Baptist, O., Leuschner, H.: To buy or not to buy: influence of seller photos and reputation on buyer trust and purchase behavior. Int. J. Hum. Comput. Stud. **70**(1), 1–13 (2012). https://doi.org/10.1016/j.ijhcs.2011.08.005
25. Li, X., Guo, X., Wang, C., Zhang, S.: Do buyers express their true assessment? Antecedents and consequences of customer praise feedback behaviour on taobao. Internet Res. **26**, 1112–1133 (2016). https://doi.org/10.1108/IntR-03-2015-0063
26. Thierer, A., Koopman, C., Hobson, A., Kuiper, C.: How the internet, the sharing economy, and reputational feedback mechanisms solve the lemons problem. U. Miami L. Rev. **70**, 830 (2015). https://doi.org/10.2139/ssrn.2610255
27. Chen, X., Huang, Q., Davison, R., Hua, Z.: What drives trust transfer? The moderating roles of seller-specific and general institutional mechanisms. Int. J. Electron. Commer. **20**, 261–289 (2015). https://doi.org/10.1080/10864415.2016.1087828

28. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003). https://doi.org/10.1162/jmlr.2003.3.4-5.993
29. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012). https://doi.org/10.1145/2107736.2107741
30. Tang, J., Meng, Z., Nguyen, X.L., Mel, Q., Zhang, M.: Understanding the limiting factors of topic modeling via posterior contraction analysis, vol. 1, pp. 337–345 (2014). https://api.semanticscholar.org/CorpusID:6527691
31. Reisenbichler, M., Reutterer, T.: Topic modeling in marketing: recent advances and research opportunities. J. Bus. Econ. **89**, 1–30 (2019). https://doi.org/10.1007/s11573-018-0915-7
32. Yamada, A., Hashimoto, S., Nagata, N.: A Text mining approach for automatic modeling of kansei evaluation from review texts, pp. 319–328 (2018). https://doi.org/10.1007/978-981-10-8612-0_34
33. Koltcov, S.: Application of rényi and tsallis entropies to topic modeling optimization. Physica A: Stat. Mech. Appl. **512**, 1192–1204 (2018). https://doi.org/10.1016/j.physa.2018.08.050
34. Koltcov, S., Ignatenko, V.: Renormalization analysis of topic models. Entropy **22**(5), 556 (2020). https://doi.org/10.3390/e22050556
35. Nakayama, M., Hatanakaka, C., Konakawa, H., Suzuki, Y., Koh, A., Sugihara, Y., Kawai, T.: Japanese dictionary for sentiment analysis of counselling text. In: Proceedings of the 9th International Conference on Human-Agent Interaction, pp. 311–315 (2021). https://doi.org/10.1145/3472307.3484663
36. Fan, Z.-P., Che, Y.-J., Chen, Z.-Y.: Product sales forecasting using online reviews and historical sales data: a method combining the bass model and sentiment analysis. J. Bus. Res. **74**, 90–100 (2017). https://doi.org/10.1016/j.jbusres.2017.01.010