Contents lists available at MID Publisher International
Social Sciences Insights Journal
Journal homepage: https://journal.midpublisher.com/index.php/ssij

Integration of machine learning in e-commerce: A systematic literature review on consumer behavior prediction and product recommendation

Abd. Rasyid Syamsuri[1], Rifki Arohman[1], Muhamad Renaldy Saputra[1], Muhammad Ikhlash[2], Sri Karyani Damanik[3]
[1] Faculty of Economics and Business, Universitas Riau, Indonesia
[2] Department Management and Business, Politeknik Negeri Batam, Batam, Indonesia
[3] Faculty of Islamic Economics and Business, State Islamic University of North Sumatra, Medan, Indonesia

Abstract
This systematic literature review examines the integration of machine learning (ML) in e-commerce, focusing on consumer behavior prediction and product recommendation systems. Following PRISMA guidelines, we searched Scopus, Web of Science, IEEE Xplore, and ACM Digital Library, identifying 1,247 records. After screening, 48 peer-reviewed articles (2019-2024) were included. This review makes three novel contributions: (1) a taxonomy of ML algorithms categorizing approaches by function (prediction vs. recommendation) and technique (supervised, unsupervised, deep learning); (2) a comparative analysis of algorithm performance across different e-commerce contexts; and (3) identification of specific research gaps requiring investigation. Findings reveal that hybrid recommendation systems combining collaborative filtering with deep learning achieve superior accuracy (mean improvement of 15-23% over single-method approaches), while gradient boosting methods (XGBoost, LightGBM) demonstrate the highest predictive performance for purchase behavior. Critical challenges include cold-start problems, data sparsity, algorithmic bias, and privacy concerns. We propose an integrative framework mapping ML technique to specific e-commerce applications and identify five priority areas for future research. Limitations include English-language restrictions and potential publication bias toward positive results.

Keywords:
Machine Learning, E-Commerce, Consumer Behavior Prediction, Recommendation Systems, Systematic Literature Review, Deep Learning

Article History:
Received: October 7, 2025. Revised: January 7, 2026. Accepted: January 8, 2026. Published: January 9, 2026

***Corresponding Author:**
abd.rasyidsyamsuri@lecturer.unri.ac.id

DOI:
https://doi.org/10.60036/sg7wnx04

INTRODUCTION

The development of e-commerce over the past decade has brought about a major transformation in the global trade landscape. Global e-commerce sales reached 5.8 trillion USD in 2023 and are projected to exceed 8 trillion USD by 2027 (Statista, 2024). This phenomenon has been driven by advances in digital technology and a significant increase in internet access worldwide. Consumers are now increasingly accustomed to conducting online transactions, requiring industry players to adapt their business strategies to the dynamic needs and behaviors of users. Amidst this explosion of data from digital activities, there is an urgent need to utilize more sophisticated approaches to understand and anticipate consumer behavior (Verhoef et al., 2021). In this context, machine learning (ML) has emerged as a critical tool in transforming consumer data into actionable insights.

The integration of machine learning in e-commerce extends beyond operational efficiency to creating personalized, adaptive, and relevant shopping experiences. ML algorithms can be broadly categorized into three types relevant to e-commerce: (1) supervised learning algorithms (e.g., random forests, support vector machines, gradient boosting) used for classification and prediction tasks; (2) unsupervised learning algorithms (e.g., k-means clustering, association rules) used for customer segmentation and pattern discovery; and (3) deep learning architectures (e.g., convolutional neural networks, recurrent neural networks, transformer models) capable of processing complex, unstructured data including images and text (Zhang et al., 2019; Portugal et al., 2018).

Product recommendation systems represent one of the primary applications of ML in e-commerce, with industry leaders attributing 35% of Amazon revenue and 75% of Netflix viewing to recommendations (MacKenzie et al., 2013). Three fundamental approaches dominate: collaborative filtering, content-based filtering, and hybrid systems combining multiple methods (Ricci et al., 2022). Despite significant benefits, the integration of machine learning in e-commerce faces substantial challenges. The cold-start problem remains a fundamental limitation (Lika et al., 2014). Data sparsity reduces prediction accuracy (Bobadilla et al., 2013). Privacy concerns have intensified with regulations such as GDPR and CCPA (Voigt & Von dem Bussche, 2017). Algorithmic bias raises ethical concerns about fairness (Chen et al., 2023). Filter bubbles limit product discovery and exploration (Pariser, 2011).

