

# Exploring the Influence of Machine Learning in e-Commerce: A Systematic and Bibliometric Review

Javier Gamboa-Cruzado<sup>1</sup>, Thalia Mosqueira-Cerda<sup>2</sup>, Anibal Torre Camones<sup>1</sup>,  
Roberto Quispe Mendoza<sup>2</sup>, Angel F. Navarro Raymundo<sup>3</sup>,  
Jesús Jiménez García<sup>4</sup>, Blanca Cecilia López-Ramírez<sup>5,\*</sup>

<sup>1</sup> Universidad Nacional Mayor de San Marcos,  
Peru

<sup>2</sup> Universidad Nacional de Trujillo,  
Peru

<sup>3</sup> Universidad Nacional Tecnológica de Lima Sur,  
Peru

<sup>4</sup> Universidad Señor de Sipán,  
Peru

<sup>5</sup> Tecnológico Nacional de México/IT de Roque, Celaya,  
Mexico

jgamboac@unmsm.edu.pe, 2019002754@unfv.edu.pe,  
aatorrec@unac.edu.pe, rquispe@unitru.edu.pe,  
anavarro@untels.edu.pe, jjimenezg@uss.edu.pe, bllopez@itroque.edu.mx

**Abstract.** The use of Machine Learning (ML) in e-commerce has revolutionized key processes such as service personalization, dynamic pricing optimization, and sales forecasting, generating a direct impact on both operational efficiency and user experience quality. The objective of this paper is to rigorously identify, analyze, and synthesize the findings of the most relevant studies related to the application of Machine Learning and its impact within the field of e-Commerce. A systematic review was conducted on 66 papers extracted from recognized academic databases—Springer, Scopus, IEEE Xplore, Web of Science, and EBSCOhost—covering the period from 2018 to 2024. The methodology adopted was based on Kitchenham's (2009) guidelines, with detailed documentation of search equations, exclusion criteria, and quality assessment parameters to ensure the consistency, transparency, and reliability of the results obtained. The thematic analysis revealed that the categories "Intelligent Detection" and "Advanced Machine Learning" are particularly prominent in the scientific literature. Furthermore, it was observed that papers published in higher-quartile journals tend to offer conclusions with a greater degree of objectivity and methodological rigor. It is recommended to promote interdisciplinary studies that leverage the high frequency

of co-authorship identified, thereby fostering stronger scientific collaboration networks. Likewise, the homogeneity observed in paper titles reveals consolidated thematic lines, opening opportunities to explore innovative approaches in the field of e-commerce and machine learning.

**Keywords.** Machine learning; neural networks; natural language processing; artificial learning; e-commerce; systematic review.

## 1 Introduction

Machine Learning (ML) has consolidated itself as a key technology in e-commerce, facilitating the analysis of large volumes of data, the prediction of purchasing patterns, and the personalization of the user experience. Its ability to optimize operational processes and anticipate consumer preferences positions ML as an essential resource in highly competitive environments. In this context, understanding its transformative role is crucial to

maintaining competitiveness and responding agilely to changing market demands. Several recent studies illustrate significant advances in this area: for instance, a framework combining multi-factor authentication and machine learning has been proposed to strengthen the security of online financial transactions, employing two-factor authentication alongside facial recognition, achieving high accuracy rates and highlighting the approach's effectiveness and adaptability [1]; likewise, FAE has been introduced, a framework that optimizes the training of recommendation models by storing the most accessed embeddings in GPU memory, reducing data transfers and accelerating processing, decreasing training time by up to 2.3 times without compromising accuracy [2].

Furthermore, a study analyzes online shoppers' perceptions regarding last-mile delivery (LMD) services in Saudi Arabia, using ML models to identify preferences for fast deliveries, dissatisfaction with early delivery windows, and the rise of digital payments, underlining the need to improve LMD services to foster customer loyalty [3]. Additionally, a novel ML framework for customer review categorization based on a bag of features and hybrid deep neural networks has been proposed, which, when evaluated on AliExpress and Amazon datasets, achieved a 91.5% accuracy and an 8.46% fallout, outperforming existing models in sensitivity, specificity, and accuracy [4].

Regarding sales forecasting, an LSTM model has been presented that incorporates cross-information from related time series, improving demand prediction on Walmart.com by considering nonlinear patterns and product grouping strategies [5]. In the field of fake review detection, an approach combining aspect extraction with CNN and LSTM models has been proposed, achieving significant improvements in processing time and accuracy compared to traditional methods [6].

Through a systematic review of 225 papers and expert interviews, 21 key application areas of ML in retail have been identified, highlighting its role in decision-making and operational tasks, and proposing a structural framework for its effective implementation [7].

In the context of online education, the integration of Linked Open Data and predictive ML

models has been suggested to address the "cold-start" problem and adapt prior knowledge in course recommendations, achieving 90% accuracy [8]. Simultaneously, hybrid models have been proposed for hierarchical product classification in e-commerce, combining flat and local approaches through ML, improving the weighted F1-score and introducing valuable datasets in Spanish [10]; additionally, a task called Category-Aware Session-Based Recommendation (CSBR) has been introduced, implemented through an Intention Adaptive Graph Neural Network (IAGNN), outperforming previous methods across multiple real-world datasets [11]. In the area of customer service, ICS-Assist emerges as an intelligent framework employing two-stage machine learning and knowledge distillation to recommend solutions in real-time, significantly improving performance indicators at Alibaba [12]. Similarly, a recurrent neural network (RNN) survival model has been proposed to predict user return rates on websites, overcoming limitations of traditional techniques [13].

Shipping cost optimization and damage reduction are addressed through a multi-stage approach applied at Amazon, achieving substantial cost savings and carbon footprint reductions [14]. Moreover, content-based recommendation has been enhanced by integrating semantic information into the TF-IDF model, improving item similarity precision [15]. Regarding online fraudulent store detection, an efficient model has been proposed, achieving up to 96.93% accuracy with a reduced number of features [16].

The implementation of ML in API management, error detection, and organizational process optimization has been analyzed through a review of 85 recent studies [74]. In the area of sentiment analysis on product reviews, research indicates that SVM- and Random Forest-based models are the most effective for opinion classification [78], while other systematic reviews evaluate the effectiveness of ML techniques in fraud detection, identifying gaps and emerging trends [79], as well as the need for real-time prevention strategies [80].

Intelligent quality evaluation systems for courier services have also been developed [81], along with reviews of recommendation algorithms based on deep learning [82], and new taxonomies have been

proposed to classify ML applications in e-commerce [83]. In terms of sentiment analysis, the need to develop universal models and detect phenomena such as sarcasm is highlighted [84], while other studies emphasize the dominance of SVM and LSTM techniques [85], extending these findings to the hospitality and tourism sectors [89].

Despite significant progress, important gaps persist in the scientific literature on ML in e-commerce: many studies fail to adequately consider the adaptability of models to dynamic scenarios or their scalability in heterogeneous contexts, and there is a noticeable dispersion in methodological approaches and limited systematic integration of findings. This situation justifies the need for a rigorous systematic review that identifies existing limitations, consolidates available knowledge, and guides future research towards more robust, adaptable, and effective models in real-world digital commerce environments.

Consequently, this paper aims to analyze the effectiveness and applicability of Machine Learning techniques in sentiment analysis, identify gaps in the literature, and propose future research lines that optimize their implementation in e-commerce. The structure of the paper is organized as follows: Section II presents the theoretical framework; Section III describes the methodology applied; Section IV presents the main findings and discussion; and finally, Section V summarizes the conclusions and offers recommendations for future research.

## 2 Theoretical Background

Given the accelerated growth of Machine Learning and its influence on e-commerce, it is essential to understand the key concepts underpinning this approach before examining its most relevant current trends.

### 2.1 Machine Learning

Machine Learning (ML) is defined as the systematic study of algorithms and mathematical models that enable computational systems to learn to perform specific tasks based on patterns present in data, without the need for explicit programming

[15]. Its application is common in scenarios such as fraud detection, although it is also necessary to analyze the organizational context in which these models are implemented [17]. ML is part of artificial intelligence systems, whose objective is to develop agents capable of interacting with their environment, dynamically adapting, and adjusting their behavior through continuous feedback [18, 19]. Operationally, it is described as the ability of machines to generate programs based on observed data, recognizing patterns across multiple input-output pairs [20].

There are multiple methodological approaches within ML, such as supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, inverse reinforcement learning, and imitation learning, all of which have proven useful in error prediction and process improvement in areas such as API deployment or organizational functionalities [74,72]. In summary, ML constitutes a fundamental technique for data analysis and modeling, aimed at predicting outcomes with a high degree of accuracy in complex and dynamic environments [82].

