INTERNATIONAL JOURNAL OF SOCIAL SCIENCE, EDUCATIONAL, ECONOMICS, AGRICULTURE RESEARCH AND TECHNOLOGY (IJSET) | ISSN (E): 2827-766X
Website: https://ijset.org/ | Volumes 5 No. 7 (2026)

THE EFFECT OF PRODUCT ATTRIBUTES AND CUSTOMER REVIEWS ON SALES PERFORMANCE ON THE TOKOPEDIA E-COMMERCE PLATFORM

Aditya Prasetio[1], Dedy Dwi Prastyo[2]
[1] Interdisciplinary School of Management and Technology, Institut Teknologi Sepuluh Nopember, Indonesia;
[2]Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia;
E-mail: [1]prasetio.aditya@gmail.com, [2]dd.prastyo@its.ac.id

Received: 01/04/2026 | **Revised** : 11/04/2026 | **Accepted:**14/05/2026 | **Published** : 24/05/2026

Abstract

This study analyzes the influence of product attributes and customer reviews on sales performance on the Tokopedia e-commerce platform. Data were collected via web scraping from Tokopedia's electronics category (laptops and smartphones), yielding 463 products as the final sample. Product attributes (price, store status, brand clarity, and product description) and customer review indicators (review volume, average star rating, and sentiment score) were used as independent variables, while sales performance (sold count) served as the dependent variable. Sentiment analysis was conducted using a lexicon-based text mining approach with an Indonesian sentiment lexicon. Data analysis was performed using Partial Least Squares Structural Equation Modeling (PLS-SEM) with SmartPLS 4. Results indicate that product attributes have a positive and significant effect on sales performance ($\beta = 0.141$, $t = 5.706$, $p < 0.05$), and customer reviews have a highly significant effect ($\beta = 0.823$, $t = 46.768$, $p < 0.05$). Together, both variables explain 76.9% of the variance in sales performance ($R^2 = 0.769$). Customer reviews, particularly review volume, are the dominant determinant, while store status is the most influential product attribute indicator.

Keywords: customer reviews, e-commerce, product attributes, sales performance, text mining

INTRODUCTION

The development of the digital economy in Indonesia has shown a very rapid trend over the past decade. The Indonesian Internet Service Providers Association (APJII, 2023) reported that more than 215 million Indonesians are connected to the internet, with online shopping as one of the primary activities. The e-Conomy SEA 2023 report by Google, Temasek, and Bain & Company estimated that e-commerce transactions in Indonesia reached more than USD 62 billion, making Indonesia the largest e-commerce market in Southeast Asia. Tokopedia, now part of the GoTo ecosystem, is one of the dominant players with more than 100 million monthly active users. This rapid growth has intensified competition among sellers in the marketplace. Consumers are faced with thousands of similar product choices, making non-physical factors such as product attributes and customer reviews crucial in influencing purchasing decisions. Product attributes such as price, brand, store status (official store or power merchant), and product description quality are the primary information that consumers examine before buying (Li et al., 2019; Halim et al., 2022). On the other hand, customer reviews as a form of electronic word-of-mouth (e-WOM) play an important role in reducing consumer uncertainty. Research shows that the valence and volume of reviews can significantly influence consumer perceptions and ultimately sales performance (Li & Shimizu, 2018; Li, Wu & Mai, 2019; Wang et al., 2022).

Despite the promising potential of the e-commerce market, field phenomena reveal anomalies that are confusing for sellers. Products with seemingly perfect reputations, such as a 5-star rating or Official Store status, are often found to have low or stagnant sales performance (sold count). Conversely, products with simple descriptions or without premium store labels sometimes record high sales figures. This inconsistency indicates that conventional product attributes and simple review metrics may no longer be the sole determinants of consumer purchasing decisions in an increasingly saturated competitive landscape. However, most prior studies on product attributes and customer reviews have been conducted on global marketplaces such as Amazon and Taobao (Chong et al., 2016; Li et al., 2016; Fan et al., 2017). In the Indonesian context, available research tends to focus on purchase intention or on technical aspects of sales prediction using machine learning (Meiryani & Warganegara, 2021; Sunarya et al.,

2025). Accordingly, there is a research gap in directly testing the relationship between product attributes and customer reviews against actual sales performance on a local e-commerce platform such as Tokopedia. This study aims to address that gap by measuring the real impact of the combination of product attributes and review quality on sales performance using a quantitative approach based on data scraped from Tokopedia.

