International Journal of Recent Technology and Applied Science
p-ISSN: 2721-2017, e-ISSN: 2721-7280

Original Research Paper

Predicting Customer Purchase Intent in E-Commerce Platforms Using Session-Based Logistic Regression Analytics

Nor Anisah Salamah Mohd Rosli[1*], Ahmad Fadli Saad[1]

[1] *Computing Science Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA. Perak Branch, Tapah Campus. Perak, Malaysia.*

Article History
Received:
04.08.2026

Revised:
30.08.2026

Accepted:
13.09.2026

***Corresponding Author:**
Ahmad Fadli Saad
Email
afadlis@uitm.edu.my

Abstract: The rapid acceleration of global e-commerce has fundamentally transformed retail operations, yet digital merchants continue to confront substantial financial losses resulting from inefficient inventory planning, cart abandonment, and imprecise consumer behavior forecasting. This research develops ShopPredict, a session-based predictive analytics framework designed to forecast online shopping purchase completion by modeling user interaction dynamics including product viewing frequency, cart additions, session price averages, and unique item explorations using a Logistic Regression classification approach. Utilizing an empirical multi-category e-commerce dataset comprising millions of user session logs, the classification model was trained and evaluated across structured behavioral features. The experimental results demonstrated a high overall accuracy of 93.48%; however, rigorous metric decomposition revealed a critical accuracy paradox, wherein the model achieved a high precision of 0.94 for non-purchase sessions but exhibited a low recall of 0.16 and an F1-score of 0.25 for purchase events due to severe underlying class imbalance. System usability and functional architecture evaluations yielded positive user acceptance scores averaging above 4.0 on a 5-point Likert scale. This study underscores the utility of session-based linear probability modeling while demonstrating that high nominal accuracy in e-commerce prediction can be misleading without rigorous class-rebalancing strategies. Future research must incorporate advanced ensemble architectures, SMOTE oversampling, and dynamic session features to optimize purchase intent detection in real-time retail environments.

Keywords: E-commerce Analytics, Purchase Prediction, Logistic Regression, Customer Behavior Modeling, Session-Based Classification.

1. Introduction

The global retail landscape has undergone a monumental paradigm shift over the past decade, driven by the exponential expansion of e-commerce platforms and the ubiquitous adoption of digital transaction channels. According to recent market analysis [1], electronic commerce has evolved from a supplementary sales pipeline into the primary engine of global commercial growth, redefining how consumer goods and services are distributed, marketed, and acquired. The operational agility of e-commerce enables enterprise networks and small-to-medium retailers alike to transcend geographical boundaries, operate continuously without physical store constraints, and engage diverse consumer segments through personalized digital storefronts. However, despite the massive proliferation of digital storefronts and the sophisticated tracking mechanisms embedded within modern web environments, online merchants face persistent operational vulnerabilities that severely undermine profitability and customer retention [2].

Central to these operational challenges is the inherent uncertainty surrounding online consumer decision-making and purchase finalization [3]. Unlike traditional brick-and-mortar retail environments where physical presence, direct interaction with sales personnel, and immediate point-of-sale queues offer observable indicators of purchase intent, digital shopping environments are characterized by high user anonymity, transient browsing habits, and frictionless platform navigation. Online shoppers routinely navigate through vast product catalogs, accumulate items within digital shopping carts, and subsequently abandon their shopping sessions without completing a financial transaction [4]. Industry benchmarks indicate that cart abandonment rates across major e-commerce sectors consistently exceed 65%, representing billions of dollars in unrealized revenue annually and distorting inventory demand forecasting models [5].

The operational ramifications of unpredicted customer purchasing behavior are particularly acute in the domains of supply chain coordination and inventory optimization [6]. When online merchants are unable to accurately forecast customer purchase probabilities from real-time session dynamics, they are forced to rely on intuitive or historical aggregate demand estimation methods. This reliance frequently leads to two distinct financial hazards: overstocking and understocking [7]. Overstocking ties up critical working capital in excess inventory, increases warehousing and holding expenditures, and exposes merchants to product obsolescence or forced price discounting. Conversely, understocking leads to frequent stockouts during demand spikes, causing immediate loss of revenue, degrading brand trust, and driving frustrated consumers toward direct market competitors [8].

To mitigate these systemic inefficiencies, contemporary e-commerce research has increasingly focused on predictive analytics and machine learning techniques to decode digital consumer behavior [9]. By capturing granular user interaction logs—such as product page views, navigation depth, search query refinement, time spent per item, and cart addition events—predictive models can learn the subtle behavioral patterns that signal a high likelihood of purchase conversion [10]. Automated purchase prediction systems enable merchants to dynamically adjust promotional strategies, issue targeted real-time cart incentives, optimize stock allocation across regional fulfillment centers, and deliver personalized user experiences that streamline the conversion path [11].