### 2.2 E-Commerce

E-commerce is understood as the use of the Internet and digital networks to sell, purchase, exchange services, and share information [4], through electronic devices that enable data transmission [9]. It plays a key role in platform operations, especially within the retail sector [13], and serves as a driver of economic development in the digital era [15]. These platforms feature complex architectures with multiple vulnerabilities [79]. As a digital marketing strategy, e-commerce has gained prominence with the global expansion of the Internet [81]. In short, it has become the main environment where users discover, compare, and acquire products efficiently [83].

### 2.3 Tools Used

In the development of this study, Mendeley Desktop was used as the reference management software, facilitating the organization and handling of the papers.

Additionally, the RAJ Research Assistant tool, developed by Dr. Javier Gamboa Cruzado, was

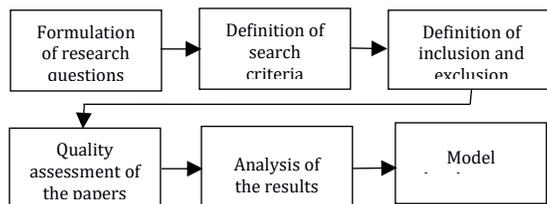


Fig. 1. SLR process

Table 1. Research Problems and Objectives

Research Question	Objective
RQ1. What Machine Learning algorithms are currently used in studies related to e-commerce, and what is their impact in this field?	Identify the algorithms used to provide a clear understanding of the tools preferred by the scientific community and industry.
RQ2. What quartile levels are presented by the journals publishing studies on the application of Machine Learning in e-commerce?	Analyze these levels to assess the relevance and credibility of research in the fields of Machine Learning and e-commerce.
RQ3. Which studies are distinguished by presenting conclusions with a high degree of objectivity and moderate polarity, and how are they distributed according to the publication quartile in Machine Learning applied to e-commerce?	Identify papers with conclusions of high objectivity and moderate polarity.
RQ4. Which authors appear recurrently as co-authors in scientific publications addressing the impact of Machine Learning on e-commerce?	Identify authors and their collaboration networks to detect leaders in the field, potential collaborators, and research trends.
RQ5. What are the main thematic categories addressed in the scientific literature on Machine Learning and its impact on e-commerce?	Classify the topics for better organization and understanding of the field, highlighting the most researched areas and those requiring further attention

Table 2. Search Descriptors and Their Synonyms

Descriptor	Description
machine learning / neuronal networks / automatic learning / predictive informatics	Independent Variable (A)
e-commerce / digital trading / retail commerce / digital business	Dependent Variable (B)

employed to support the systematic analysis of the scientific literature.

approach, a rigorous examination of the influence of Machine Learning on e-commerce is sought, with the purpose of providing substantiated answers to the formulated research questions.

### 3 Review Method

This study employs a systematic literature review (SLR) approach based on the methodological guidelines proposed by B. Kitchenham [67] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Through this

To ensure traceability and reproducibility of the process, a documentary database was constructed including the search equations, exclusion criteria, and quality assessment parameters used, thus achieving coherent and methodologically consistent results (see Figure 1).

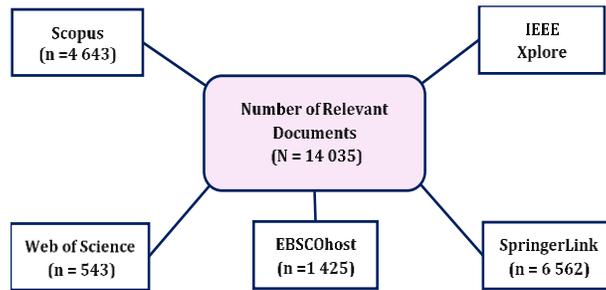


Fig. 2. Number of relevant documents

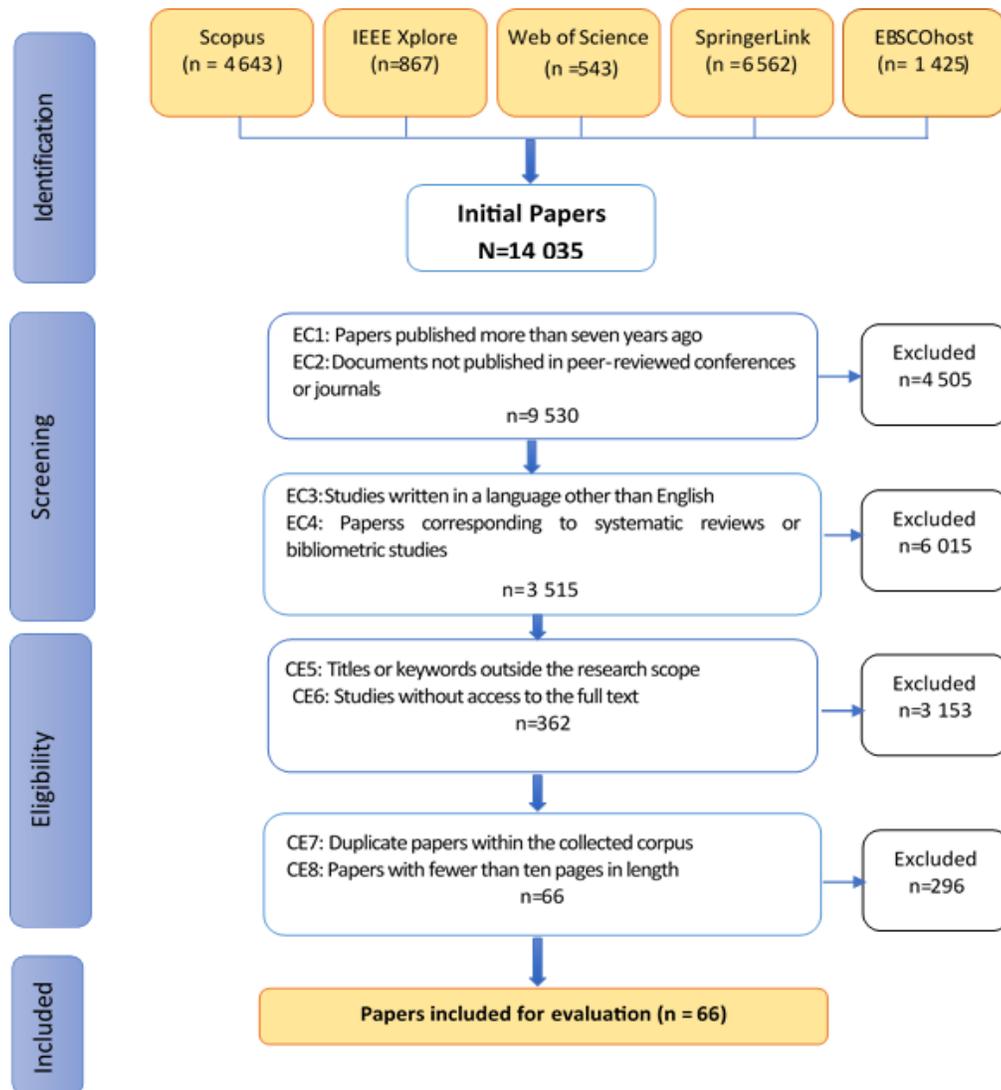


Fig. 3. PRISMA Chart

**Table 3.** Information Sources and Search Equations.

Source	Search Equation	N° of Results
Scopus	TITLE-ABS-KEY (("machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND ("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business"))	4 643
IEEE Xplore	((("Document Title": " machine learning") OR ("Document Title": "neural networks") OR ("Document Title": "automatic learning") OR ("Document Title": "predictive informatics")) AND (("Document Title": "e-commerce") OR ("Document Title": "digital trading") OR ("Document Title": "retail commerce") OR ("Document Title": "digital business"))) OR (((("Abstract": " machine learning") OR ("Abstract": "neural networks") OR ("Abstract": "automatic learning") OR ("Abstract": "predictive informatics")) AND (("Abstract": "e-commerce") OR ("Abstract": "digital trading") OR ("Abstract": "retail commerce") OR ("Abstract": "digital business"))) OR (((("Author Keywords": " machine learning") OR ("Author Keywords": "neural networks") OR ("Author Keywords": "automatic learning") OR ("Author Keywords": "predictive informatics")) AND (("Author Keywords": "e-commerce") OR ("Author Keywords": "digital trading") OR ("Author Keywords": "retail commerce") OR ("Author Keywords": "digital business"))))	867
Web of Science	(TI=("machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND TI=("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (AB=(" machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND AB=("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (AK=(" machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND AK=("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business"))	543
SpringerLink	(Title:( " machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND Title:( "e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (Abstract:( "machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND Abstract:( "e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (Keyword:( " machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND Keyword:( "e-commerce" OR "digital trading" OR "retail commerce" OR "digital business"))	6 562
EBSCOhost	(TI (" machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND TI ("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (AB (" machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND AB ("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business")) OR (SU (" machine learning" OR "neural networks" OR "automatic learning" OR "predictive informatics") AND SU ("e-commerce" OR "digital trading" OR "retail commerce" OR "digital business"))	1 425

### 3.1 Research Questions and Objectives

Given the relevance of studies on the impact of Machine Learning in e-commerce, it is essential to define an appropriate search strategy that efficiently collects information from each study. For this purpose, specific research questions (RQ) were formulated, serving as a fundamental guide for the analysis. Five research questions along with their respective objectives are presented below (see Table 1).