LITERATURE REVIEW

E-commerce is a form of electronic trade that utilizes the internet as the primary medium for buying and selling goods and services. According to Kotler and Keller (2016), e-commerce is the process of buying and selling products through electronic systems, including the internet and computer networks. E-commerce enables consumers to obtain product information, compare prices, and transact conveniently without limitations of space and time. Tokopedia, as one of the main players in the Indonesian e-commerce market and now part of the GoTo ecosystem, has millions of sellers and tens of millions of monthly active users. This open marketplace model creates intense competition where product attributes and customer reviews become key determinants of sales success.

Electronic Word-of-Mouth (e-WOM) is a development of the traditional word-of-mouth (WOM) concept occurring in an online environment. According to Hennig-Thurau et al. (2004), e-WOM is any positive or negative statement made by potential, actual, or former consumers about a product or company, made available to many people through the internet. The fundamental difference between traditional WOM and e-WOM lies in the reach and speed of information dissemination: conventional WOM only reaches a limited scope such as family or friends, whereas e-WOM can reach thousands to millions of consumers in a short time through e-commerce platforms, forums, and social media (Cheung & Thadani, 2012). In the e-commerce context, customer reviews are one of the most dominant forms of e-WOM (Halim et al., 2022), typically comprising star ratings, narrative text, and visual evidence such as photos and videos. The key dimensions of e-WOM relevant to this study are: (1) review volume — the number of reviews available, serving as an indicator of product popularity (Li et al., 2020); (2) review valence — the tendency toward positive or negative reviews, influencing quality perception (Li & Shimizu, 2018); and (3) review credibility — the extent to which reviews are considered helpful and authentic by readers (Wang et al., 2022).

Product attributes are the product characteristics that form the basis of consumer consideration in the purchase process. Kotler and Armstrong (2017) define product attributes as encompassing quality, features, design, and other supporting aspects. In e-commerce, these attributes are manifested digitally: product title, price, description, images, ratings, and special badges such as the Official Store label. According to Cue Utilization Theory (Olson & Jacoby, 1972), consumers use various cues to assess product quality. In e-commerce, where consumers cannot physically touch products, the role of extrinsic cues such as price and store reputation becomes dominant. Li et al. (2019) affirm that price and rating are the dominant attributes in purchase decisions, while Halim et al. (2022) found that product attributes and seller attributes significantly influence purchase intention.

Sales performance is a measure of the extent to which a product or seller successfully achieves sales results in accordance with market targets or potential. According to Kotler and Keller (2016), sales performance can be seen from various indicators, such as sales volume, market share, and sales growth rate. In e-commerce, sales performance is often measured by sales rank, number of units sold, or gross merchandise value. Fan et al. (2017) combined historical sales data with sentiment analysis to improve sales prediction accuracy. Meiryani and Warganegara (2021) used an artificial neural network and successfully predicted sales volume on Tokopedia with high accuracy (95%). Chong et al. (2016; 2017) showed that online reviews are a stronger predictor than price promotions in determining product sales in the digital era.

A review of the literature reveals several research gaps. First, studies using actual Tokopedia data remain scarce, particularly those based on real sales data. Second, most Indonesian studies still use purchase intention as the dependent variable; research based on actual sales is very limited. Third, customer reviews are often studied in isolation, while product attributes are relatively rarely combined in the same model. Fourth, Indonesian studies are dominated by questionnaire-based and SEM approaches, with few applying web scraping combined with data-driven analysis. This study addresses these gaps by jointly modeling product attributes and customer reviews against actual sales performance on Tokopedia using scraped data and PLS-SEM.

METHOD

This study uses a quantitative approach with explanatory research design. The quantitative approach was chosen because the primary objective of the study is to test the influence of independent variables on the dependent variable in a measurable and objective manner. According to Sugiyono (2019), explanatory research aims to explain the causal relationship between variables through hypothesis testing. The research design focuses on testing the effect

of product attributes and customer reviews (as independent variables) on sales performance (as the dependent variable) on the Tokopedia e-commerce platform. This study utilizes secondary data obtained through web scraping techniques, comprising product attributes, customer reviews, and actual sales performance indicators. Research data were collected using the web scraping method from the Tokopedia platform. This method was selected because it can efficiently obtain large volumes of actual data. The object of this study is product data available on Tokopedia, specifically in the electronics category (laptops and smartphones). The data collected include: (1) product attributes — product price, description length, and store status (Official Store or Power Merchant); (2) customer reviews — review volume, average star rating, and review valence obtained through sentiment analysis; and (3) sales performance — the number of units sold (sold count) as displayed by Tokopedia.

