IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CONSUMER BEHAVIOUR ANALYSIS AND PREDICTION IN E-COMMERCE USING MACHINE LEARNING

Jiya Patel
Student
Department of DS
Drs. Kiran & Pallavi Patel Global University,
Vadodara, Gujarat, India
jiyapatel1108@gmail.com

Vruti Parikh
Assistant Professor
Department of CSE
Drs. Kiran & Pallavi Patel Global University,
Vadodara, Gujarat, India
vrutiparikh.cse.kset@kpgu.ac.in

Abstract **-** The rapid growth of e-commerce platforms has generated massive volumes of consumer interaction data, including clickstreams, browsing histories, transaction logs, and online reviews. Accurately analyzing this data to understand consumer behaviour and predict purchase intention is crucial for improving recommendation systems, customer retention, and revenue optimization. Traditional statistical approaches struggle to model the non-linear, high-dimensional, and temporal nature of consumer behaviour data. Machine learning (ML) and deep learning (DL) techniques such as Random Forest, Gradient Boosting, LightGBM, neural networks, transformers, and graph neural networks—have emerged as powerful data-driven solutions.

This review systematically surveys recent research (2020–2025) on ML-based consumer behaviour analysis and purchase intention prediction in e-commerce. We examine datasets, modeling techniques, performance metrics, and key findings across representative studies. The review highlights that advanced models frequently achieve high predictive performance (often exceeding 90% accuracy or AUC). However, challenges related to generalization, scalability, real-time deployment, and interpretability remain. This paper identifies critical research gaps and outlines future research directions toward robust, scalable, and explainable e-commerce intelligence systems.

Keywords **-** ***Consumer Behaviour, Purchase Intention, E-commerce, Machine Learning, Deep Learning, Clickstream Analytics.***

I. INTRODUCTION

E-commerce platforms such as Amazon, Flipkart, Alibaba, and eBay have fundamentally transformed the retail ecosystem by enabling personalized shopping experiences and data-driven marketing strategies. These platforms continuously collect large-scale consumer behavioural data, including clickstream logs, session sequences, search queries, purchases, ratings, and textual reviews. This data provides valuable insights into consumer preferences, intent, and decision-making processes.

Predicting purchase intention is a core problem in e-commerce analytics, directly impacting recommendation systems, targeted advertising, inventory management, and customer relationship management. Traditional rule-based and statistical models often fail to capture complex non-linear relationships and evolving consumer preferences [1]. In contrast, machine learning techniques can automatically learn hidden patterns from large-scale, high-dimensional data.

Recent studies demonstrate that supervised learning models such as Random Forest, XGBoost, CatBoost, and LightGBM effectively predict purchase behaviour using clickstream and transaction data [3]. Moreover,

deep learning models—including recurrent neural networks, attention mechanisms, transformers, and graph neural networks—have shown superior performance in modeling sequential and relational consumer interactions [8]. Given the rapid evolution of this field, a systematic review is essential to consolidate recent advancements, identify limitations, and guide future research.

II. LITERATURE REVIEW

A. Machine Learning Approaches for Consumer Behaviour Analysis

Supervised machine learning models are widely used for purchase intention prediction due to their robustness and interpretability. Random Forest, XGBoost, LightGBM, and CatBoost models have consistently achieved high accuracy on clickstream and transaction datasets [11]. These ensemble methods effectively capture non-linear interactions among behavioural features such as session duration, page views, revisit frequency, and product categories.

Unsupervised learning techniques, including K-means and hierarchical clustering, are commonly applied for customer segmentation [5]. By grouping users based on behavioural similarity, these methods support targeted marketing and personalized recommendations. Several studies report improved conversion rates when clustering is integrated with predictive models.

B. Deep Learning and Advanced Models

Deep learning approaches have demonstrated superior performance in modeling sequential and temporal consumer behaviour. Recurrent neural networks (RNNs), LSTMs, and GRUs effectively learn session-level dependencies in clickstream data [12]. Attention mechanisms further enhance performance by identifying key interactions that influence purchase decisions.

Recent research increasingly adopts Transformer-based architectures and Graph Neural Networks (GNNs). Transformers capture long-range dependencies in user sessions [8], while GNNs model complex user–item and item–item relationships [13]. Multimodal approaches that integrate behavioural data with textual sentiment from reviews have shown further improvements in prediction accuracy [6]. Reinforcement learning-based models enable adaptive recommendation strategies by learning from continuous user feedback [15].

