FIMEK

Journal of Agronomy, Technology and Engineering Management
ISSN 2620-1755

Article

From Prediction to Decision: Managing Machine Learning for Cart Abandonment in E-Commerce

Brankica Pažun [1,2,*], Aleksandra Golubović-Stojanović [3] and Irena Brajović [4]

[1] Faculty of Engineering Management, Union Nikola Tesla University, Bulevar vojvode Mišića 43, 11000 Belgrade, Serbia.
[2] The Academy of Applied Studies Polytechnic, Katarine Ambrozić 3, 11000 Belgrade, Serbia.
[3] Prof. dr Radomir Bojković Business School of Applied Studies, Topličina 12, 37000 Kruševac, Serbia.
[4] Eighth Belgrade Grammar School, Grčića Milenka 71, 11000 Belgrade, Serbia.
* Correspondence: brankica.pazun@fim.rs

Received: 21 July 2026; Accepted: 1 October 2026

Abstract: This paper examines whether machine learning can help predict shopping cart abandonment in e-commerce. The study uses a publicly available dataset from an electronics store and includes 41,036 sessions in which at least one product was added to the cart. Six indicators of browsing behavior and product prices were analyzed using logistic regression and Random Forest. Both models showed limited predictive performance. Logistic regression achieved a ROC-AUC of 0.563, whereas Random Forest achieved a ROC-AUC of 0.556. Product prices were the most influential indicators in the Random Forest model. On average, prices were higher in sessions that did not result in a recorded purchase, although this finding does not imply that higher prices caused cart abandonment. The results indicate that the available data are insufficient for reliable predictions at the individual-session level. Machine learning may support marketing and management decisions, but its use in an e-commerce information system requires additional data and further evaluation before automated actions are introduced.

Keywords: *Engineering management; management information systems; machine learning; e-commerce process; marketing analytics; cart abandonment.*

1. Introduction

Online stores collect data on the products customers view, the items they add to their carts, and the purchases they complete. These data can help companies better understand customer behavior and support marketing and business decision-making.

Shopping cart abandonment occurs when a customer adds a product to the cart but does not complete a purchase during the recorded session. Adding a product to the cart indicates purchase interest, although it does not necessarily mean that the customer has decided to buy it. Previous studies have examined this issue from the perspectives of consumer behavior, marketing, e-commerce, and information systems [1,2].

Customers may leave a cart for various reasons, including product characteristics, website experience, additional costs, or uncertainty about the purchase [2,3]. Data on customer activity in an online store can provide useful information about this behavior [4,5]. However, such data must be analyzed and interpreted before they can meaningfully support business decisions.

Machine learning provides one approach for analyzing large numbers of customer sessions. It can help identify patterns and estimate whether a session is likely to end in a purchase. Such estimates may be useful for marketing and management, but they should be evaluated carefully before being used to guide actions toward customers [6].

The aim of this study is to examine whether data available when a product is first added to the cart can be used to predict whether a purchase will follow. Logistic regression and Random Forest were compared, and the implications of their predictive performance for the use of machine learning in an e-commerce information system were evaluated. The study also considers what additional data and validation procedures would be required before such models could support routine marketing decisions.

2. Literature and Research Background

Customers often view and compare several products before making an online purchase. They may also add a product to the cart to save it for later. Therefore, a cart that is not followed by a purchase within the same recorded session does not necessarily mean that the customer has lost interest.

Previous studies have linked cart abandonment to trust, perceived risk, website experience, product characteristics, available information, and additional costs [1–3,7]. Studies based on recorded customer activity also show that browsing and cart data can help explain how customers use online stores [4,5]. However, these data show what a customer did, not why the behavior occurred. For example, viewing many products does not necessarily mean that the customer is undecided.

Machine learning can help analyze large numbers of shopping sessions and identify patterns associated with cart abandonment. Rausch, Derra, and Wolf [8] examined this type of prediction, while more recent research has considered additional session data and more complex models [9]. In this study, logistic regression and Random Forest are compared using information available when a product is first added to the cart. The objective is to assess how well these data predict the recorded outcome of a session.