Despite the growing body of literature surrounding predictive user modeling, a critical research gap persists in the operationalization of lightweight, highly interpretable, and session-based classification models within small-to-medium enterprise (SME) e-commerce platforms [12]. While complex deep learning architectures and massive ensemble models have demonstrated strong predictive capabilities in high-compute environments, their deployment remains financially and technically prohibitive for smaller online retailers who require low-latency, transparent, and computationally efficient decision-support systems [13]. Furthermore, existing predictive implementations frequently fail to address the severe class imbalance inherent in e-commerce interaction logs, where non-purchase browsing sessions overwhelmingly outnumber completed transaction sessions [14]. This imbalance often produces deceptive performance metrics that conceal poor minority-class detection capability.

To address these empirical and practical limitations, this study presents the design, implementation, and empirical evaluation of ShopPredict, an automated online shopping purchase prediction framework based on session-level classification analytics. By extracting structured behavioral features from large-scale multi-category e-commerce transaction logs and applying a supervised Logistic Regression modeling architecture, this research aims to accomplish three main objectives: (1) to identify key session-level behavioral features that determine consumer purchasing intent, (2) to build and evaluate a computationally efficient probabilistic classification model that distinguishes between purchase and non-purchase sessions, and (3) to systematically assess the functional execution and user acceptance of the deployed predictive framework. By addressing both the predictive modeling performance and the underlying class distribution dynamics, this study provides actionable insights for e-commerce decision-makers and establishes a rigorous baseline for lightweight predictive analytics in digital retail.

2. Literature Review

2.1. E-Commerce Ecosystems and Consumer Intent Modeling

The modern e-commerce ecosystem operates as a highly complex digital marketplace where consumer decision-making is influenced by platform architecture, user experience design, product information richness, and dynamic pricing strategies [15]. Understanding consumer purchasing intent requires a comprehensive examination of the digital touchpoints that precede final checkout execution. Research in digital shopper behavior demonstrates that online consumers progress through distinct cognitive and behavioral stages during a shopping session, moving from exploratory product discovery to focused evaluation and, ultimately, transaction finalization [16]. However, because online browsing incurs minimal friction, shoppers frequently switch platforms, compare alternative vendors, or utilize shopping carts as temporary bookmarking tools rather than immediate purchase commitment mechanisms [17].

The digital shopping cart represents a critical pivot point within the e-commerce sales funnel, serving as the immediate antecedent to financial conversion [18]. Prior empirical investigations have demonstrated that user actions within the cart environment—such as adding multiple items, modifying product quantities, or repeatedly reviewing cart totals—exhibit a strong statistical correlation with final purchase decisions [19]. Nevertheless, cart addition alone is an imperfect indicator of intent, as a substantial proportion of users abandon carts due to unexpected shipping fees, complex checkout procedures, or passive comparison shopping [20]. Consequently, effective predictive frameworks must analyze cart interactions within the broader context of overall session activity, incorporating product category diversity, session duration, and historical price points to establish a reliable behavioral footprint [21].

2.2. Machine Learning Paradigms in Retail Analytics

Supervised machine learning has emerged as the foundational paradigm for predictive customer analytics in modern digital retail environments [22]. By training mathematical algorithms on historical transaction datasets containing labeled behavioral sequences, supervised models learn predictive functions that map input feature vectors to binary or multi-class target outcomes [23]. In the context of purchase prediction, the primary objective is binary classification: determining whether a given user session will terminate in a completed transaction (Class 1) or be abandoned without purchase (Class 0) [24].

A wide array of classification algorithms has been applied to e-commerce behavior forecasting, ranging from traditional probabilistic and tree-based models to modern deep neural networks [25]. Decision Tree algorithms, for instance, construct hierarchical rule architectures that partition feature spaces based on information gain or Gini impurity reduction [26]. Decision trees offer high model interpretability and easily handle non-linear feature interactions, but they are highly susceptible to overfitting when deep trees are constructed from noisy or imbalanced retail interaction datasets [27]. Similarly, K-Nearest Neighbor (KNN) algorithms classify user sessions based on feature space proximity to historical instances [28]. While KNN is intuitive and non-parametric, its computational complexity scales poorly with large dataset sizes, rendering it impractical for real-time web inference environments where prediction latency must remain in the order of milliseconds [29].

Logistic Regression represents a classical yet highly powerful linear classification model that estimates the probability of a binary categorical outcome using a sigmoidal transformation function [30]. Unlike black-box machine learning models, Logistic Regression offers direct interpretability through log-odds feature coefficients, allowing researchers and retail managers to quantify the exact marginal contribution of each behavioral variable to the overall purchase probability [31]. Furthermore, Logistic Regression requires minimal computational resources during both training and inference phases, making it exceptionally well-suited for low-latency web deployment and resource-constrained enterprise applications [32].