### 3.2 Information Sources and Search Equations

The databases used to collect relevant studies were Scopus, IEEE Xplore, Web of Science,

SpringerLink, and EBSCOhost. Table 2 presents the search approach employed, which was based on the structured exploration of keywords and their synonyms.

The search process was conducted using a set of relevant terms that facilitated the identification, extraction, and analysis of information. Table 3 details the search equations, adapted according to the characteristics of each database consulted.

### 3.3 Identified Studies

After completing the search in each selected database, the initial set of relevant studies for the review was compiled (see Figure 2).

**Table 4.** Quality assessment results

Ref.	Tipo	QA1	QA2	QA3	QA4	QA5	QA6	QA7	Score
[1]	Journal	3	3	3	3	3	3	3	21
[2]	Journal	3	3	3	3	3	3	3	21
[3]	Journal	3	3	3	3	3	3	3	21
[4]	Journal	3	3	3	3	3	3	3	21
[5]	Journal	3	2	3	2	3	2	2	17
[6]	Journal	2	3	2	3	3	3	2	18
[7]	Journal	2	2	2	3	3	3	3	18
[8]	Journal	1	3	2	2	1	3	2	14
[9]	Journal	2	1	3	3	2	3	3	17
[10]	Journal	3	2	3	2	1	2	3	16
[11]	Journal	2	1	3	2	2	3	3	16
[12]	Journal	2	2	1	3	3	2	3	16
[13]	Journal	2	2	3	1	1	3	3	15
[14]	Journal	3	2	3	2	1	1	2	14
[15]	Journal	2	2	3	3	1	1	3	15
[16]	Journal	3	3	3	1	1	2	2	15
[17]	Journal	3	2	1	3	2	1	3	15
[18]	Journal	3	2	1	3	2	1	3	15
[19]	Journal	3	2	1	3	2	1	2	14
[20]	Conference	3	2	1	3	2	1	3	15
[21]	Conference	3	2	2	3	2	1	2	15
[22]	Conference	2	2	2	2	3	3	3	17
[23]	Journal	3	3	3	3	3	3	3	21
[24]	Journal	2	2	2	3	3	3	1	16
[25]	Conference	3	2	1	3	2	1	3	15
[26]	Journal	3	2	1	3	2	1	3	15
[27]	Journal	3	2	1	3	2	1	2	14
[28]	Journal	3	2	1	3	2	1	2	14
[29]	Journal	3	2	1	3	2	1	3	15
[30]	Journal	3	2	1	3	2	1	3	15
[31]	Journal	3	2	1	3	2	1	3	15
[32]	Journal	3	2	1	3	2	1	3	15
[33]	Journal	2	3	1	3	2	1	3	15

### 3.4 Exclusion Criteria

Exclusion criteria (EC) were defined to rigorously evaluate the quality of the selected literature. The retrieved papers were considered for this review based on an objective exclusion list. A total of eight criteria were applied, detailed below to delimit the study corpus:

CE1: Papers published more than seven years ago.

CE2: Documents not published in peer-reviewed conferences or journals.

CE3: Studies written in a language other than English.

CE4: Papers corresponding to systematic reviews or bibliometric studies.

CE5: Titles or keywords outside the research scope.

CE6: Studies without access to the full text.

CE7: Duplicate papers within the collected corpus.

CE8: Papers with fewer than ten pages in length.

**Table 4.** (Continuation)

Ref.	Tipo	QA1	QA2	QA3	QA4	QA5	QA6	QA7	Score
[34]	Journal	3	2	1	3	2	1	2	14
[35]	Journal	3	2	2	2	1	3	3	16
[36]	Conference	3	3	3	2	2	2	2	17
[37]	Conference	2	2	3	3	1	3	2	16
[38]	Journal	3	2	1	3	2	1	2	14
[39]	Journal	3	2	1	3	2	1	3	15
[40]	Journal	3	2	1	3	2	1	2	14
[41]	Conference	3	2	1	3	2	1	2	14
[42]	Journal	3	2	1	3	2	1	2	14
[43]	Journal	3	2	1	3	2	1	3	15
[44]	Journal	3	2	1	3	2	1	3	15
[45]	Journal	3	2	1	3	2	1	3	15
[46]	Conference	1	2	3	1	2	3	1	13
[47]	Journal	3	2	1	3	2	1	2	14
[48]	Journal	3	2	1	3	2	1	2	14
[49]	Journal	3	2	1	3	2	1	2	14
[50]	Journal	3	2	1	3	2	1	3	15
[51]	Journal	3	3	2	2	1	3	2	16
[52]	Journal	3	2	1	3	2	1	3	15
[53]	Journal	3	2	1	3	2	1	3	15
[54]	Journal	3	2	1	3	2	1	3	15
[55]	Journal	3	2	1	3	2	1	2	14
[56]	Journal	3	2	1	3	2	1	3	15
[57]	Journal	3	2	1	3	2	1	3	15
[58]	Journal	3	2	1	3	2	1	3	15
[59]	Journal	3	2	1	3	2	1	2	14
[60]	Journal	3	2	1	3	2	1	2	14
[61]	Journal	3	2	1	3	2	1	2	14
[62]	Journal	3	2	1	3	2	1	3	15
[63]	Journal	2	3	1	2	3	3	2	16
[64]	Journal	3	2	1	3	2	1	3	15
[65]	Journal	3	2	1	3	2	1	3	15
[66]	Journal	1	2	3	1	2	3	3	15

### 3.5 Study Selection

Initially, 14 035 papers were identified based on keyword searches. After applying the defined filters, 66 pertinent studies were selected for analysis, as shown in Figure 3.

### 3.6 Quality Assessment

During this stage, the selected papers were analyzed by applying seven quality criteria (QA). In

this final selection and filtering period, the official list of papers was established, including a quality assessment to ensure that each paper's research was understandable and reliable.

The criteria applied were:

QA1: Does the study clearly and explicitly identify its research objectives?

QA2: Is the methodology applied appropriate and accepted within the field of study?

Table 4. (Continuation)