A prediction is useful to a business only if it supports a better decision. A high predicted probability of cart abandonment is not, by itself, sufficient justification for offering a discount or sending a message. The timing of a marketing action can also affect its outcome [10]. For this reason, model results should be considered together with information about the customer, product, and planned marketing action [6].

The use of such a model also requires decisions about how data are collected and how model performance is evaluated. An online store needs consistent records of product views, cart additions, and purchases. Before predictions are used in practice, managers should assess predictive performance and test whether a proposed action improves business results. The information-systems team would manage the data, analysts would evaluate the model, and the marketing team would plan and assess the intervention.

This paper does not evaluate a model already deployed by an online store and does not test a marketing campaign. Instead, it uses existing data to assess the limits of a simple prediction model and to identify what should be evaluated before such a model is used to support routine decisions.

3. Materials and Methods

The study uses the publicly available *eCommerce Events History in Electronics Store* dataset [11]. The analyzed file contains 885,129 events recorded between September 24, 2020, and February 28, 2021. These include 793,748 product views, 54,035 cart additions, and 37,346 purchases across 490,398 customer sessions.

The analysis includes sessions in which at least one product was added to the cart. A total of 234 sessions were excluded because a purchase was recorded before the first relevant cart addition. The final analytical sample therefore comprised 41,036 sessions.

Sessions were divided into two outcome groups. A session was classified as a completed purchase if a purchase was recorded after the first cart addition. If no purchase was recorded afterward within the same session, the session was classified as cart abandonment. This operational

definition therefore captures the absence of a subsequent purchase within the recorded session and does not establish whether the customer purchased the product in a later session.

Table 1. Indicators available at the first cart addition.

Indicator	Definition
Number of views	Product views before first cart addition
Distinct products	Distinct products viewed
Categories	Distinct categories viewed
Brands	Distinct brands viewed
Mean viewed-product price	Mean price of viewed products
First cart-item price	Price of first item added to cart

First, the six indicators were compared between sessions with a recorded subsequent purchase and sessions without one. Because the comparison concerned two independent outcome groups, the Mann–Whitney U test was used to examine between-group differences.

Second, logistic regression and Random Forest were used to predict the binary session outcome after the first cart addition, defined as a recorded subsequent purchase versus no recorded subsequent purchase. The data were divided into a training set (80%) and a test set (20%), while retaining approximately the same proportions of the two outcome groups in both sets. Model development was performed on the training set, and predictive performance was evaluated on the held-out test set.

Model performance was assessed using accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (ROC-AUC). Random Forest feature importance was also examined to identify which of the six indicators contributed most strongly to the model predictions. Feature importance was interpreted only as a measure of model reliance on an indicator and not as evidence that the indicator caused a customer to abandon the cart.

4. Results and Discussion

The final sample includes 41,036 sessions in which a customer added at least one product to the cart. A purchase was subsequently recorded in 20,528 sessions (50.02%), while no purchase was recorded in 20,508 sessions (49.98%). These figures refer only to the sessions included in this study.

Table 2. Indicators by recorded session outcome.

Indicator	Subsequent purchase	No subsequent purchase	p-value
Number of views	1.82	1.74	0.064
Distinct products	1.42	1.40	0.148
Categories	1.041	1.047	<0.001
Brands	0.916	0.944	<0.001
Mean viewed-product price	141.19	173.56	<0.001
First cart-item price	140.59	173.27	<0.001

Table 2 shows that the two groups did not differ significantly in the number of product views or the number of distinct products viewed. Differences in the numbers of categories and brands were statistically significant but small in magnitude.

Larger differences were observed for product prices. The mean price of viewed products was 173.56 in sessions without a subsequent purchase, compared with 141.19 in sessions with a purchase. The price of the first product added to the cart was also higher in sessions without a subsequent purchase (173.27 versus 140.59). However, the dataset does not provide sufficient information to conclude that higher prices caused customers to leave their carts. Other factors, such as product type or the customer's need for additional information, may also be relevant [7].