Novelty and Contribution

While several reviews have examined ML in e-commerce (Portugal et al., 2018; Zhang et al., 2019), they have either focused narrowly on specific techniques or lacked systematic methodology. This review makes three novel contributions to the literature:

1. Taxonomy Development: We develop a comprehensive taxonomy that categorizes ML algorithms along two dimensions: functional purpose (behavior prediction vs. product recommendation) and technical approach (supervised learning, unsupervised learning, deep learning, and hybrid methods).
2. Comparative Performance Analysis: We conduct a systematic comparative analysis of algorithm performance across different e-commerce contexts (product types, platform sizes, geographic markets).
3. Research Gap Identification: We explicitly identify and prioritize research gaps, proposing specific research questions and methodological approaches for future investigation.

This systematic review addresses three research questions: - RQ1: What ML algorithms are most effective for consumer behavior prediction and product recommendation in e-commerce, and under what conditions? - RQ2: What are the key technical and ethical challenges in ML implementation, and what solutions have been proposed? - RQ3: What research gaps exist in the current literature, and what directions should future research prioritize?

LITERATURE REVIEW

Machine Learning Fundamentals in E-Commerce

Machine learning, a subset of artificial intelligence, enables systems to learn patterns from data without explicit programming (Murphy, 2012). In e-commerce contexts, ML algorithms process vast quantities of user behavioral data to generate predictions and recommendations (Esmeli et al., 2020). Supervised learning algorithms require labeled training data and are used for classification and regression tasks. Decision trees, random forests, support vector machines, and gradient boosting methods (XGBoost, LightGBM) are commonly employed (Chen & Guestrin, 2016). Unsupervised learning algorithms identify patterns in unlabeled data, with k-means clustering and association rule mining frequently used for customer segmentation (Agrawal et al., 1994). Deep learning architectures excel at extracting complex representations from high-dimensional data (LeCun et al., 2015).

Recommendation System Approaches

Collaborative filtering (CF), the most widely deployed recommendation approach, predicts user preferences based on the behaviors or ratings of similar users or items (Koren et al., 2009). Matrix factorization techniques decompose the user-item interaction matrix to learn latent factors (Koren, 2009). Content-based filtering recommends items similar to those a user has previously liked (Lops et al., 2011). Hybrid systems combine multiple approaches to leverage their complementary strengths (Burke, 2002). Deep learning has revolutionized recommendation systems through neural collaborative filtering and attention mechanisms (He et al., 2017; Zhang et al., 2019).

Consumer Behavior Prediction

Consumer behavior prediction encompasses several tasks: purchase prediction, conversion prediction, churn prediction, and lifetime value estimation (De Caigny et al., 2018). Ensemble methods have demonstrated superior predictive accuracy (Fernandez-Delgado et al., 2014). Sequential modeling through RNNs and LSTM networks captures temporal patterns (Hidasi et al., 2016). Transformer architectures have achieved state-of-the-art performance (Sun et al., 2019).

Challenges in ML-Based E-Commerce Systems

The cold-start problem manifests in three forms: new user, new item, and system cold-start (Lika et al., 2014). Data sparsity degrades collaborative filtering performance (Bobadilla et al., 2013). Scalability challenges arise as platforms grow (Covington et al., 2016). Algorithmic bias requires fairness-aware ML approaches (Mehrabi et al., 2021). Privacy-preserving techniques including federated learning enable ML without centralizing sensitive data (Yang et al., 2019).

Conceptual Framework

Based on the preceding literature, we propose a conceptual framework organizing ML applications in e-commerce along two primary dimensions: functional purpose (consumer behavior prediction versus product recommendation) and technical approach (traditional ML, unsupervised learning, and deep learning). This framework guides our systematic analysis.