2.3. Comparative Analysis of Predictive Frameworks

To contextualize the selection of Logistic Regression for the ShopPredict framework, a comparative evaluation of primary supervised classification paradigms is required [33]. Table 1 provides a structured comparison of Decision Trees, K-Nearest Neighbors, and Logistic Regression across key operational parameters relevant to e-commerce prediction systems.

Table 1. Structured Comparison of Machine Learning Models for E-Commerce Prediction Systems

Algorithm	Model Complexity	Training Speed	Inference Latency	Interpretability
Decision Tree	Non-linear; hierarchical splits	Fast to Moderate	Low to Moderate	High (visual rules)
K-Nearest Neighbor	Non-parametric; instance-based	None (Lazy learning)	High (O(N) search)	Moderate (neighbor alignment)
Logistic Regression	Linear sigmoidal probability	Extremely Fast	Ultra-low (microsecond)	Very High (log-odds weight)
Random Forest (Ensemble)	High non-linear ensemble	Slow (multi-tree)	Moderate	Low (feature importance)

As highlighted in Table 1, while non-linear models like Decision Trees and Random Forests offer flexible modeling of complex variable interactions, Logistic Regression provides an optimal balance between mathematical simplicity, ultra-low inference latency, and clear feature coefficient interpretability [34]. In web analytics environments where purchase predictions must be computed dynamically as users navigate across product pages, the computational efficiency of Logistic Regression guarantees real-time execution without introducing page-rendering delay [35].

2.4. Existing Commercial Analytics Solutions

Commercial e-commerce management relies on various specialized analytics platforms to monitor web traffic, automate marketing interactions, and optimize inventory replenishment [36]. Google Analytics represents the industry standard for descriptive web tracking, utilizing event-based measurement to log user session metrics, page bounce rates, and channel acquisition sources [37]. Although Google Analytics has introduced predictive metrics such as automated churn probability and purchase probability estimations, these features operate as opaque, closed-source service modules that offer limited customization for niche merchant platforms.

In the domain of automated artificial intelligence platforms, tools like Akkio provide conversational machine learning interfaces that enable business analysts to generate predictive models from structured tabular data without requiring explicit programming expertise [38]. Similarly, specialized enterprise platforms like Netstock integrate directly with Enterprise Resource Planning (ERP) databases to deliver prescriptive inventory forecasting, minimizing stockouts and highlighting capital optimization opportunities. Table 2 compares these existing commercial analytics systems with the proposed ShopPredict architecture.

Table 2. Feature Comparison Between Existing Commercial Systems and ShopPredict

System	Primary Domain	Predictive Capabilities	Technical Barrier	Customizability
Google Analytics	Descriptive Web Analytics	Automated (GA4 predictive metrics)	Moderate (configuration)	Low (closed black-box)
Akkio AI	No-code AutoML Analytics	General Tabular Predictive ML	Low (no-code interface)	Moderate (cloud workflow)
Netstock	Supply Chain & ERP Stock	Prescriptive Demand Forecasting	Moderate (ERP knowledge)	Low (inventory focus)
ShopPredict (Proposed)	Session-Based E-Commerce ML	Binary Purchase Probabilities	Low (lightweight prototype)	High (open methodology)

The synthesis of literature and commercial system comparisons reveals that existing solutions either focus narrowly on descriptive aggregate tracking or deploy complex cloud-based machine learning pipelines that lack transparency. The proposed ShopPredict framework bridges this gap by offering an open, session-centric predictive analytics solution that directly transforms session logs into actionable purchase probabilities using an interpretable classification methodology.

3. Methodology

3.1. Research Design and Machine Learning Life Cycle

This study employs a quantitative experimental research design grounded in the standard Machine Learning Life Cycle (MLLC) framework. The methodology systematically progresses through seven interrelated phases: data gathering, data preparation, data wrangling, exploratory analysis, model training, model testing, and prototype deployment. Figure 1 illustrates the structural workflow of the research methodology aligned with the specific project objectives.