C. Comparative literature review

The comparative literature review reveals significant advancements in consumer behaviour analysis and purchase intention prediction using machine learning and deep learning techniques.

Zhou and Hudin (2024) employed real clickstream logs and proposed an Attention-based Graph Neural Network model, achieving an F1-score of 96%, which effectively captured both temporal dependencies and relational user–item data; however, the approach was computationally intensive and less suitable for real-time applications [4].

Sakalauskas and Kriksciuniene (2024) focused on advertising clickstream data and implemented a multi-step classifier, reporting an accuracy of approximately 89%. Their method proved effective for customer segmentation and high-value targeting, but its applicability was limited due to the restricted nature of advertisement datasets [5].

Gooljar et al. (2024) combined sentiment analysis from Amazon reviews with behavioral data using BERT and classical machine learning models, resulting in nearly a 92% performance uplift; however, this approach required rich textual data, limiting its use in data-scarce environments [6].

Park et al. (2024) applied sequential neural networks to e-commerce session data and achieved about 94% accuracy by successfully capturing co-purchase patterns, though the method required large-scale session logs for optimal performance [7].

Ma et al. (2024) integrated CNN and BiLSTM architectures using review and session data, achieving around 93% accuracy by combining semantic and sequential features, although the approach involved high computational requirements [12].

Chachra (2024) explored large-scale commerce logs using a combination of gradient boosting and neural networks, achieving an AUC of 0.94, but noted that real-world deployment performance varied [6].

Mallapragada et al. (2025) proposed a multimodal Transformer-based framework that achieved an AUC of approximately 97% by integrating diverse e-commerce data sources, though it required large labeled datasets [8].

Shi et al. (2025) introduced persona embeddings from user session logs and reported around 89% accuracy, though the static nature of personas remained a limitation [10].

Kuang et al. (2025) combined clustering with GRU and attention mechanisms for retail time-series data, achieving 85–90% accuracy but faced parameter sensitivity issues [9].

Aditi M. Jain (2025) proposed a DQN-inspired LSTM-based framework on 885k sessions, achieving 88%

accuracy and an AUC of 0.88, while highlighting the requirement for large-scale behavioral logs [15].

III. IDENTIFIED RESEARCH GAPS

Despite significant progress, several challenges remain. Many models are trained on platform-specific datasets and exhibit limited generalization to other e-commerce environments [1]. Most studies rely primarily on clickstream data, underutilizing multimodal signals such as textual reviews, images, and contextual information.

High-accuracy deep learning models often suffer from scalability and real-time deployment issues due to high computational complexity [9]. Additionally, the lack of interpretability in complex models restricts trust and adoption in business decision-making. Dynamic persona modeling and reinforcement learning-based personalization remain underexplored.

IV. POSSIBLE EXTENSIONS

Future research should focus on multimodal learning frameworks that integrate behavioural, textual, and visual data using transformer architectures [8]. Lightweight and low-latency models such as optimized gradient boosting and hybrid ML–DL systems should be explored for real-time deployment [11].

Scalable graph learning techniques can further enhance relational behavior modeling [4]. Reinforcement learning approaches offer promising directions for adaptive and personalized recommendation systems. Finally, incorporating explainable AI techniques will improve transparency and trust, facilitating real-world adoption.

V. PROPOSED METHODOLOGY

The proposed methodology introduces an integrated framework for analyzing consumer behaviour and predicting purchase intention in e-commerce platforms using machine learning and deep learning techniques. The framework is designed to address key challenges identified in existing studies, such as limited generalization, scalability constraints, and lack of interpretability. By combining both unsupervised and supervised learning approaches, the framework aims to simultaneously understand consumer segments and predict future purchasing actions.

The first stage of the framework focuses on data collection and integration from multiple heterogeneous sources, including clickstream logs, session-level interactions, transaction records, and product metadata. These data sources are synchronized to construct a unified view of the consumer journey, enabling the tracking of user interactions from initial site entry to final purchase or exit. This comprehensive data foundation ensures that both temporal and behavioural patterns are preserved for subsequent analysis.

Following data integration, preprocessing and feature engineering are performed to improve data quality and enhance predictive capability. This stage involves handling missing values, removing noise, and normalizing numerical attributes. High-level behavioural features such as session duration, navigation depth, browsing frequency, and cart-to-view ratios are derived to capture user engagement and intent more effectively than raw click data.