Table 3. Predictive performance on the test set.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic regression	0.549	0.564	0.429	0.487	0.563
Random forest	0.548	0.549	0.531	0.540	0.556

Neither model predicted the outcome of individual sessions reliably. Logistic regression achieved a ROC-AUC of 0.563, slightly higher than the Random Forest value of 0.556. Random Forest showed higher recall and F1 score. Overall, neither model performed well enough to support automatic marketing actions based solely on the six indicators used in this study.

The observed between-group differences did not translate into accurate predictions at the individual-session level. Product prices differed between sessions with and without a subsequent purchase, but these differences were insufficient for reliable prediction of individual outcomes. Similarly, the statistically significant differences in categories and brands were too small to suggest strong practical value on their own. In addition, the study defines cart abandonment as the absence of a subsequent purchase within the same recorded session; consequently, some customers who completed a purchase in a later session may have been classified as abandonment cases.

Table 4. Proposed engineering management gates for an e-commerce model.

Gate	Evidence required before progression	Accountable function	Status in this study
Data and process	Event order, session linkage, missingness and reconciliation with orders	Information systems	Partial: event order checked; no operational audit
Prediction	Temporal holdout, discrimination, calibration and cost-sensitive error review	Analytics and management	Not met: ROC-AUC 0.563 and 0.556; calibration unreported
Intervention	Controlled comparison of conversion, margin, contact cost and customer feedback	Marketing and management	Not tested
Operations	Monitoring frequency, data drift limits, owner, rollback procedure and privacy review	Operations and information systems	Not tested

Table 4 presents a sequence of steps that an online retailer could follow before using a prediction model in routine operations. First, the retailer should verify that cart and purchase events are recorded consistently and can be linked to other relevant business data. A revised model should then be validated on later sessions. Finally, any proposed marketing intervention should be evaluated in relation to its cost and its effects on purchases and customer experience.

Prediction errors are also relevant from a management perspective. A customer incorrectly classified as likely to leave the cart might receive an unnecessary promotion, whereas a customer incorrectly classified as likely to purchase might not receive useful assistance. The present data do not allow the financial cost of either type of error to be estimated. Before implementation, a retailer would therefore need to compare model-guided actions with simpler approaches and with taking no action.

Table 5. Random Forest feature importance.

Indicator	Relative importance
Mean viewed-product price	44.48%
First cart-item price	41.07%
Number of views	6.88%
Distinct products	3.31%
Brands	3.12%
Categories	1.14%

The mean price of viewed products and the price of the first product added to the cart together accounted for 85.55% of the reported feature importance in the Random Forest model. Thus, the model relied most strongly on the two price-related indicators. This does not show that price caused cart abandonment, and the two price indicators may also contain overlapping information.

For higher-priced products, retailers could consider providing clearer specifications, comparisons, warranty terms, payment options, and delivery information. This represents a possible marketing response rather than an effect tested in the present study. Likewise, the model results do not justify the automatic use of discounts. Previous research suggests that the timing of a marketing message can influence its outcome [10].

Mikić et al. [12] discuss how AI-based personalization may support customer experience and business performance and emphasize the importance of data quality and responsible use. The present study examines a more limited set of data and does not test personalization or a marketing campaign. Its findings therefore illustrate the limitations of a simple cart-abandonment prediction model rather than the effects of using AI to communicate with customers.

Future research could incorporate purchase history, promotions, delivery conditions, and other relevant information available to a retailer, provided that such data are used appropriately. Models should be validated on data from later periods and, where possible, across different online stores. Most importantly, any proposed marketing intervention should be evaluated in a controlled test to determine whether it improves business outcomes.

5. Conclusions

This study examined whether data recorded before the first cart addition can be used to predict a subsequent purchase in an online store. The analysis included 41,036 customer sessions. Mean product prices were higher in sessions without a recorded subsequent purchase, but this finding does not demonstrate that price caused cart abandonment.