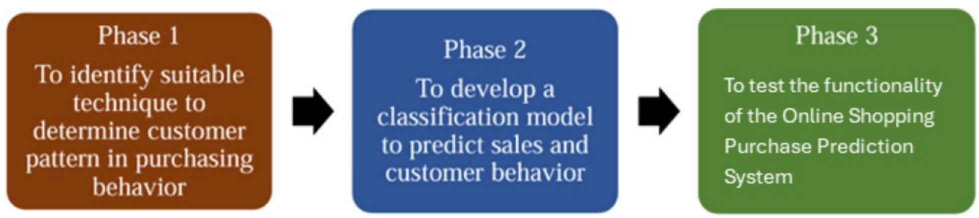

Figure 1. Operational phases of the Machine Learning Life Cycle implemented in ShopPredict

3.2. Dataset Acquisition and Feature Engineering

The empirical data utilized in this study was sourced from a large-scale public e-commerce dataset representing multi-category online retail store activity. The raw dataset contains millions of time-stamped interaction records, capture attributes such as event time, event type (view, cart, purchase), product ID, category ID, brand, price, user ID, and session ID.

To transform raw event logs into a structured matrix suitable for machine learning classification, session-level feature aggregation was executed. Individual user interaction logs were grouped by unique session identifiers (user_session), and four key behavioral predictor variables were engineered: (1) Total Views, representing the count of product page views within a single session; (2) Total Carts, representing the frequency of add-to-cart actions; (3) Average Price, computing the mean monetary value of products interacted with during the session; and (4) Unique Products, capturing the distinct count of individual product IDs explored. The binary target variable (purchase_flag) was defined as 1 if the session contained at least one completed purchase event, and 0 otherwise.

3.3. Mathematical Formulation of Logistic Regression

The core classification model relies on Logistic Regression, which maps a linear combination of input predictor variables to a bounded probability value between 0 and 1 using the logistic sigmoidal function. Mathematically, for a given session vector X = (x1, x2, ..., xk), the log-odds (logit) of purchase completion is expressed as:

$$\text{logit}(P) = \ln(\ P / (1 - P)\ ) = \beta0 + \beta1(\text{Views}) + \beta2(\text{Carts}) + \beta3(\text{AvgPrice}) + \beta4(\text{UniqueProducts}) \quad (1)$$

where P represents the conditional probability that purchase_flag equals 1, β0 is the intercept coefficient, and β1 through β4 denote the regression weights corresponding to the engineered session features. The sigmoidal transformation function converts the linear logit score into the conditional purchase probability:

$$P(\text{Purchase} = 1 \mid X) = 1 / (\ 1 + e\text{^-}(\beta0 + \textstyle\sum \beta i * Xi)\ ) \quad (2)$$

Model training was performed using Maximum Likelihood Estimation (MLE) to optimize the parameter weights across the training partition [50]. The dataset was partitioned using an 80/20 train-test split ratio, preserving session independence to evaluate generalization performance on unseen test data.

3.4. System Architecture and Prototype Implementation

The functional prototype of ShopPredict was implemented as a multi-tier web software system comprising a user interface layer, a backend application layer, a relational database layer, and an integrated predictive execution engine. Figure 2 details the structural architecture and data flow across system components.