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations	
[13] Cui et al. (2022)	Intention Graph Neural Network (IAGNN)	Adaptive Neural Network	To propose Category-aware Session-Based Recommendation (CSBR) leveraging user-specified categories for improved recommendations	Diginetica, Yoochoos e, Jdata	Introduced CSBR task and IAGNN model, utilizing category-aware graphs for dynamic user intention capture	Does not account for noisy information from unrelated categories; limited to specific datasets
[14] Dewey et al. (2023)	Linear Analysis, Trees	Discriminant Decision	To assess accounts receivable risk in SMEs using machine learning techniques	Customer data from John J. Jerue Truck Broker	Demonstrated the feasibility of tailored risk assessments for SMEs using accessible ML tools	Limited to internal data, potential for customer misclassification, and lacks advanced features like credit limit automation
[15] Fu et al. (2020)	Two-stage learning knowledge network (Panel-Student)	machine model, distillation (Panel-Student)	Recommend suitable customer service solutions at runtime	Historical customer service logs	Proposed ICS-Assist framework for real-time scenario recognition and solution mapping	Limited to East Asian and Southeast Asian data; potential biases in training data
[16] Govin daraj et al. (2023)	LSTM, architecture, Attention mechanism, Bag of Words, Beam search decoding	Seq2Seq	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[17] Grob et al. (2019)	RNN survival model combining RNNs and survival analysis		Predict user return time using time series of user sessions	User sessions from ASOS.com	Developed a novel RNN survival model that incorporates censored data and learns features from raw time series	Cannot account for external factors influencing user behavior; relies on historical data for predictions
[18] Grote (2024)	LSTM, architecture, Attention mechanism, Bag of Words, Beam search decoding	Seq2Seq	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[19] Gurum oorthy et al. (2021)	Multi-stage approach with linear time algorithm and binary search for hyper-parameter tuning		Optimize package type selection to minimize shipment and damage costs	Historical shipment data	Developed a scalable, efficient model for package type recommendation, leading to significant cost savings	Limited to existing product attributes; may not generalize to all product types
[20] Huang (2023)	Improved content recommendation algorithm integrating semantic information with TF-IDF		Enhance semantic analysis in content-based recommendation systems	Course title, overview, teaching objectives	Proposed a novel algorithm combining TF-IDF and word embedding for better item similarity	Limited to specific datasets; may not generalize across all domains
[21] Hwang et al. (2020)	Workload characterization, chiplet-based hybrid sparse-dense accelerator design		Address memory and compute bottlenecks in personalized recommendation systems	Not specified	Introduced Centaur, an accelerator that improves memory throughput and compute efficiency for embedding and MLP layers	Limited to the specific architecture of Intel HARPv2; may not generalize to other systems
[22] In et al. (2024)	Self-guided framework (SG-GSR), graph augmentation, group-training	GSR	To enhance GNN robustness against adversarial attacks by utilizing a clean sub-graph from the attacked graph	Not specified	Introduced SG-GSR addressing limitations of existing GSR methods, effective under various attack types	Assumes existence of clean sub-graphs, may not generalize to all real-world scenarios
[23] Janavi ċiūtė et al. (2024)	Machine learning models evaluating combinations of URL-based, content-based, and third-party features		To identify minimal feature sets for high-precision detection of fraudulent online shops	Custom dataset of 1140 records	Identified key features for effective fraud detection with minimal computational resources	Limited to publicly available features; may not cover all fraudulent shop tactics
[24] Ji et al. (2023)	LSTM, architecture, Attention mechanism, Bag of Words, Beam search decoding	Seq2Seq	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.

Table 4. (Continuation)

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations
[34] Maashi et al. (2023)	Ensemble learning with GRFO optimization, ELM, BiLSTM, AE	Design an intelligent credit card fraud detection system using sustainable practices	Kaggle credit card transaction dataset	Introduced a novel GRFO-based feature selection and POA for hyperparameter tuning	Limited by dataset size and potential overfitting issues
[35] Mahendra et al. (2024)	Knowledge graph embedding co-learning with Graph Neural Networks	Assess the efficacy of knowledge-aware deep recommender models	Amazon-Book, Last-FM	Introduced complex-valued KG embeddings; Enhanced recommendation accuracy and reduced training time	Limited exploration of trade-offs between performance gain and computational cost
[36] DENG et al. (2023)	Deep Forest algorithm, symmetric sampling, SMOTE	Develop a detection model for the "Ride Item's Coattails" attack in e-commerce recommendation systems	Real recommendation system attack data	Introduced a novel attack detection model using Deep Forest; addressed class imbalance with symmetric sampling and SMOTE	Limited to specific attack types; may not generalize to all e-commerce platforms
[37] Mohamed et al. (2022)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[38] N et al. (2024)	ABNAM, TF-IDF, N-gram, CNN, linear SVM, random forest, Naïve Bayes	To develop a novel classification model for aspect extraction in e-commerce product reviews	Product review dataset from open-source repositories	Introduced ABNAM as a novel model for efficient aspect extraction in e-commerce	Existing models could not classify into 16 distinct categories; limitations in dataset size not addressed
[39] Navaneethan et al. (2023)	HIASTLO algorithm, deep learning	To enhance data integrity and security in IoT-based e-commerce blockchain applications	Not specified	Proposed a novel deep learning optimization model for detecting various cyberattacks in e-commerce	Limited sample size; specific datasets not mentioned; potential scalability issues
[40] Nishimura et al. (2023)	Shape-restricted optimization model, algorithms for transitive reduction	Examine the relationship between user pageview histories and item-choice behavior on e-commerce websites	Real-world clickstream data	Proposed a model that estimates item-choice probabilities using PV sequences and monotonicity constraints	Complexity of constraints can be extremely large; computational efficiency may still be a concern
[41] Arouj et al. (2020)	Secure Federated Submodel Learning (SFSL), Private Set Union (PSU) protocol	To enhance federated learning efficiency and privacy for mobile clients in e-commerce recommendations	30-day Taobao user data	First to propose FSL framework for on-device intelligence, enabling tunable privacy and efficient aggregation	Introduces new privacy risks; requires careful tuning of privacy and efficiency
[42] Nkongolo et al. (2023)	Fuzzy logic-based feature selection, ensemble learning	Defend critical infrastructures against zero-day threats	NSL-KDD, UGRansome	Proposed a computational framework enhancing ML model performance for IDS	Limited generalizability to other datasets; reliance on specific feature selection methods
[43] Punnett et al. (2021)	Collaborative filtering	Propose a health recommender system to recommend remedies based on symptoms	Custom dataset of remedies for various diseases	Developed a collaborative filtering-based recommender system for healthcare, created a comprehensive dataset	Limited to the diseases included in the dataset; may not generalize to all medical conditions
[44] Quang et al. (2022)	Linear Regression, Random Forest, LightGBM	Develop a machine learning model for price suggestion in e-commerce	Mercari dataset	Demonstrated effectiveness of LightGBM for pricing suggestions	Limited to specific dataset and may not generalize to all e-commerce platforms

QA3: Is the context in which the research was conducted accurately described?

QA5: Are the methods used to interpret the results pertinent and justified?

QA4: Does the document present a coherent, logical, and well-organized structure?

QA6: Are the experimental findings clearly presented and documented?

Table 4. (Continuation)

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations
[25] Jiang et al. (2022)	GNN-based recommendations, ROI concept, multi-level attention module	Improve training/serving efficiency and recommendation quality at Taobao	User-query-item relevance graph	Introduced ROI for GNNs, scalable system for web-scale recommendations	Limited by dynamic user interests and potential information overload
[26] Khmel nitsky et al. (2023)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[27] Kumar et al. (2023)	Statistical methods, user-oriented questionnaire, model trees, bagging	Analyze usability and security of e-commerce websites in Jharkhand, India	Data from e-commerce websites	Developed a framework for assessing usability and security using AHP, VIKOR, and TOPSIS methods	Limited to specific regional websites; does not account for broader e-commerce trends
[28] Li et al. (2022)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[29] Liu et al. (2024)	ARIMA, LSTM, Random Forest, PSO, Bayesian optimization	To predict demand for agricultural products in e-commerce using a single model-based approach	E-commerce agricultural products dataset	Developed optimized demand prediction models for agricultural products using data mining techniques	Limited by the specificity of agricultural product characteristics and potential data noise
[30] Liu et al. (2022)	Combined Compromise Solution (CoCoSo), Analytic Hierarchy Process (AHP), Single-Valued Neutrosophic Sets (SVNSs)	Reviewed recent advances in medical imaging using adversarial training, offering insights for researchers	Telecommunication industry chatbots	Developed a novel MCDM framework for chatbot selection under uncertainty, integrating CoCoSo with AHP and SVNSs, demonstrating improved decision-making in complex environments.	Does not account for social influences relevant to the voluntary consumer context; it might not be appropriate for institutional applications that require the integration of information technology.
[31] Liu et al. (2024)	VATA model (Variational Autoencoder, Transformer, Attention Mechanism)	Analyze user shopping behavior for improved e-commerce experiences	E-Commerce Dataset, Behavior Trajectory Dataset, Social Media Consumption Dataset	Proposed a novel VATA model for intelligent classification and analysis of user shopping behavior	Does not address potential overfitting; may require extensive computational resources
[32] Liu et al. (2024)	Eye-tracking technology, statistical analysis, machine learning models	Investigate online consumer behavior using eye-tracking data to enhance e-commerce recommendation systems	Eye-tracking data from online shopping	Integration of eye-tracking data improves recommendation system performance; insights on consumer attention allocation	Limited to healthy participants aged 18-35; may not generalize to broader populations
[33] Ma et al. (2020)	AUI algorithm, multidimensional trajectory set model	To improve anonymous user identification using clickstream data	Online behavior log data from CNNIC	Proposed a novel method for user recognition based on multidimensional trajectory sets, outperforming decision trees and user documents	Limited to specific user behaviors; further research needed for broader applicability

QA7: Is the study relevant and appropriate given its objectives and thematic scope?