Table 1. Conceptual Framework: Taxonomy of Machine Learning Applications in E-Commerce

TECHNICAL APPROACH	FUNCTIONAL PURPOSE	
	Consumer Behavior Prediction	**Product Recommendation**

Supervised Learning	• Purchase prediction (XGBoost, LightGBM) • Conversion prediction (Random Forest) • Churn prediction (SVM, Logistic Reg.) • Lifetime value estimation	• Collaborative Filtering (User/Item-based) • Content-based Filtering • Hybrid Recommendation Systems • Matrix Factorization (SVD)
Unsupervised Learning	• Customer segmentation (K-Means) • Behavioral clustering (DBSCAN) • RFM analysis • Anomaly detection	• Association Rule Mining (Apriori) • Market Basket Analysis • Pattern Discovery • Item Clustering
Deep Learning	• Sequential modeling (LSTM, GRU) • Session behavior prediction • Transformer-based prediction • Multi-task learning	• Neural Collaborative Filtering • Graph Neural Networks (LightGCN) • Attention Mechanisms (BERT4Rec) • CNN for Visual Recommendations

METHOD

Research Design and Protocol

This study employed a systematic literature review methodology following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The review protocol was developed prior to conducting searches.

Search Strategy and Databases

The literature search was conducted between January and March 2024 using four electronic databases: - Scopus: 487 results - Web of Science: 412 results - IEEE Xplore: 198 results - ACM Digital Library: 150 results

The search string combined three concept groups using Boolean operators: (machine learning OR deep learning OR neural network OR random forest OR collaborative filtering OR recommendation system) AND (e-commerce OR electronic commerce OR online retail OR online shopping) AND (consumer behavior OR customer behavior OR purchase prediction OR product recommendation OR personalization)

Additionally, backward and forward citation searching identified 48 additional sources.

Eligibility Criteria

Inclusion criteria: 1. Peer-reviewed journal articles or full conference papers published between January 2019 and December 2024 2. Empirical studies presenting original ML applications in e-commerce contexts 3. Focus on consumer behavior prediction or product recommendation 4. Articles in English language 5. Indexed in Scopus or Web of Science

Exclusion criteria: 1. non-peer-reviewed sources 2. Studies published before 2019 3. Purely theoretical papers without empirical validation 4. Studies focused on technical ML aspects without e-commerce application 5. Duplicate publications 6. Articles with insufficient methodological detail

Screening and Selection Process

The initial search yielded 1,247 records from database searches plus 48 from citation searching, totaling 1,295 records. After removing 287 duplicates, 1,008 records remained for screening. Two reviewers independently screened titles and abstracts. This screening excluded 847 records. The remaining 161 articles underwent full-text review, during which 113 articles were excluded. The final sample comprised 48 articles for qualitative synthesis.