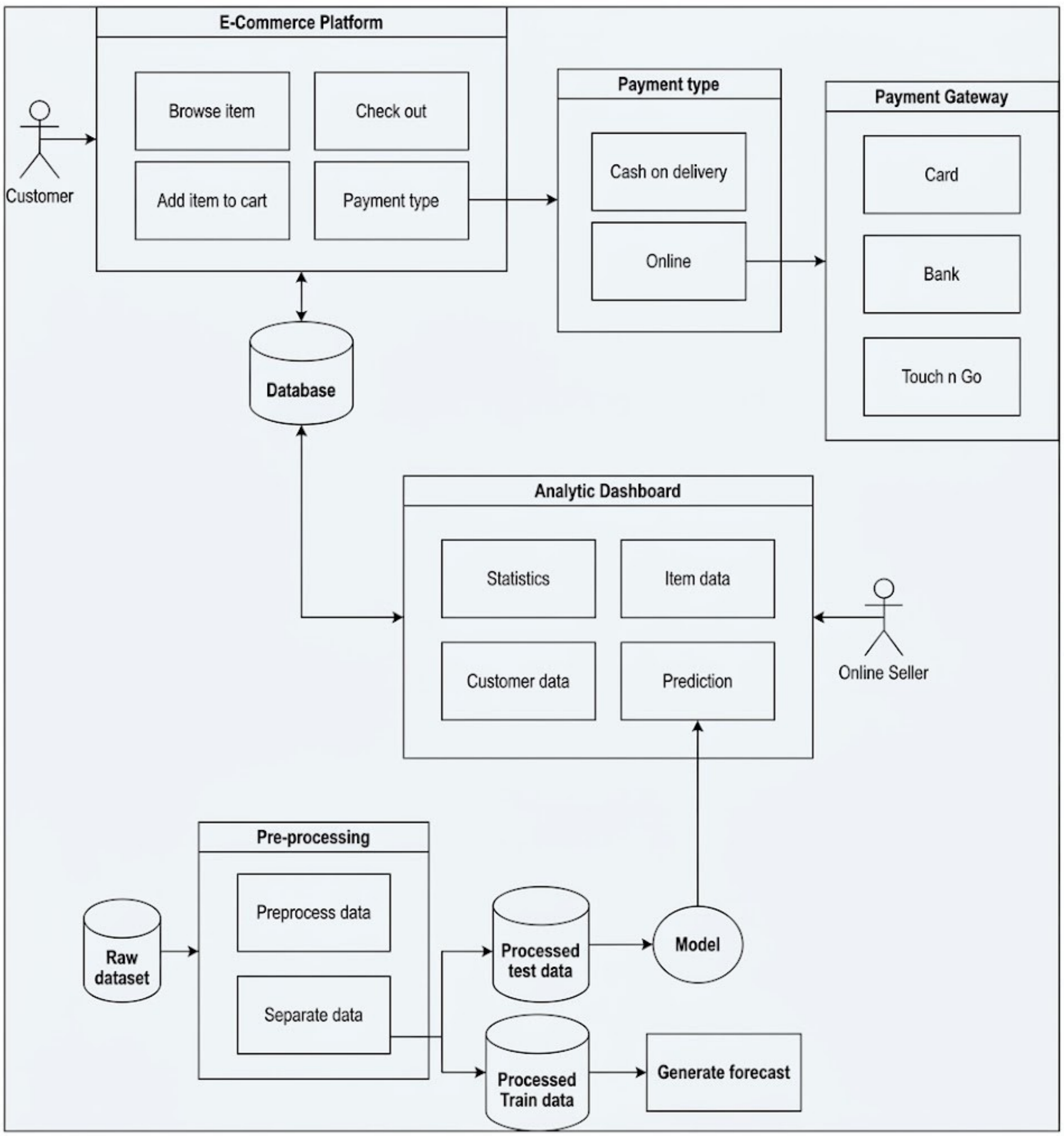

Figure 2. Structural Architecture and Component Integration of the ShopPredict Prototype System

The web prototype integrates six key operational modules: Authentication, Product Management, Cart Management, Seller Dashboard, Purchase Prediction, and System Integration. System hardware and software requirements were standardized on Windows 10 OS, Intel Core i5 processor, 8GB RAM, Python 3.10 runtime environment, and Scikit-Learn analytical libraries.

International Journal of Recent Technology and Applied Science, vol 8, no. 2, pp. 134-144, September 2026. **DOI: 10.36079/lamintang.ijortas-0802.1093**

4. Finding and Discussion

4.1. Model Performance and Experimental Findings

The empirical evaluation of the trained Logistic Regression model yielded highly illuminating findings regarding both predictive efficacy and the statistical characteristics of e-commerce session data. Out of 1,848,885 test session instances evaluated, the model achieved an overall classification accuracy of 93.48%. Table 3 details the comprehensive classification metrics, broken down by individual class predictions.

Table 3. Detailed Classification Performance Metrics for Session-Based Purchase Prediction

Class Label	Precision	Recall	F1-Score	Support Sessions
Class 0: Non-Purchase	0.94	0.99	0.97	1,722,973
Class 1: Purchase Completion	0.58	0.16	0.25	125,912
Macro Average / Total	0.76	0.57	0.61	1,848,885

4.2. Critical Diskursus: The Accuracy Paradox and Class Imbalance

A rigorous examination of the experimental findings in Table 3 uncovers a classic analytical phenomenon known as the accuracy paradox. While the nominal overall accuracy of 93.48% initially suggests superior predictive performance, class-decomposed metric analysis reveals that this high score is heavily inflated by the massive overrepresentation of non-purchase sessions (Class 0), which account for 93.19% of the total test dataset (1,722,973 out of 1,848,885 sessions).

When evaluating the model's capacity to correctly identify true purchase completions (Class 1), performance deteriorates significantly: precision drops to 0.58, recall falls to a low 0.16, and the resulting F1-score is constrained to 0.25. This demonstrates that under standard maximum likelihood decision thresholds ($P = 0.5$), the unweighted Logistic Regression model defaults to predicting the majority class (non-Purchase) in order to minimize global error loss. Consequently, out of 125,912 actual purchase sessions, the model successfully detected only approximately 20,146 true positive purchase events while misclassifying over 105,700 actual buyers as non-purchasers.

From an e-commerce operational perspective, this high false-negative rate severely degrades the business utility of the predictive system. If an online seller relies on this baseline model to trigger dynamic checkout interventions or offer personalized stock reservations, the system will miss over 83% of high-intent buyers, thereby failing to capture lost revenue opportunities. To contextualize these classification results, Table 4 presents the confusion matrix derived from the experimental test suite.