Each paper was evaluated based on these seven criteria using a scale from 1 to 3 (1 = Poor,

2 = Fair, 3 = Excellent). The minimum score for exclusion was 13.

As a result of the evaluation of the 66 papers, all primary studies met each of the QAs with a score greater than or equal to 11.5. Through this evaluation, the definitive number of publications

Table 4. (Continuation)

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations
[45] Rodrigues et al. (2021)	Machine learning methods for schema matching networks, bootstrapping for training data generation, user feedback integration	To identify semantic correspondences in multiple schemas simultaneously, improving schema matching quality	Not specified	Proposed methods outperform traditional matching, leveraging constraints and user feedback for improved accuracy	Requires large labeled datasets, potential inconsistency in matches across schemas
[46] Samonte et al. (2022)	Literature review, LSTM, Random Forest	To develop a prediction model for retail sales using deep learning techniques	Various datasets from selected studies	Identified LSTM as the most effective deep learning technique for sales forecasting	Limited to studies published between 2017-2021; does not account for external factors affecting sales
[47] Singh et al. (2021)	Knowledge-based collaborative filtering algorithm	To develop an efficient recommendation system using user interaction data for E-commerce	Amazon E-commerce dataset	Proposed a novel recommendation system integrating user activity data for improved accuracy	Limited to specific user behaviors and may not generalize across different domains
[48] Ray (2021)	Logistic regression, random forest, XGBoost, artificial neural networks, support vector machines	Develop statistical acceptance prediction models for loan offers in OLCB platforms	Data on customer behavior, demographics, financial variables, loan offer characteristics	First conversion prediction models for OLCB platforms, enhancing understanding of consumer behavior in loan decisions	Lack of extensive research on OLCB-specific consumer behavior; potential data handling challenges
[49] Tran et al. (2023)	Operational model formulation, literature review	To derive research topics and practical challenges for fraud detection in e-commerce	Not specified	Identified 6 research topics and 12 practical challenges for fraud detection, highlighting the need for an organization-centric view	Lacks empirical validation of proposed solutions; does not address specific implementation challenges in diverse e-commerce environments
[50] Tsuboi et al. (2024)	LSTM, architecture, Seq2Seq Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability and user experience.
[51] Umer et al. (2021)	Variational inference method for multiple treatment estimation	To estimate effects of multiple treatments using observational data while considering causal relationships	Artificial and real-world datasets	Proposed a novel method for estimating treatment effects from diverse measures, including continuous parameters	Limited by the need for accurate proxy variables and potential biases from non-confounders
[52] Upreti et al. (2024)	LSTM, architecture, Seq2Seq Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability and user experience.
[53] Walter (2024)	Survey of security, privacy, and defense techniques in federated learning	Investigate security and privacy in AI, focusing on federated learning and its challenges	Not specified	Highlighted the need for a unified approach to security, privacy, and trustworthiness in AI	Lack of empirical validation; challenges in implementation and scalability
[54] Wang (2024)	LSTM, architecture, Seq2Seq Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.

included in the study was established (see Table 4).

Table 4. (Continuation)

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations
[55] Wasilewski (2024)	Machine Learning, Multi-Objective Optimization, ANFIS, Particle Swarm Optimization	Enhance supply chain performance in Cross-Border E-Commerce (CBEC) through demand volume prediction	Sales records from six retailers	Proposed an IoT-based framework for CBEC, utilizing ANFIS and PSO for demand prediction	Limited to specific retailers; may not generalize across all CBEC contexts
[56] Xie et al. (2024)	Framework development, AI clustering	To create a framework for multivariate e-commerce user interfaces and address personalization challenges	Customer behavior data	Proposed a framework for personalized UIs, introduced PCR metric, demonstrated practical implementation	Limited research on UI variant effectiveness, challenges in data collection and privacy
[57] Xu et al. (2022)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[58] Xu et al. (2024)	Markov decision process, deep learning (CNN, RNN)	Develop a recommendation algorithm for cross-border e-commerce logistics	Simulated shopping environment	Proposed a hybrid model combining CNN and RNN for improved recommendation accuracy	Limited by data sparsity and model complexity
[59] Zhang et al. (2022)	Federated learning, privacy algorithm	Develop a federated recommendation framework that preserves user privacy while maintaining recommendation accuracy	MovieLens, Amazon Product Reviews	Introduced a novel framework integrating federated learning with differential privacy, ensuring user data remains local	Challenges in optimizing privacy-utility trade-off; noise may affect recommendation quality
[60] Zhang et al. (2023)	Competitive Graph Neural Networks (CGNN)	To detect fraud behaviors on e-commerce platforms, specifically Taobao, using a novel approach that eliminates reliance on predefined fraud patterns	Two Taobao datasets, two public datasets	Introduced eFraudCom, which uses weak supervision from normal behaviors and mutual information regularization to enhance fraud detection	Limited by the need for representative normal samples; may not generalize to all e-commerce platforms
[61] Zhang et al. (2021)	Dual Graph Multitask framework, Graph Neural Networks	To improve Delivery Time Estimation (DTE) in e-commerce by addressing data imbalance	Taobao logistics datasets	Proposed a novel framework that effectively predicts delivery time for both high-shot and low-shot data	Limited to specific datasets; may not generalize across different e-commerce platforms
[62] Zhang et al. (2021)	Conformal prediction framework for PQA	Improve answer reliability and handle unanswerable questions in product reviews	Amazon dataset	Proposed a rejection model to enhance PQA reliability and accuracy	Limited to the Amazon dataset; does not account for varying question phrasing
[63] Zhang et al. (2022)	Grain framework integrating diversified influence maximization with GNNs	Improve data efficiency in GNNs through data selection methods	Public datasets	First to connect GNN data selection with social influence maximization; introduces novel influence and diversity functions	Existing methods ineffective for GNNs; challenges in scaling to large graphs
[64] Zhao et al. (2024)	Proposed framework for optimizing GPU-centric training of recommender systems through interleaving, and caching	To enhance training throughput and hardware utilization for wide-and-deep recommender systems	Industrial recommendation models	Introduced a novel optimization framework that significantly improves GPU utilization and reduces training delays	Limited to specific hardware configurations; may not generalize to all recommender system architectures
[65] Zhou et al. (2024)	Multi-angle perceptual convolutional neural network, deep perception model (DPM)	To evaluate the logistics service quality of cross-border e-commerce enterprises	Not specified	Proposed a novel evaluation framework integrating shallow and deep features for service quality assessment	Does not specify the datasets used; lacks detailed performance metrics
[66] Špicas et al. (2023)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.

## 4 Results and Discussion

This section presents and examines the findings obtained, contextualizing them in relation to the literature reviewed and the objectives outlined in this research.

### 4.1 Overview of the Studies

Table 5 provides a comprehensive summary of the reviewed studies, highlighting the methodological variety, research objectives, datasets used, key contributions, and identified limitations. This synthesis allows for the analysis of common patterns, advances, and gaps in the application of Machine Learning in e-commerce.

**Methods Used – Algorithmic Diversity and Emerging Approaches:** The studies exhibit a rich methodological diversity, highlighting techniques such as deep neural networks (LSTM, Seq2Seq), traditional statistical models (regressions, trees), and recent frameworks like Graph Neural Networks, GANs, and federated learning. This heterogeneity reflects a growing maturity of the field but also a fragmentation that still prevents convergence toward replicable protocols. An experimental approach predominates, focused on performance improvement, although there is little cross-methodology comparison.

**Research Objectives:** The research objectives address real-world e-commerce challenges: personalized recommendation, demand forecasting, transaction security, user experience, and semantic classification. A practical orientation toward enhancing user-system interaction is observed, but emerging topics such as privacy, fraud, and last-mile logistics are also present. However, many papers lack explicit hypotheses, which limits the potential for replicability and theoretical contrast.

**Datasets – Heterogeneity:** Datasets vary widely in type, size, and origin: from public sources (Kaggle, Amazon) to unspecified private repositories. This lack of standardization compromises reproducibility and hinders comparative analysis between studies. Furthermore, many models are trained on highly contextual data (e.g., Taobao, Mercado Libre Peru), raising serious concerns about their generalization to other markets or languages.

### Key Contributions – Functional Advances, Prototypical Models, and New Architectures:

Key contributions highlight improvements in computational efficiency, recommendation accuracy, proactive fraud detection, and logistics optimization. Innovative architectures (such as VATA or eFraudCom), hybrid models, and multi-objective systems are introduced. However, most contributions are technical or architectural in nature, with little attention to the ethical, social, or regulatory implications of using ML in real-world e-commerce environments.