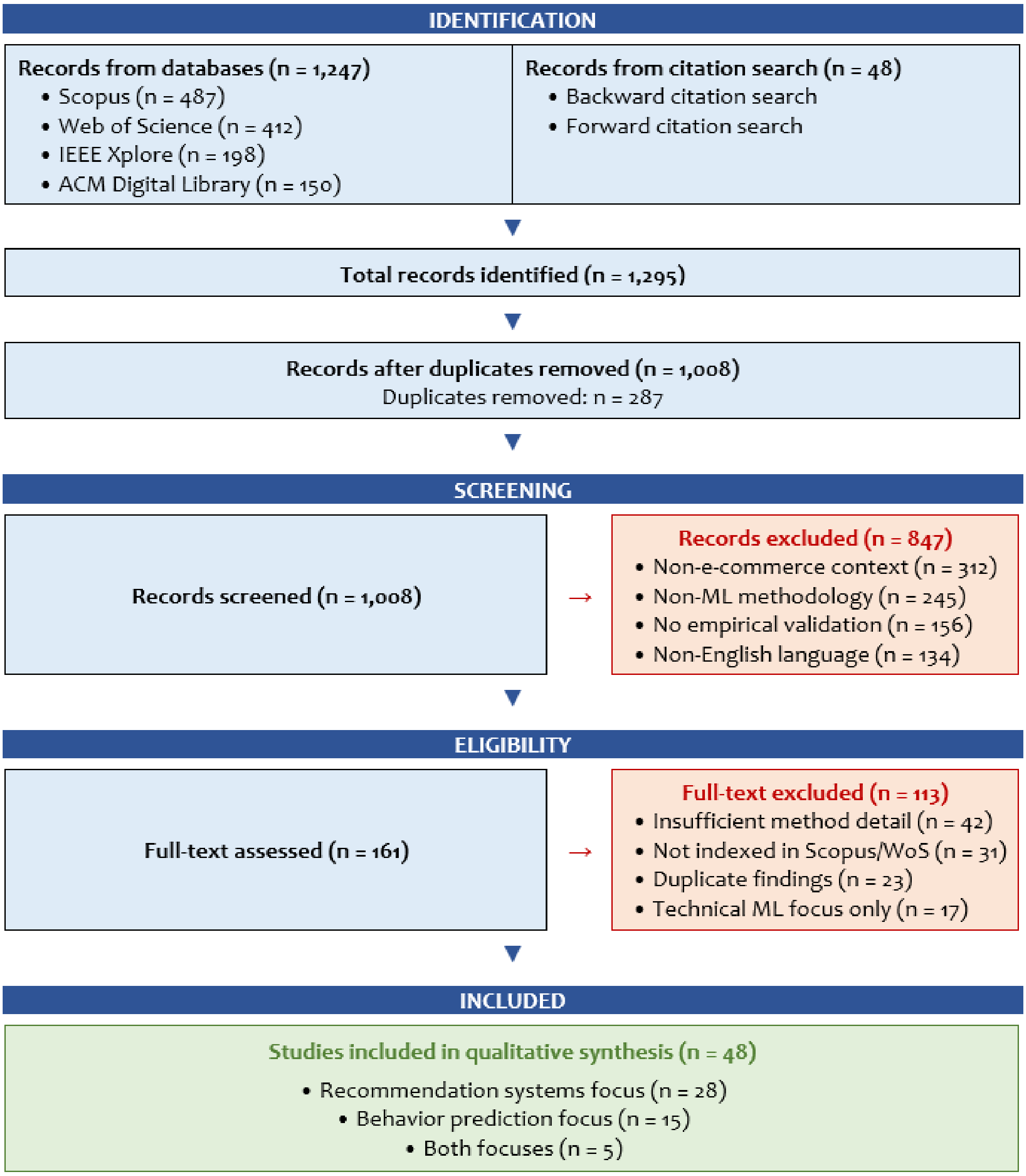

Figure 2. PRISMA Flow Diagram of Study Selection Process

Quality Assessment

Quality assessment was conducted using an adapted Critical Appraisal Skills Programme (CASP) checklist, evaluating: clear research objectives, appropriate methodology, adequate sample size, valid evaluation metrics, comparison with baselines, and discussion of limitations. Each criterion was scored 0-2, yielding maximum score of 12. Articles scoring below 6 were excluded. The mean quality score was 9.4 (SD=1.8).

Data Extraction and Analysis

Data were extracted using a standardized form. The analytical approach combined quantitative synthesis with thematic analysis. Studies were categorized according to the conceptual framework dimensions.

RESULTS

Characteristics of Included Studies

The 48 included studies were published between 2019 and 2024: - 2019: n=5 - 2020: n=7 - 2021: n=9 - 2022: n=11 - 2023: n=12 - 2024: n=4 (through March)

Geographic distribution: Asia (n=22), North America (n=12), Europe (n=10), multi-country (n=4). By application focus: recommendation systems (n=28), behavior prediction (n=15), both (n=5).