The population in this study is all products listed on the Tokopedia e-commerce platform. Purposive sampling was applied with the following criteria: (1) products in the electronics category, specifically laptops and smartphones; (2) products with a minimum of 10 textual reviews; (3) products without textual reviews were excluded as sentiment score calculation was not possible; and (4) products available in active stores. The final sample used in the analysis was 463 products, after removing products that did not meet the minimum review criteria and products with unavailable description data. This sample size exceeds the minimum threshold required by the Power Analysis and the 10-times rule recommended by Hair et al. (2019) for PLS-SEM analysis.

This study uses three groups of variables. The first exogenous variable (X1) is Product Attribute Quality, formed by four formative indicators: product price (ATTR_PRICE), store status (ATTR_SHOP: 0=Regular Merchant, 1=Official Store), brand clarity (ATTR_BRAND: 0=no brand/generic, 1=registered brand), and description quality (ATTR_DESC, measured by character length as a proxy for information completeness). The second exogenous variable (X2) is Customer Reviews, formed by three formative indicators: star rating (REP_STAR), review volume (REP_VOL), and sentiment score (REP_SENT). The endogenous variable (Y) is Sales Performance, measured by units sold (PERF_SALES). Price, review volume, and sold count were log-transformed ($\ln(x+1)$) prior to analysis to correct for skewed distributions.

Prior to statistical analysis, a preprocessing stage was conducted on textual review data. A lexicon-based text mining method using an Indonesian sentiment dictionary (InSet Lexicon) was applied to convert review text into numerical sentiment scores (REP_SENT) ranging from -1 (negative) to +1 (positive). Sentiment scores were calculated per product using a weighted average of all available reviews. Given the mixed-scale characteristics of the data and the use of formative indicators, the data analysis technique employed is Partial Least Squares Structural Equation Modeling (PLS-SEM) using SmartPLS 4 software (Ringle et al., 2024). Formative measurement models were adopted because the indicators form the latent construct rather than reflecting it (Hair et al., 2019). The measurement model evaluation includes multicollinearity testing using Variance Inflation Factor (VIF) and significance testing of outer weights via bootstrapping with 10,000 sub-samples. The structural model evaluation includes the coefficient of determination (R^2) and hypothesis testing through bootstrapping. A path is declared significant if the t-statistic exceeds 1.96 or the p-value is less than 0.05 at the 95% confidence level.

Research Model and Hypotheses

This study adopts a direct effect model in which Product Attribute Quality (X1) and Customer Reviews (X2) are positioned as exogenous variables that independently and simultaneously influence the endogenous variable of Sales Performance (Y). X1 is formed by four indicators: price (ATTR_PRICE), store status (ATTR_SHOP), brand clarity (ATTR_BRAND), and description quality (ATTR_DESC). X2 is formed by three indicators: star rating (REP_STAR), review volume (REP_VOL), and sentiment score (REP_SENT). Y is measured by units sold (PERF_SALES).