Logistic regression and Random Forest showed limited predictive performance, with ROC-AUC values of 0.563 and 0.556, respectively. The six indicators used in this study are therefore insufficient for reliable predictions at the individual-session level or for supporting automatic marketing actions.

Before implementing such a model in practice, online retailers would require more informative purchase, product, and customer-related data, together with further validation of predictive performance. Managers should also evaluate the business effects of any marketing action based on model predictions before introducing it into routine operations.

The study is limited to data from one electronics store and a single observation period. In addition, cart abandonment was defined within the boundaries of the recorded session, so later purchases made in another session could not be identified. Further research should test whether purchase history, promotions, and delivery information improve prediction, validate models on later periods and in other online stores, and assess whether model-supported decisions improve business outcomes.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Declaration of Generative AI Use: During the preparation of this manuscript, the authors used ChatGPT to assist with English translation and language editing. The authors reviewed and edited the output and take full responsibility for the content of this publication.

References

1. Wang, S.; Cheah, J.H.; Lim, X.J. Online shopping cart abandonment: A review and research agenda. *Int. J. Consum. Stud.* **2023**, *47*, 453–473. https://doi.org/10.1111/ijcs.12876.
2. Chopra, I.P.; Jebarajakirthy, C.; Jain, T.; Maseeh, H.I. Electronic shopping cart abandonment: What do we know and where should we be heading? *Electron. Mark.* **2024**, *34*, 25. https://doi.org/10.1007/s12525-024-00697-0.
3. Patharia, I.; Jain, T. Antecedents of electronic shopping cart abandonment during online purchase process. *Bus. Perspect. Res.* **2024**, *12*, 400–418. https://doi.org/10.1177/22785337221148810.
4. Kukar-Kinney, M.; Scheinbaum, A.C.; Orimoloye, L.O.; Carlson, J.R.; He, H. A model of online shopping cart abandonment: Evidence from e-tail clickstream data. *J. Acad. Mark. Sci.* **2022**, *50*, 961–980. https://doi.org/10.1007/s11747-022-00857-8.
5. Rausch, T.M.; Brand, B.M. Gotta buy 'em all? Online shopping cart abandonment among new and existing customers. *Int. J. Electron. Bus.* **2022**, *17*, 109–134. https://doi.org/10.1504/IJEB.2022.121913.
6. Basu, R.; Lim, W.M.; Kumar, A.; Kumar, S. Marketing analytics: The bridge between customer psychology and marketing decision-making. *Psychol. Mark.* **2023**, *40*, 2588–2611. https://doi.org/10.1002/mar.21908.
7. Rodrigues, M.A.; Proença, J.; Soares, I. The influence of perceived risk on mobile shopping cart abandonment. *J. Strateg. Mark.* **2024**, *32*, 152–169. https://doi.org/10.1080/0965254X.2022.2160480.
8. Rausch, T.M.; Derra, N.D.; Wolf, L. Predicting online shopping cart abandonment with machine learning approaches. *Int. J. Mark. Res.* **2022**, *64*, 89–112. https://doi.org/10.1177/1470785320972526.
9. Esmeli, R.; Gokce, A. An analysis of consumer purchase behavior following cart addition in e-commerce utilizing explainable artificial intelligence. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 28. https://doi.org/10.3390/jtaer20010028.
10. Li, J.; Luo, X.; Lu, X.; Moriguchi, T. The double-edged effects of e-commerce cart retargeting: Does retargeting too early backfire? *J. Mark.* **2021**, *85*, 123–140. https://doi.org/10.1177/0022242920959043.

11. Kechinov, M. *eCommerce Events History in Electronics Store*; Open CDP/REES46: Public Dataset; Kaggle. Available online: https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-electronics-store (accessed on 1 October 2026).
12. Mikić, J.; Gardašević, J.; Brkić, I. The impact of artificial intelligence on creating a personalized consumer experience. *J. Agron. Technol. Eng. Manag.* **2026**, *9*, 1870–1885. https://doi.org/10.55817/GHKR1604.

© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).