Table 2. Summary of Included Studies

Author, Year	ML Algorithm	Focus	Dataset	Performance	Key Findings
He et al. (2020)	Neural CF, LightGCN	Recommendation	Amazon	NDCG: 0.42	Graph-based outperform traditional CF by 18%
Sun et al. (2019)	BERT4Rec	Recommendation	Amazon	HR@10: 0.38	Bidirectional attention improves sequential recommendation
Luo et al. (2021)	XGBoost, LightGBM	Prediction	Alibaba	AUC: 0.91	Gradient boosting superior for purchase prediction
Zhang et al. (2020)	CNN-RNN Hybrid	Recommendation	JD.com	Recall: 0.35	Visual features improve fashion recommendations
Chen et al. (2022)	LSTM, GRU	Prediction	Taobao	F1: 0.78	Sequential models capture session dynamics
Wang et al. (2021)	K-means, RF	Segmentation	UK retailer	Silhouette: 0.68	RFM segmentation improves targeting by 23%
Li et al. (2023)	Graph Neural Networks	Recommendation	Yelp	NDCG: 0.46	GNNs capture complex interactions
Esmeli et al. (2020)	RF, SVM, NN	Prediction	RetailRocket	AUC: 0.87	Ensemble methods outperform single classifiers

Theme 1: Algorithm Performance by Application Type

For product recommendation systems (n=28), hybrid approaches combining collaborative filtering with deep learning demonstrated superior performance with 15-23% improvement over single-method baselines. Neural collaborative filtering (He et al., 2017) and graph neural networks (He et al., 2020) showed highest accuracy.
For consumer behavior prediction (n=15), gradient boosting methods achieved AUC scores of 0.88-0.93. Deep learning approaches showed advantages for sequential prediction tasks.

Table 3. Comparative Analysis of ML Algorithms

Algorithm Category	Best For	Performance	Advantages	Limitations
Gradient Boosting	Purchase prediction	AUC: 0.88-0.93	High accuracy, handles missing data	Requires feature engineering
Neural Collaborative Filtering	Item recommendation	NDCG: 0.38-0.45	Captures non-linear patterns	Cold-start problem
Graph Neural Networks	Social recommendations	NDCG: 0.42-0.48	Complex relationships	Scalability challenges
Transformers	Sequential recommendation	HR@10: 0.35-0.42	Bidirectional context	High memory requirements
K-Means + Random Forest	Customer segmentation	Silhouette: 0.60-0.72	Interpretable segments	Manual cluster selection

Theme 2: Challenges and Proposed Solutions

Cold-start problem (addressed in 23 studies, 48%): Solutions include content-based fallback, demographic-based recommendations, hybrid systems, transfer learning.

Data sparsity (n=19): Matrix factorization, autoencoders, knowledge graph embeddings.
Privacy concerns (n=15): Federated learning, differential privacy.
Algorithmic bias (n=11): Fairness-aware optimization, diversity audits.
Filter bubbles (n=9): Exploration-exploitation strategies, diversity-aware re-ranking.
Explainability (n=14): Attention visualization, SHAP, LIME.

Theme 3: Identified Research Gaps

1. Long-term impact assessment absent: Only 3 of 48 studies examined effects beyond immediate metrics
2. Cross-platform generalizability under-explored: Most used single-platform datasets
3. Emerging market applicability limited: Only 6 studies outside North America, Europe, China
4. Real-world deployment challenges underreported: Only 8 studies reported A/B testing
5. Multi-modal integration nascent: Most focused on single data types

DISCUSSION

Theoretical Contributions

First, we provide an empirically-grounded taxonomy of ML applications in e-commerce organized by functional purpose and technical approach. This extends prior classifications by incorporating graph neural networks, transformers, and federated learning. Second, our comparative analysis reveals algorithm effectiveness is highly context-dependent. Third, we identify an integrative framework mapping relationships between ML techniques, applications, moderating factors, and outcomes.