In the next phase, unsupervised learning techniques such as K-Means or hierarchical clustering are applied to segment users into distinct behavioural groups. This segmentation helps identify different shopping personas, such as high-intent buyers, exploratory users, and price-sensitive customers. These behavioural cohorts provide valuable contextual information that enhances the accuracy and relevance of downstream predictive models.

Finally, supervised machine learning and deep learning models, including tree-based algorithms and sequential architectures such as LSTM or Transformer networks, are employed to predict purchase intention. Model performance is evaluated using appropriate metrics for imbalanced e-commerce data, such as F1-score, ROC-AUC, and NDCG. The overall framework emphasizes real-time applicability, scalability, and interpretability, aiming to support personalized recommendations, targeted marketing strategies, and improved decision-making in e-commerce systems.

VI. CONCLUSION

This review has examined recent advancements in machine learning and deep learning techniques for consumer behaviour analysis and purchase intention prediction in e-commerce. The surveyed studies demonstrate that ensemble learning, sequential models, transformers, and graph neural networks significantly enhance predictive performance, often exceeding 90% accuracy or AUC.

However, challenges related to generalization, scalability, interpretability, and multimodal integration remain unresolved. Addressing these issues will be critical for developing robust, deployable, and explainable e-commerce intelligence systems. This review provides a comprehensive foundation for future research and supports the development of data-driven frameworks for personalized and intelligent e-commerce platforms.

VII. REFERENCES

[1] Z. Wen, W. Lin, and H. Liu, "Machine-learning-based approach for anonymous online customer purchase intentions using clickstream data," *Systems*, vol. 11, no. 5, p. 255, 2023.

[2] W. Wang *et al.*, "A user purchase behavior prediction method based on XGBoost," *Electronics*, vol. 12, no. 9, p. 2047, 2023.

[3] I. A. Khandokar *et al.*, "A gradient boosting classifier for purchase intention prediction of online shoppers," *Heliyon*, vol. 9, no. 11, e15163, 2023.

[4] S. Zhou and N. S. Hudin, "Advancing e-commerce user purchase prediction using time-series attention and graph neural networks," *PLOS ONE*, vol. 19, no. 4, e0299087, 2024.

[5] V. Sakalauskas and D. Kriksciuniene, "Personalized advertising in e-commerce using clickstream data to target high-value customers," *Algorithms*, vol. 17, no. 1, p. 27, 2024.

[6] V. Gooljar, T. Issa, and S. Hardin-Ramanan, "Sentiment-based predictive models for online purchases in the era of marketing 5.0: A systematic review," *Journal of Big Data*, vol. 11, p. 107, 2024.

[7] M. Park *et al.*, "Enhancing e-commerce recommendation systems with concurrent purchases," *Applied Sciences*, vol. 14, no. 16, p. 7255, 2024.

[8] S. Mallapragada *et al.*, "Multi-modality transformer for e-commerce: Inferring user purchase intention to bridge the query-product gap," *arXiv preprint* arXiv:2501.14826, 2025.

[9] Y. Kuang *et al.*, "CATS: Clustering-aggregated and time series for business customer purchase intention prediction," *arXiv preprint* arXiv:2505.13558, 2025.

[10] Y. Shi *et al.*, "You are what you bought: Generating customer personas for e-commerce applications," *arXiv preprint* arXiv:2504.17304, 2025.

[11] J. Du, "Predictions for consumer behaviour of e-commerce sales data based on the LightGBM model," in *Proc. SciTePress*, 2024, pp. 1–9.

[12] X. Ma *et al.*, "E-commerce review sentiment analysis and purchase intention prediction based on deep learning technology," *Journal of Organizational and End User Computing*, vol. 36, no. 1, pp. 1–18, 2024.

[13] C. Zhang *et al.*, "Multi-aspect enhanced graph neural networks for recommendation," *Neural Networks*, vol. 158, pp. 25–38, 2023.

[14] B. Guo and C. Sismeiro, "Between click and purchase: Predicting purchase decisions using clickstream data," *Journal of Marketing Research*, vol. 57, no. 2, pp. 1–18, 2020.

[15] A. M. Jain, "Predicting e-commerce purchase behavior using a DQN approach," *arXiv preprint* arXiv:2506.17543, 2025.