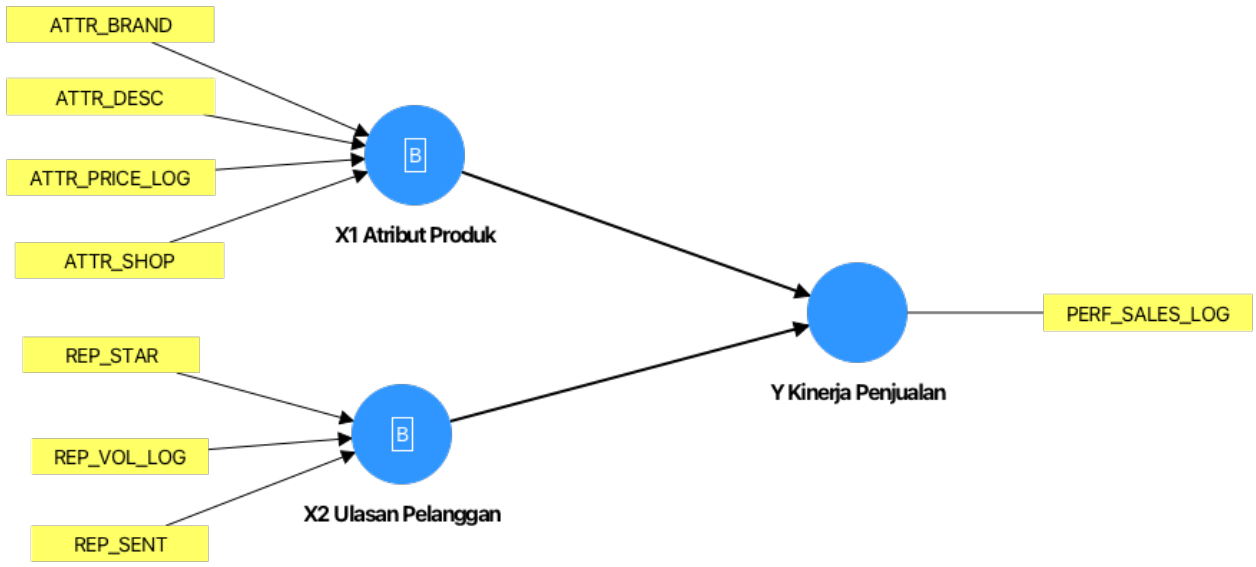

Image: Research Model

Based on the conceptual framework above, the research hypotheses are formulated as follows:
H1: Product attributes have a significant positive effect on sales performance on Tokopedia.
H2: Customer reviews have a significant positive effect on sales performance on Tokopedia.
H3: Product attributes and customer reviews simultaneously have a significant effect on sales performance on Tokopedia.

RESULTS AND DISCUSSION

Data and Descriptive Statistics

Data were obtained through web scraping from Tokopedia in the electronics category (laptops and smartphones), yielding 463 products as the final sample after excluding 4 products without description data (268 laptops, 57.9%; 195 smartphones, 42.1%). Log transformation successfully corrected skewed distributions: PERF_SALES skewness decreased from 4.55 to 0.79. Descriptive findings show that 91.8% of products are sold by Regular Merchants, the average star rating is 4.82/5 — indicating rating inflation — and the average sentiment score is 0.485, reflecting a dominance of positive reviews in the Tokopedia electronics ecosystem.

Measurement Model Evaluation (Outer Model)

All formative indicators showed VIF values below 3.0, confirming the absence of multicollinearity. Inner VIF for both constructs was 1.105, indicating no structural multicollinearity. Table 1 presents the outer weights and significance of each indicator.

Table 1. Outer Weights and Significance of Formative Indicators

Indicator	Weight	T-stat	P-value	Status
ATTR_PRICE_LOG	-0.214	1.767	0.077	Not Significant
ATTR_SHOP	0.832	10.394	0.000	Significant
ATTR_BRAND	0.348	2.953	0.003	Significant
ATTR_DESC	0.282	2.015	0.044	Significant
REP_STAR	0.018	0.656	0.512	Not Significant
REP_VOL_LOG	0.996	113.750	0.000	Significant
REP_SENT	0.148	4.208	0.000	Significant

Two indicators — ATTR_PRICE_LOG and REP_STAR — were not statistically significant but were retained in the model. Following Hair et al. (2019), formative indicators are non-interchangeable; their removal

would reduce the content validity of the construct. Theoretically, price and star rating are fundamental elements of e-commerce purchase decisions, and their inclusion is necessary to preserve the conceptual domain of each variable.

Structural Model Evaluation (Inner Model)

The coefficient of determination $R^2 = 0.769$ (adjusted $R^2 = 0.768$) indicates that Product Attributes (X1) and Customer Reviews (X2) jointly explain 76.9% of the variance in Sales Performance, classified as strong according to Hair et al. (2017). Table 2 presents the effect size of each construct, and Table 3 presents the hypothesis testing results.