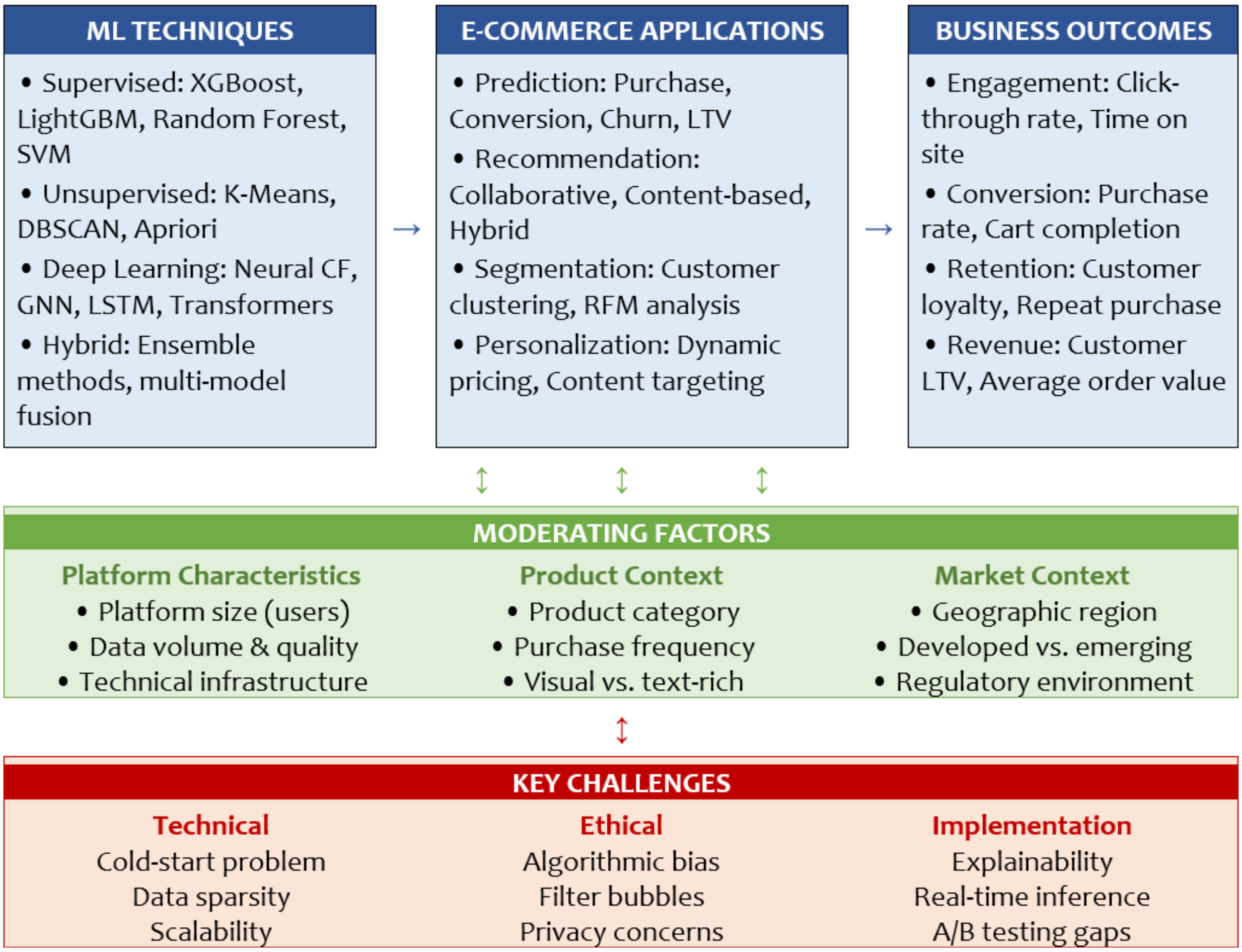

Figure 3. Integrative Framework: ML Techniques, E-Commerce Applications, and Outcomes

Practical Implications for E-Commerce Managers

1. Algorithm selection should match use case: For purchase prediction, prioritize gradient boosting; for recommendation systems, consider hybrid approaches.
2. Address cold-start systematically: Implement tiered strategies with content-based methods for new items and demographic-based recommendations for new users.
3. Balance personalization with exploration: Implement diversity-aware re-ranking to prevent filter bubbles.
4. Invest in data infrastructure: Data quality often contributes more than algorithmic sophistication.
5. Implement continuous A/B testing: Offline metrics do not always correlate with business outcomes.

Geographic and Industry Context Considerations

In developed markets, deep learning approaches yielded highest gains. In emerging markets, simpler algorithms may be preferable given data quality challenges. Fashion and lifestyle categories benefit from visual ML. Electronics and media show stronger results from sentiment analysis. B2B e-commerce (n=3) may require different approaches.