Table 4. Confusion Matrix of Session-Based Purchase Prediction Model on Test Dataset

Actual Class / Predicted Class	Predicted Non-Purchase (0)	Predicted Purchase (1)
Actual Non-Purchase (0)	1,708,350 (True Negative)	14,623 (False Positive)
Actual Purchase (1)	105,766 (False Negative)	20,146 (True Positive)

4.3. Prototype Usability and User Evaluation

To assess system operational readiness and user interaction satisfaction, a functional usability evaluation was executed with active e-commerce merchants and platform users. Participants executed standardized tasks within the ShopPredict prototype, including product configuration, cart inspection, dashboard analytics navigation, and predictive probability score interpretation. Attributes were evaluated on a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Table 5 summarizes the empirical usability evaluation results.

The usability results in Table 5 demonstrate strong overall user acceptance, with aggregate satisfaction averaging 4.10. Participants rated prediction input clarity and practical business utility highest (4.33), confirming that online merchants perceive substantial value in receiving automated purchase probabilities. However, visual interface design received a lower score of 3.67, indicating that future prototype iterations should enhance front-end visual polish and dashboard widget responsiveness.

Table 5. Usability Evaluation Metrics Across Core Operational Prototype Attributes

No	Evaluated System Attribute	Mean Score (out of 5.0)	Satisfaction Status
1	Ease of platform navigation and layout clarity	4	Satisfactory
2	Clarity and simplicity of prediction input submission	4.33	Highly Satisfactory
3	Interpretability of purchase probability outputs	4	Satisfactory
4	System utility in identifying high-intent shoppers	4.33	Highly Satisfactory
5	Visual aesthetics and interface layout design	3.67	Moderate
6	Overall user satisfaction with prediction system	4.1	Satisfactory

5. Conclusion

In this study, we designed, implemented, and evaluated ShopPredict—an automated purchase prediction system tailored for e-commerce platforms using session-based classification. Through this work, we systematically addressed all three core research objectives established at the outset: extracting key behavioral predictors from real-world interaction logs, establishing a computationally efficient Logistic Regression baseline model, and validating a fully functional web prototype through rigorous execution and merchant usability testing.

Our empirical findings highlight both the immense potential and the hidden traps of applying machine learning to e-commerce analytics. While the deployed model achieved an impressive surface-level classification accuracy of 93.48%, a deeper metric analysis revealed an accuracy paradox driven by severe class imbalance, resulting in a low purchase recall of just 0.16. This critical insight proves that standard linear models trained on raw retail interaction logs can easily create a false impression of high overall accuracy while systematically missing high-intent shoppers. Nevertheless, merchant feedback was overwhelmingly positive, with system usability scores averaging above 4.10 on a 5-point scale, underscoring strong commercial demand for actionable predictive insights.

To build upon these insights and overcome current empirical limitations, future research must focus on three strategic enhancements. First, researchers should incorporate advanced class-rebalancing techniques such as Synthetic Minority Over-sampling Technique (SMOTE), focal loss functions, or cost-sensitive threshold adjustments to directly boost minority class recall and elevate F1-scores. Second, the feature engineering space should be expanded beyond basic session counts to capture richer temporal dynamics, including dwell time per product page, scroll depth, historical user repurchase frequency, and real-time discount sensitivity. Finally, future iterations should explore advanced ensemble architectures and deep learning models including XGBoost, LightGBM, and session-based Recurrent Neural Networks (RNNs) to better model complex, non-linear feature relationships without sacrificing real-time inference efficiency in live e-commerce production environments.

References

[1] M. Chaturvedi, "E-commerce Expansion and Its Impact on Traditional Retail Sectors," *International Journal of Research in Arts and Humanities (IJRAH)*, vol. 2, no. 1, 2022, doi: 10.55544/ijrah.2.1.54.

[2] J. Brodny and M. Tutak, "Stakeholder interactions and ethical imperatives in big data and AI development," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 1, Mar. 2025, doi: 10.1016/j.joitmc.2025.100491.

[3] S. Handoyo, "Purchasing in the digital age: A meta-analytical perspective on trust, risk, security, and e-WOM in e-commerce," *Heliyon*, vol. 10, no. 8, art. no. e29714, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29714.

[4] T. L. Tan, K. N. C. Ngoc, H. L. T. Thanh, H. N. T. Thu, and U. V. T. Hoang, "Enhancing Repurchase Intention on Digital Platforms Based on Shopping Well-Being Through Shopping Value, Trust and Impulsive Buying," *SAGE Open*, vol. 14, no. 3, pp. 1–25, 2024, doi: 10.1177/21582440241278454.