Table 2. Effect Size (f^2)

Construct	f^2	T-stat	P-value	Interpretation
Product Attributes (X1)	0.078	2.652	0.008	Small
Customer Reviews (X2)	2.649	7.604	0.000	Large

Table 3. Hypothesis Testing Results (Bootstrapping)

Hypothesis	β	T-stat	P-value	Result
H1: Product Attributes → Sales Performance	0.141	5.706	0.000	Supported
H2: Customer Reviews → Sales Performance	0.823	46.768	0.000	Supported

Effect of Product Attributes on Sales Performance (H1)

H1 is supported: product attributes have a positive and significant effect on sales performance ($\beta = 0.141$, $t = 5.706$, $p < 0.001$). Among the indicators, store status (ATTR_SHOP) is the most dominant (weight = 0.832), confirming Cue Utilization Theory (Olson & Jacoby, 1972): when consumers cannot physically evaluate a product, they rely on extrinsic signals such as Official Store status as a quality proxy. Brand clarity (0.348) and description quality (0.282) also contributed significantly. Price (ATTR_PRICE_LOG) showed a negative weight (-0.214), indicating that in the high-involvement electronics segment, higher prices function as a monetary sacrifice that reduces the overall attractiveness of product attributes relative to competitors. This finding is consistent with Halim et al. (2022) and Çalı & Baykasoğlu (2019).

Effect of Customer Reviews on Sales Performance (H2)

H2 is supported: customer reviews have a positive and highly significant effect on sales performance ($\beta = 0.823$, $t = 46.768$, $p < 0.001$). The magnitude of this coefficient is far greater than that of H1, confirming that customer reviews are the dominant determinant of sales performance on Tokopedia. Review volume (REP_VOL_LOG) is the most dominant indicator (weight = 0.996), consistent with Li et al. (2016). High review volume functions as a strong social proof signal: consumers perceive products with thousands of reviews as "market-tested," thereby reducing perceived purchase risk. Review sentiment (REP_SENT) also contributed significantly (weight = 0.148, $t = 4.208$, $p < 0.001$), consistent with Li, Wu & Mai (2019), confirming that positive textual content strengthens consumer confidence. Star rating (REP_STAR) showed negligible contribution (weight = 0.018) due to the rating inflation phenomenon (mean = 4.82/5), causing ratings to lose discriminatory power, as also observed by Li & Shimizu (2018).

Simultaneous Effect and Comparison of Contributions (H3)

H3 is supported: both variables simultaneously explain 76.9% of sales performance variance ($R^2 = 0.769$), classified as strong (Hair et al., 2017). Effect size comparison reveals a stark contrast: Customer Reviews ($f^2 = 2.649$, large) versus Product Attributes ($f^2 = 0.078$, small). This confirms that in the Tokopedia marketplace ecosystem, consumer-generated e-WOM has become the primary determinant of sales outcomes, surpassing conventional

product attributes. PLSpredict analysis yielded Q^2predict = 0.763 (> 0), confirming the model has predictive relevance. However, the RMSE of the PLS-SEM model (0.632) was marginally higher than the linear model (0.628), indicating low predictive power in terms of point-prediction accuracy, while the model remains strong in explanatory power.

CONCLUSION

This study empirically demonstrates that both product attributes and customer reviews significantly influence sales performance on Tokopedia. H1 is supported: product attributes have a positive and significant effect ($\beta = 0.141$, $t = 5.706$, $p < 0.001$), with Official Store status as the most dominant indicator, confirming Cue Utilization Theory in the digital marketplace context. H2 is supported: customer reviews have a positive and highly significant effect ($\beta = 0.823$, $t = 46.768$, $p < 0.001$), with review volume as the primary social proof signal, while star rating has lost discriminatory power due to rating inflation. H3 is supported: both variables simultaneously explain 76.9% of sales variance ($R^2 = 0.769$). Customer reviews are the far more dominant determinant ($f^2 = 2.649$ vs. $f^2 = 0.078$), indicating that e-WOM has become the primary driver of sales outcomes in the mature Tokopedia marketplace.

These findings carry important practical implications. For sellers, managing review volume and investing in Official Store certification deliver the highest return, with the impact of customer reviews approximately six times greater than product description optimization. For Tokopedia, developing a robust review verification system and mitigating rating inflation would preserve reviews as an objective quality signal. Theoretically, this study validates the influence of product attributes and customer reviews on actual sales performance — not merely purchase intention — in the Indonesian context, and contributes methodologically through the integration of web scraping, NLP sentiment analysis, and PLS-SEM as a reference for future data-driven research.