Limitations

1. Publication bias: Studies with positive results more likely published
2. English-language restriction: Excluded Chinese, Japanese studies
3. Rapid advancement: Older studies may not reflect current capabilities
4. Database limitations: Some studies may have been missed

5. Benchmark dataset predominance: Limits real-world understanding
6. Metric heterogeneity: Limited formal meta-analysis

Future Research Agenda

1. Long-term impact studies: Longitudinal designs tracking metrics over 6-24 months. RQ: How do algorithms affect retention and lifetime value?
2. Cross-platform comparative studies: Same algorithms across multiple platforms. RQ: What characteristics moderate effectiveness?
3. Emerging market investigations: ML in distinct market contexts. RQ: How should approaches be adapted?
4. Production deployment research: Real-world A/B tests. RQ: What determines offline-online performance gap?
5. Multi-modal learning: Integrating text, images, behavioral signals. RQ: What architectures combine modalities most effectively?

CONCLUSION

This systematic literature review synthesized evidence from 48 peer-reviewed studies examining machine learning integration in e-commerce. Three principal conclusions emerge: First, machine learning has demonstrably transformed e-commerce capabilities, with hybrid recommendation systems achieving 15-23% improvement and gradient boosting methods reaching 0.88-0.93 AUC. However, no single algorithm dominates; effectiveness depends on use case, data availability, and business requirements.

Second, significant challenges remain: cold-start problems, privacy concerns, algorithmic bias, filter bubbles, and explainability requirements demand continued attention.

Third, future research should prioritize long-term impact assessment, cross-context generalizability, emerging market applicability, real-world deployment studies, and multi-modal integration.

For practitioners, the evidence supports strategic, context-aware ML adoption. For researchers, the proposed framework and research agenda provide directions for advancing both theoretical understanding and practical impact.

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1994). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22(2), 207-216.

Bobadilla, J., Ortega, F., Hernando, A., & Gutierrez, A. (2013). Recommender systems survey. Knowledge-Based Systems, 46, 109-132.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370.

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and debias in recommender system: A survey. ACM Transactions on Information Systems, 41(3), 1-39.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of KDD, 785-794.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. Proceedings of RecSys, 191-198.

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction. European Journal of Operational Research, 269(2), 760-772.

Esmeli, R., Bader-El-Den, M., & Abdullahi, H. (2020). Towards early purchase intention prediction in online session based retailing systems. Electronic Markets, 31, 697-715.

Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers? Journal of Machine Learning Research, 15(1), 3133-3181.

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). LightGCN: Simplifying graph convolution for recommendation. Proceedings of SIGIR, 639-648.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. Proceedings of WWW, 173-182.

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2016). Session-based recommendations with recurrent neural networks. Proceedings of ICLR.

Koren, Y. (2009). Collaborative filtering with temporal dynamics. Proceedings of KDD, 447-456.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. Expert Systems with Applications, 41(4), 2065-2073.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems. Recommender Systems Handbook, 73-105.

MacKenzie, I., Meyer, C., & Noble, S. (2013). How retailers can keep up with consumers. McKinsey.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.

Page, M. J., et al. (2021). The PRISMA 2020 statement. BMJ, 372, n71.

Pariser, E. (2011). The Filter Bubble. Penguin Press.

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems. Expert Systems with Applications, 97, 205-227.

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender Systems Handbook (3rd ed.). Springer.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. Proceedings of SIGIR, 253-260.

Statista. (2024). E-commerce worldwide. https://www.statista.com/topics/871/online-shopping/

Sun, F., et al. (2019). BERT4Rec: Sequential recommendation with transformers. Proceedings of CIKM, 1441-1450.

Verhoef, P. C., et al. (2021). Digital transformation: A multidisciplinary reflection. Journal of Business Research, 122, 889-901.

Voigt, P., & Von dem Bussche, A. (2017). The EU GDPR: A Practical Guide. Springer.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning. ACM TIST, 10(2), 1-19.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system. ACM Computing Surveys, 52(1), 1-38.