### Limitations – Data, Context, and Evaluation

**Constraints:** The most common limitations include dependence on historical data, geographic bias, overfitting, lack of scalability, and absence of longitudinal evaluation. Several studies acknowledge methodological constraints or limited applicability, but few systematically address the robustness of their models in dynamic, adversarial, or diverse user environments. Moreover, there is a frequent reuse of benchmark datasets with little analysis of their representativeness.

This review reveals a vibrant ecosystem of innovation in Machine Learning applied to e-commerce, but also a lack of methodological convergence that hinders knowledge consolidation. The absence of standardized datasets and comparative frameworks limits cross-validation of models. It is recommended to promote open benchmarks, incorporate social impact metrics, and strengthen transversal evaluation in real-world contexts. This would help guide the responsible and effective adoption of intelligent solutions in global digital platforms.

Figure 4 shows the annual evolution of the number of papers related to Machine Learning in e-commerce, disaggregated by scientific source. A sustained growth in publications is evident from 2019 to 2024.

The results show a steady increase in scientific production, with particularly significant growth in 2023 and 2024, driven by publications in IEEE Xplore and Scopus. Scopus has maintained a constant presence since 2020, while EBSCOhost and SpringerLink have gained prominence in recent years.

**Table 5.** Summary of methodological characteristics and key findings of the reviewed studies

Study	Methods used	Research Objectives	Datasets	Key Contributions	Limitations
[1] Abbas et al. (2023)	Point Hierarchical Factorization, Process, Poisson	Develop a personalized recommendation algorithm considering time-sensitive user behavior and preferences	User interaction data	Introduced a novel framework for time-sensitive recommendations using point processes, enhancing user engagement	Does not account for external factors influencing user preferences over time
[2] Aburb eian et al. (2024)	Multi-factor authentication, machine learning (logistic regression, decision trees, random forest, naive Bayes)	To propose a framework integrating multi-factor authentication and machine learning for securing online financial transactions	Credit card transactions dataset from Kaggle	Developed a dual-layer security framework combining MFA and ML for enhanced transaction safety	Limited exploration of user experience with MFA systems; dataset imbalance issues addressed but may still affect performance
[3] Adnan et al. (2021)	Proposed FAE framework for hot-embedding aware data layout	To optimize training of recommender models by leveraging popular choices in embedding access patterns	Real-world datasets	Introduced a framework that reduces CPU-GPU communication and optimizes embedding storage	Requires recalibration of hotness for different datasets; potential accuracy impact from mini-batch scheduling
[4] Aljohani (2024)	Machine learning classification and regression models	Examine online shoppers' perceptions of last-mile delivery (LMD) services and its impact on shopping behavior	Online survey data from Saudi Arabia	Provides insights into consumer preferences for timely LMD services and the shift towards digital payments	Limited to urban areas in Saudi Arabia; may not generalize to other regions
[5] Alotaibi (2023)	Bag-of-features approach, hybrid DNN framework	Propose a machine-learning framework for classifying customer reviews on social media	AliExpress, Amazon	Developed a novel opinion extraction mechanism using NLP and DNN for improved classification	Does not account for audio and visual review components
[6] Banda et al. (2019)	Long Short-Term Memory (LSTM) network, systematic pre-processing framework, product grouping strategies	To improve sales forecasting in E-commerce by leveraging cross-series information from related products	Real-world online marketplace dataset from Walmart.com	Introduced a unified model incorporating cross-series relationships and systematic preprocessing for E-commerce sales forecasting	Limited by the dynamic nature of E-commerce and potential overfitting in complex models
[7] Bathla et al. (2022)	LSTM, Seq2Seq architecture, Attention mechanism, Bag of Words, Beam search decoding	Authors claim that using GAN and its variants improves medical image segmentation accuracy	Dialog dataset	Proposed an intelligent chatbot model enhancing response accuracy using LSTM and attention mechanisms.	Overemphasis on threat, which may not be the primary motivator for all individuals considering technologies; does not consider the technology's usability.
[8] Beltzung et al. (2020)	Machine Learning for structural analysis	Classify fraudulent online shops based on source code structure	Archived fake-shop and legitimate shop datasets	Developed Fake-Shop Detection API and Middleware for real-time risk assessment	Limited to DACH region; may not generalize to other languages or cultures
[9] Brackmann et al. (2023)	Structured literature review and expert interviews	Identify application areas for ML in retail	Not specified	Developed a framework for practitioners to determine ML use in retail	Limited to identified application areas; does not cover all potential ML applications
[10] Chana et al. (2024)	Machine learning for prerequisite identification and recommendation	To enhance e-learning recommendations by addressing cold-start issues and ensuring learners meet prerequisite knowledge	Not specified	Proposed a novel prerequisite-based recommendation system using Linked Open Data and machine learning	Limited to the effectiveness of prerequisite identification and may not generalize across all learning contexts
[11] Sidelnik et al. (2019)	Systematic literature review, scoring method	Evaluate the quality of economic and financial data to satisfy shareholder demands and assess the influence of shareholding structure	Annual and interim financial reports	Highlights the need for qualitative economic and financial communication in decision-making	Limited to companies listed on the Bucharest Stock Market; Focused on stakeholder demands only
[12] Cotacallapa et al. (2024)	Comparative analysis of seven ML algorithms (MNB, LSVC, MLR, RF, XGBoost, FastText, Voting Ensemble)	To enhance hierarchical product classification in e-commerce by integrating flat and local approaches	Spanish-language dataset of over 1 million products from Mercado Libre Peru	Introduced a hybrid model for better classification performance and a significant Spanish dataset	Limited to specific e-commerce context; potential taxonomic inconsistencies

This increase suggests a growing interest from the academic community in ML applications in e-commerce. A clear upward trend is observed, validated by Kendall's trend analysis with a p-value

of 0.009 and a coefficient of determination  $R^2 = 0.99$ , indicating a strong positive correlation and that this growth is not random but systematic. Authors Aléxis Cárdenas-Quispe et al. [68]

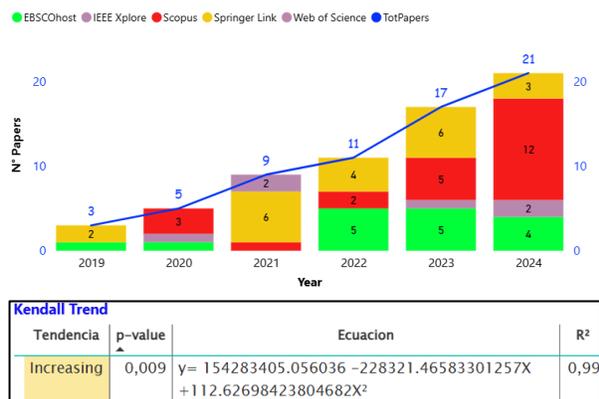


Fig. 4. Papers by year, source, and trend

highlights that the most relevant source for retrieving their papers was IEEE Xplore, with a total of 49 identified articles. In contrast, José Rojas Valdivia et al. [74] report that Wiley Online Library was the most frequently used search source in their study, accounting for 22.45%. For Lucas Micol Policarpo et al. [83], the most consulted source was Science Direct, peaking in 2018. Finally, Sabina-Cristiana Necula and Vasile-Daniel Păvăloaia [86] also identified Science Direct as their primary search source.

The increasing trend suggests that future research should consider the accelerated scientific output as a basis for more complex meta-analyses. This pattern may also be replicated in other sectors such as education, healthcare, or digital agriculture. Geographically, it is recommended to evaluate this trend in underrepresented regions. Moreover, the use of more sophisticated techniques in recent years could mark a turning point toward more advanced and applied research.

Figure 5 shows the geographical distribution of research on Machine Learning in e-commerce, classified according to the degree of objectivity in the conclusions.

China leads with the highest number of publications (19), although it shows a significant proportion of subjective conclusions (5). India and Saudi Arabia also exhibit high production levels, tending toward neutral or objective conclusions. European countries such as Germany, France, and the Netherlands, despite having fewer publications, demonstrate a high average level of objectivity, standing out for their scientific rigor.



Fig. 5. Papers by country and objectivity of conclusions

The highest average objectivity is recorded in Poland, with a single but highly objective study. The global distribution reveals a clear concentration of studies in Asia, followed by Europe and the Americas.

Authors Javier Gamboa-Cruzado et al. [73], in their research journal, state that the countries with the highest publication prevalence are the United States and China. Although José Rojas Valdivia et al. [74] does not report consistent objectivity in the conclusions of their papers, they do in their abstracts, where the United States is highlighted as the country with the most publications. In these abstracts, the author exhibits objectivity, while China, the second-leading country in publications, presents more neutral summaries in comparison.