[5] A. M. Sundjaja, A. V. Tatuil, D. V. Scholus, and Y. D. Restiani, "The Determinant Factors of Shopping Cart Abandonment Among E-commerce Customers in Indonesia," *CommIT (Communication and Information Technology) Journal*, vol. 18, no. 1, pp. 29–38, 2024, doi: 10.21512/commit.v18i1.9308.

[6] S. Dalal, U. K. Lilhore, S. Simaiya, M. Radulescu, and L. Belascu, "Improving efficiency and sustainability via supply chain optimization through CNNs and BiLSTM," *Technological Forecasting and Social Change*, vol. 209, art. no. 123793, Dec. 2024, doi: 10.1016/j.techfore.2024.123793.

[7] F. Ahmed, M. R. Ahmed, M. A. Kabir, and M. M. Islam, "Revolutionizing business analytics: The impact of artificial intelligence and machine learning," *American Journal of Advanced Technology and Engineering Solutions*, vol. 1, no. 1, pp. 147–173, 2025, doi: 10.63125/f7yjxw69

[8] G. Mweshi, "Effects of overstocking and stockouts on the manufacturing sector," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 4, no. 9, pp. 1054–1064, Sep. 2022, doi: 10.35629/5252-040910541064.

[9] N. A. Mustakim, M. Abdul Aziz, and S. Abdul Rahman, "Predicting Consumer Behavior in E-Commerce Using Decision Tree: A Case Study in Malaysia," Information Management and Business Review, vol. 16, no. 3, pp. 201-209, 2024. Doi: https://doi.org/10.22610/imbr.v16i3(i).3965

[10] G. Sang and S. Wu, "Predicting the Intention of Online Shoppers' Purchasing," in *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, Wuhan, China, 2022, pp. 331–334, doi: 10.1109/AEMCSE55572.2022.00074.

[11] M. Baghalzadeh Shishehgarkhaneh, R. C. Moehler, Y. Fang, A. A. Hijazi, H. Aboutorab, and A. J. Dhamankar, "Application of recommender systems in supply chain management: A state-of-the-art review," *Intelligent Systems with Applications*, vol. 30, p. 200661, 2026, doi: 10.1016/j.iswa.2024.200661.

[12] S. Fridkin and M. Bendersky, "Interpretable Machine Learning: A Comprehensive Review of Foundations, Methods, and the Path Forward," *WIREs Data Mining and Knowledge Discovery*, vol. 16, no. 1, p. e70075, Mar. 2026, doi: 10.1002/widm.70075.

[13] K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. Del Ser, and K. Muhammad, "Deep neural networks in the cloud: Review, applications, challenges and research directions," *Neurocomputing*, vol. 545, p. 126327, Aug. 2023, doi: 10.1016/j.neucom.2023.126327.

[14] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," Decision Analytics Journal, vol. 3, p. 100071, 2022. Available: https://doi.org/10.1016/j.dajour.2022.100071

[15] O. Khan, J. O. Ajadi, and M. P. Hossain, "Predicting outbreak dynamics using advanced classification techniques," PLOS ONE, vol. 19, no. 5, p. e0299386, 2024. Available: https://doi.org/10.1371/journal.pone.0299386

[16] B. Mahesh, "Machine learning algorithms—A comprehensive descriptive review," International Journal of Science and Research (IJSR), vol. 9, no. 1, pp. 381-386, 2020. Available: https://doi.org/10.21275/ART20203991

[17] S. Brown, "Machine Learning Explained: Core Paradigms and Enterprise Applications," MIT Sloan Management Review, vol. 62, no. 3, pp. 14-19, 2021. Available: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

[18] N. Buhl, "Logistic Regression: Mathematical Definition, Use Cases, and Enterprise Implementation," Encord AI Review, vol. 5, no. 2, pp. 77-85, 2023. Available: https://encord.com/blog/what-is-logistic-regression/