This study has several limitations. Data are cross-sectional, making causal inference limited due to potential reverse causality between review volume and sales. The sold count variable is interval-censored (lower bound only). The heterogeneous sample limits generalizability to specific product segments. The lexicon-based NLP approach has limited capacity to capture sarcasm and mixed-language expressions. Product age was not controlled as a confounding variable. Future research is recommended to adopt longitudinal designs, use more homogeneous samples, apply transformer-based NLP models such as IndoBERT or IndoRoBERTa, add moderating variables such as discount and competition level, and replicate findings on other Indonesian platforms such as Shopee and TikTok Shop.

REFERENCES

Chong, A.Y.L., Ch'ng, E., Liu, M.J., & Li, B. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. International Journal of Operations & Production Management, 36(4), 358–383.

Cheung, C.M.K., & Thadani, D.R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. Decision Support Systems, 54(1), 461–470.

Fan, Y., Zhang, X., Zhu, S., & Wang, J. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. Electronic Commerce Research and Applications, 25, 102–112.

Halim, F., Santoso, L., & Pratama, R. (2022). Online customer reviews as a marketing tool to generate customer purchase intention in e-commerce. International Journal of Data and Network Science, 6(4), 1049–1060.

Li, B., Ch'ng, E., Chong, A.Y.L., & Bao, H. (2016). Predicting online e-marketplace sales performances: A big data approach. Computers in Human Behavior, 64, 68–75.

Li, X., & Shimizu, A. (2018). Impact of online customer reviews on sales outcomes: An empirical study based on prospect theory. Review of Socionetwork Strategies, 12(2), 173–188.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. Information & Management, 56(2), 246–256.

Meiryani, M., & Warganegara, D.L. (2021). Implementation of artificial neural network in forecasting sales volume in Tokopedia Indonesia. International Journal of Scientific & Technology Research, 10(4), 209–214.

Sunarya, F.U.A., & Kusnawi. (2025). Evaluating classification models for predicting product success in Indonesian e-commerce. International Journal of Advanced Computer Science and Applications, 16(2), 1017–1025.

Wang, Q., Zhang, W., Li, J., Mai, F., & Ma, Z. (2022). Effect of online review sentiment on product sales: The moderating role of review credibility perception. Information & Management, 59(4), 103659.

Çalı, S., & Baykasoğlu, A. (2019). A Bayesian based approach for analyzing customer's online sales data to identify weights of product attributes. Electronic Commerce Research and Applications, 33, 100818.

Chong, A.Y.L., Li, B., Ngai, E.W.T., Ch'ng, E., & Lee, F. (2017). Predicting consumer product demands via big data: The roles of online promotional marketing and online reviews. International Journal of Production Research, 55(17), 5142–5163.

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.

Kotler, P., & Armstrong, G. (2017). Principles of Marketing (17th ed.). Pearson Education.

Kotler, P., & Keller, K.L. (2016). Marketing Management (15th ed.). Pearson Education.

Olson, J.C., & Jacoby, J. (1972). Cue utilization in the quality perception process. Proceedings of the Third Annual Conference of the Association for Consumer Research, 167–179.

Hair, J.F., Risher, J.J., Sarstedt, M., & Ringle, C.M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24.

Ringle, C.M., Wende, S., & Becker, J.M. (2024). SmartPLS 4. SmartPLS. https://www.smartpls.com.

Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Hair, J.F., Henseler, J., Ringle, C.M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (2nd ed.). Sage Publications.

Shmueli, G., Sarstedt, M., Hair, J.F., Cheah, J.H., Ting, H., Vaithilingam, S.S., & Ringle, C.M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. Journal of Business Research, 100, 232–245.

Bhattacharyya, S.S., Jha, S., & Fernandes, B. (2021). Exploring the factors impacting seller performance during online sales events in the e-commerce marketplace in India. Journal of Business Research, 134, 314–329.

Cai, H., Wohn, D.Y., & Mittal, A. (2023). Impact of review informativeness, readability, and multimedia richness on product sales: Evidence from Amazon video games. Electronic Commerce Research and Applications, 56, 101223.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.

Li, Y., Du, C., & Zhang, W. (2024). The effect of different types of comparative reviews on product sales. Electronic Commerce Research and Applications, 59, 101241.

Mudambi, S.M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly, 34(1), 185–200.

Zhou, Z., Zhang, Z., & Wang, C. (2024). Product consumptions meet reviews: Inferring consumer preferences by an explainable machine learning approach. Electronic Commerce Research and Applications, 59, 101245.