These trends suggest that Asia—particularly China and India—is an emerging hub of ML research in e-commerce, although improvement in the objectivity of some cases is needed. Countries with high objectivity, though lower output, may serve as methodological benchmarks. This analysis may also guide future international collaborations with underexplored regions such as Africa or Latin America. Additionally, it serves as a basis for assessing how conclusions are constructed in different cultural and academic contexts.

Figure 6 illustrates a bibliometric network of keyword co-occurrence, highlighting the main concepts related to Machine Learning (ML) in the field of e-commerce, based on the frequency and strength of connections between terms (weight).

It is observed that the strongest relationship is between "ml" and "e-commerce" (weight = 7), followed by the link between "ml" and "deep learning" (weight = 4), confirming these as the central thematic axes in the literature. "deep learning" is closely linked to "neural networks" (weight = 3), reinforcing its role in complex tasks such as recommendations and fraud detection. Other combinations like "ml" with "fraud detection," "fake-shop detection," or "federated learning" (weight = 2) indicate the growth of research focused on security and privacy. The link between "e-commerce" and "iot" also stands out, indicating an interest in connected technologies.

Authors Andrés J. Aparcana-Tasayco and Javier Gamboa-Cruzado [71] show that "Machine Learning" has a strong relationship with "SDN" and "Software-Defined Networking." It also shows strong associations with the most commonly used ML techniques and key SDN concepts such as the OpenFlow protocol and security. Meanwhile, José Rojas Valdivia et al. [74] report that keywords like "artificial intelligence," "machine learning," "internet of things," and "industry 4.0" each showed three occurrences.

The evidenced connections can extend to sectors such as fintech, healthcare, or logistics, where ML and deep learning are already applicable. Additionally, regions with advanced digital development may lead the exploration of techniques such as federated learning. The relationship between ML and security suggests future research directions in data protection. Finally, these patterns can be replicated in other contexts and time frames to assess thematic evolution.

#### 4.2 Answers to the Research Questions

This section presents the answers to the research questions posed, considering the findings, discussions, and implications for future investigations.

The synthesis of results is organized according to the five research questions, clearly distinguishing between empirical and theoretical studies to differentiate evidence-based findings from proposals yet to be validated.

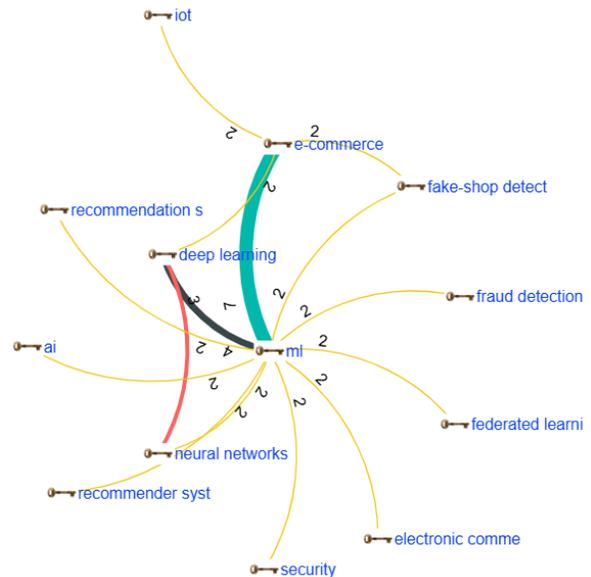


Fig. 6. Bibliometric Network of Keywords

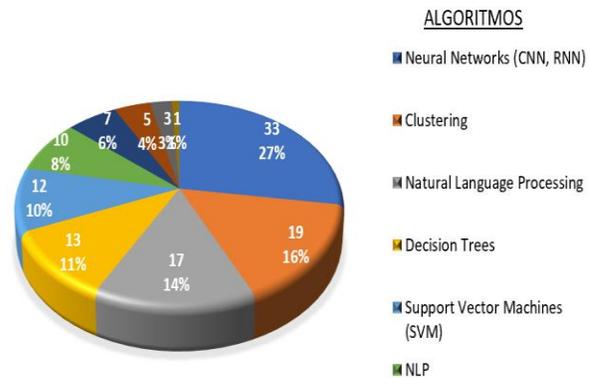


Fig. 7. Consolidated ML algorithms in e-commerce

**RQ1: What Machine Learning algorithms are currently used in research related to e-commerce, and what is their impact in this domain?**

Table 6 and Figure 7 present a percentage-based classification of the most frequently used Machine Learning algorithms in e-commerce research, allowing for a visualization of their distribution and relevance in this field.

This segmentation helps to understand which approaches are preferred by the scientific

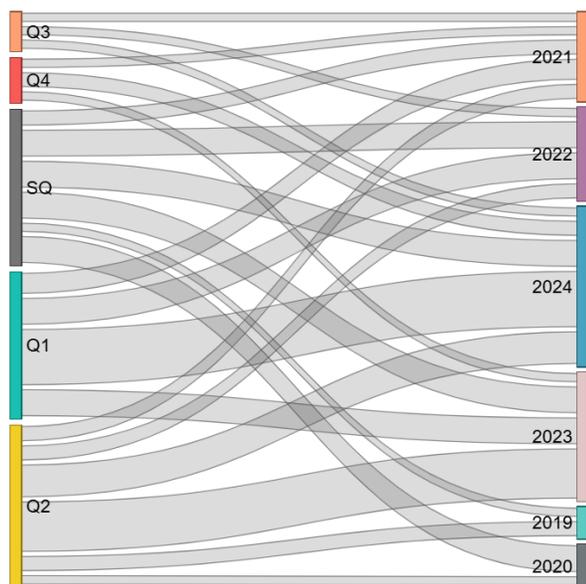


Fig. 8. Quartile level by publication medium

Table 7. Quartiles by publication source

Source	Q1	Q2	Q3	Q4	SQ	Total
Scopus	4	8	0	2	9	21
SpringerLink	6	9	1	2	3	21
EBSCOhost	5	3	2	0	6	16
IEEE Xplore	3	0	0	0	1	4
Web of Science	2	0	0	0	0	2
Total	20	20	3	4	19	66

community and which have shown greater applicability.

Neural Networks (CNN, RNN) clearly dominate, accounting for 27.5% of the studies, establishing themselves as the most widely used technique due to their ability to handle complex data. These are followed by Clustering (15.8%) and NLP (14.1%), reflecting the interest in customer segmentation and text processing, respectively. Other algorithms such as Decision Trees (10.8%), SVM (10%), and Regression Models (5.8%) also show significant presence, while methods such as Random Forests and Recommendation Algorithms are less frequently used. This methodological diversity

demonstrates the versatility of ML in predictive, classification, and recommendation tasks within e-commerce.

Authors Aléxis Cárdenas-Quispe et al. [68] highlight that the Support Vector Machine (SVM) algorithm is significantly more used for malware detection in Android, followed by Random Forest Regression and Naive Bayes, which are also commonly employed. Similarly, [71] identify Support Vector Machine as the most popular algorithm, followed by Naive Bayes, with 14.4% and 9.3%, respectively. On the other hand, [74] report that the most frequent algorithm in the reviewed papers is the Decision Tree, with 21.25%, followed by K-means, reaching 20%. Finally, Javier Gamboa-Cruzado et al. [75] note that the most widely used algorithm is Artificial Neural Network, followed by Random Forest and Decision Tree, with 36%, 28%, and 20%, respectively. According to authors Huang Huang et al. [84], the most widely accepted machine learning techniques in selected papers are Support Vector Machine (SVM) and Naive Bayes.

These findings can be extended to other sectors such as healthcare, logistics, tourism, or banking, where personalization and analysis of large datasets are critical. Moreover, the adoption of these algorithms in new geographic and temporal contexts will enhance strategic and operational decision-making. The increasing use of neural networks suggests their potential in dynamic and multivariable environments, even beyond e-commerce.

**RQ2: What quartile levels are presented by scientific journals publishing studies on The application of Machine Learning in e-commerce?**

Figure 8 and Table 7 display the distribution of the reviewed papers according to quartile levels (Q1 to Q4 and non-quartile - NQ) across the main academic databases. This representation facilitates the identification of the editorial quality levels in which knowledge related to Machine Learning in e-commerce is being published.