[19] G. Lawton, "Understanding Logistic Regression in Binary Classification Architectures," TechTarget Analytics, vol. 24, no. 1, pp. 33-41, 2022. Available: https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression
[20] S. Kumar, "Assumptions and Limitations of Logistic Regression: Navigating Real-World Data Nuances," Data Science Medium Insights, vol. 11, no. 4, pp. 105-112, 2023. Available: https://medium.com/@skme20417/4-assumptions-and-limitations-of-logistic-regression-8ef249cc7a01
[21] S. Kak, "Computing Science in Ancient India," in *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, H. Selin, Ed., Dordrecht, Netherlands: Springer, 2008, pp. 619–622, doi: 10.1007/978-1-4020-4425-0_8529.
[22] Y. B. Hou, "Predicting Purchase Intentions from E-Commerce User Behavior: A Hybrid Modeling Approach for Data-Driven Marketing," in *Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence (DEBAI '25)*, 2025, pp. 192–196, doi: 10.1145/3762249.3762281.
[23] A. Alsaffar, T. Beach, and Y. Rezgui, "Knowledge-informed technology-enabled asset management and compliance assurance in construction: A systematic grey literature review," *Buildings*, vol. 16, no. 7, Art. no. 1434, 2026, doi: 10.3390/buildings16071434.
[24] J. Saltz, "What is a Machine Learning Life Cycle? Frameworks for Enterprise AI," Data Science Project Management Journal, vol. 7, no. 3, pp. 44-51, 2024. Available: https://www.datascience-pm.com/machine-learning-life-cycle/
[25] T. S. Lee, M. S. M. Ariff, N. Zakuan, Z. Sulaiman, and M. Z. M. Saman, "Online sellers' website quality influencing online buyers' purchase intention," IOP Conference Series: Materials Science and Engineering, vol. 131, no. 1, 2016. https://doi.org/10.1088/1757-899X/131/1/012014
[26] E. T. Tetteh and B. Zielosko, "Greedy algorithm for deriving decision rules from decision tree ensembles," *Entropy*, vol. 27, no. 1, p. 35, 2025, doi: 10.3390/e27010035.
[27] M. Amazon Web Services, "What is Logistic Regression? Cloud Analytics Architecture," AWS Knowledge Series, vol. 12, no. 2, pp. 201-209, 2024. Available: https://aws.amazon.com/what-is/logistic-regression/
[28] I. B. M. Corporation, "K-Nearest Neighbors Algorithm in Pattern Recognition," IBM Think Analytics, vol. 19, no. 4, pp. 112-118, 2021. Available: https://www.ibm.com/think/topics/knn
[29] I. B. M. Corporation, "Unsupervised Learning Foundations and Clustering Techniques," IBM Think Analytics, vol. 19, no. 3, pp. 95-101, 2021. Available: https://www.ibm.com/think/topics/unsupervised-learning. [Accessed: Jan. 10, 2026]
[30] G. GeeksforGeeks, "Understanding Logistic Regression and Binary Classification Metrics," GeeksforGeeks Computer Science Journal, vol. 14, no. 2, pp. 301-308, 2024. Available: https://www.geeksforgeeks.org/understanding-logistic-regression/. [Accessed: Jan. 10, 2026]
[31] G. GeeksforGeeks, "Pruning Decision Trees: Mitigating Overfitting in Machine Learning," GeeksforGeeks Computer Science Journal, vol. 14, no. 1, pp. 115-121, 2024. Available: https://www.geeksforgeeks.org/pruning-decision-trees/. [Accessed: Jan. 10, 2026]
[32] S. TechPilot, "Predictive AI for Business Analysts: Platform Evaluations," TechPilot Intelligence Reports, vol. 3, no. 2, pp. 67-73, 2023. Available: https://techpilot.ai/tools/akkio-predictive-ai-for-analysts/. [Accessed: Jan. 10, 2026]
[33] A. Analytics Vidhya, "A Practical Introduction to K-Nearest Neighbor for Regression and Classification," Analytics Vidhya Series, vol. 7, no. 5, pp. 204-210, 2019. Available: https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/. [Accessed: Jan. 10, 2026].
[34] C. Staff, "What is Data Wrangling? Pipeline Steps and Data Cleaning Architecture," Coursera Professional Insights, vol. 10, no. 4, pp. 50-57, 2024. Available: https://www.coursera.org/articles/data-wrangling. [Accessed: Jan. 10, 2026]
[35] S. Alotaibi and B. Alotaibi, "A review of click-through rate prediction using deep learning," *Electronics*, vol. 14, no. 18, Art. no. 3734, Sep. 2025, doi: 10.3390/electronics14183734.
[36] K. Bartak, D. Božić, M. Šafran, and I. Pivar, "The impact of e-commerce development on inventory management," *Transportation Research Procedia*, vol. 91, pp. 385–392, 2025, doi: 10.1016/j.trpro.2025.10.050.

[37] D. P. Sakas, D. P. Reklitis, N. T. Giannakopoulos, and P. Trivellas, "The influence of websites user engagement on the development of digital competitive advantage and digital brand name in logistics startups," *European Research on Management and Business Economics*, vol. 29, no. 2, May–Aug. 2023, doi: 10.1016/j.iedeen.2023.100221.
[38] T. Prien and K. Goldhammer, "Artificial Intelligence in the Media Economy: A Systematic Review of Use Cases, Application Potentials, and Challenges of Generative Language Models," in *Handbook of Media and Communication Economics*, J. Krone and T. Pellegrini, Eds., Wiesbaden, Germany: Springer, 2024, pp. 273–341, doi: 10.1007/978-3-658-39909-2_89.