It is observed that papers published between 2022 and 2024 are predominantly concentrated in Q1 and Q2 journals, highlighting recent increase in publication quality. Notably, 2023 features a substantial volume of publications in Q1 journals. Non-

**Table 6.** Algorithms Used

Algorithm	Reference	Qty. (%)
Neural Networks (CNN, RNN)	[1] [2] [3] [4] [5] [6] [10] [11] [12] [13] [17] [21] [22] [23] [25] [26] [28] [34] [35] [36] [38] [39] [40] [44] [45] [48] [51] [53] [54] [56] [57] [59] [60] [61] [62] [64] [65]	33 (27.5)
Clustering	[3] [5] [6] [9] [16] [19] [28] [30] [36] [44] [45] [47] [49] [52] [54] [55] [56] [62]	19 (15.8)
Natural Language Processing	[4] [8] [12] [13] [15] [16] [19] [24] [26] [31] [33] [36] [38] [39] [53] [57] [65]	17 (14.1)
Decision Trees	[2] [7] [12] [14] [32] [34] [36] [38] [39] [42] [45] [46] [54]	13 (10.8)
Support Vector Machines (SVM)	[1] [14] [17] [20] [32] [34] [36] [38] [39] [40] [42] [44] [54]	12 (10.0)
NLP	[3] [5] [10] [17] [24] [26] [38] [51] [53]	10 (8.3)
Regression Models	[3] [7] [29] [40] [57] [60] [66]	7 (5.8)
Random Forests	[1] [17] [32] [40] [45]	5 (4.1)
Recommendation Algorithms	[10] [16] [20]	3 (2.5%)
Optimization Algorithms	[5]	1 (0.83)

quartile publications (NQ) also appear across the years, especially in 2022. In contrast, papers in lower quartiles (Q3 and Q4) are scarce and limited to specific years such as 2021. This pattern reflects a growing trend toward publication in higher-impact journals.

The results reveal a balanced distribution between Q1 (20 papers) and Q2 (20), indicating a strong commitment to publishing in high-impact journals. A significant number of studies (19) appear in non-quartile journals (NQ), suggesting dispersion in emerging or non-indexed journals. SpringerLink and Scopus host the largest volume of articles, highlighting their importance as academic sources. IEEE Xplore and Web of Science have lower representation, which could be attributed to the specific focus of their publications.

According to Alfredo Daza et al. [85], most publications in their analysis were found in Q1 journals, accounting for 35%. Meanwhile, Vanessa Duarte et al. [90] illustrate a progressive increase

in Q1 and Q2 journal output in their research, with Web of Science demonstrating the greatest stability in producing such papers.

The high concentration in Q1 and Q2 journals supports the methodological quality of the studies and can serve as a foundation for expanding this research into sectors such as healthcare, education, or manufacturing. Geographically, it may encourage publication in indexed journals from other regions. Furthermore, future studies could prioritize venues with greater visibility and rigor, promoting increased standardization and validation of results in new contexts.

**RQ3: Which studies are distinguished by presenting conclusions with a high degree of objectivity and moderate polarity, and how are they distributed according to the publication quartile in research on Machine Learning applied to e-commerce?**

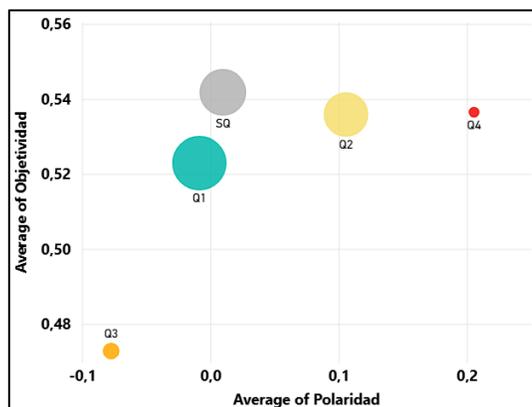


Fig. 9. Papers with high objectivity and moderate polarity in their conclusions

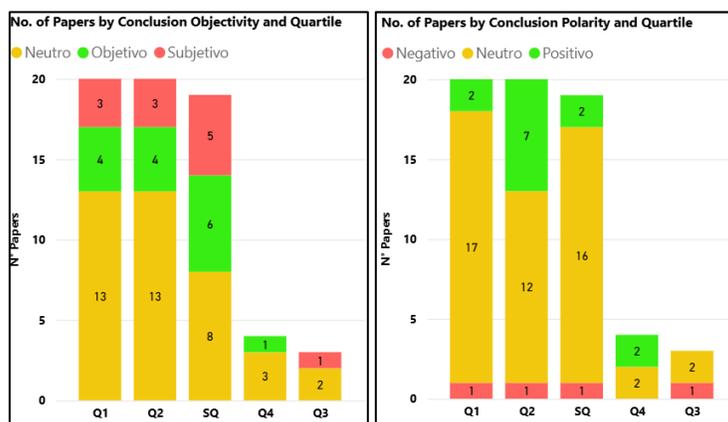


Fig. 10. Papers with high objectivity and moderate polarity

Figures 9 and 10 present a combined analysis of the objectivity and polarity of conclusions by quartile, evidenced through a scatter plot and two bar charts categorizing papers by objectivity and polarity. These visualizations enable the assessment of the consistency and potential bias of studies in relation to the journal quartile.

Papers published in Q1 journals exhibit high objectivity (0.52) and neutrality in polarity, whereas those in Q2 show similar objectivity (0.54) but a slight positive polarity tendency. Studies in non-quartile (NQ) journals demonstrate the highest objectivity (0.545), although with moderate polarity.

On the other hand, Q3 papers display lower objectivity (0.48) and negative polarity. The distribution reveals that most works with objective conclusions and neutral polarity are found in Q1

and NQ, suggesting a more rigorous and less biased approach in these quartiles.

Regarding the results presented by Andrés J. Aparcana-Tasayco and Javier Gamboa-Cruzado [71], they illustrate the objectivity and polarity of discussions and conclusions sorted by the most cited papers. It is noted that the most cited paper is subjective and exhibits positive polarity. Positive polarity indicates that the paper was formally written.

These findings reinforce the need to promote publication in higher-quartile journals to enhance argumentative quality. Moreover, they allow the replication of the analysis in other sectors such as education, healthcare, or logistics, across diverse geographical contexts and considering temporal variations. Furthermore, comparisons could be

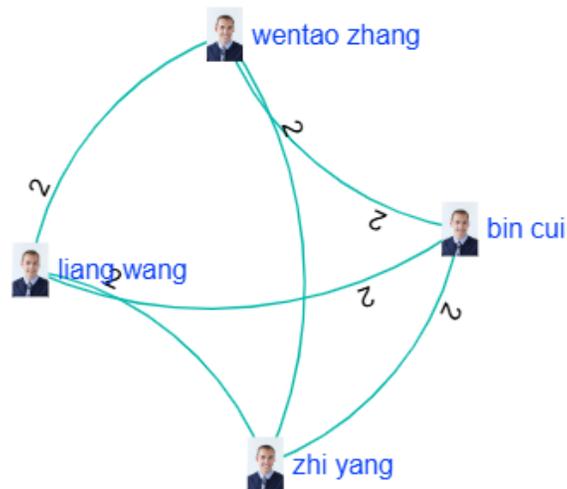


Fig. 11. Bibliometric network of authors and co-authors

drawn from studies conducted in post-pandemic contexts.

**RQ4: Which authors frequently appear as co-authors in scientific publications addressing the impact of Machine Learning on e-commerce?**

Figure 11 illustrates a bibliometric co-authorship network among researchers who have worked on topics related to the use of Machine Learning in e-commerce. The connections reflect the frequency of collaboration between authors.

For the results of RQ4, the network shows a strong interaction among four authors: Wentao Zhang, Bin Cui, Zhi Yang, and Liang Wang, who have co-authored at least two papers together. This cohesion indicates the existence of an active and possibly consolidated collaborative core in this field of study. The symmetry in the connections suggests stable collaboration relationships, which could reflect well-established research teams or sustained joint projects over time.

In contrast to our small network of authors and co-authors, Andrés J. Aparcana-Tasayco and Javier Gamboa-Cruzado [71] display a strong network of authors who shared seven papers across a single research group. Meanwhile, Nadia Giuffrida et al. [69] indicate that only five out of seventeen authors are interconnected; notably, these connected authors are also the most cited.

This co-authorship pattern can serve as a model to foster collaboration networks in other disciplines such as digital health, marketing, or smart logistics. Furthermore, identifying these key actors may facilitate the recognition of centers of excellence in ML applied to e-commerce. Lastly, future research could compare these networks across different geographic regions or before and after disruptive events such as the pandemic.

**RQ5: What are the main thematic categories addressed in the scientific literature on Machine Learning and its impact on e-commerce?**

Figure 12 and Table 8 present a thematic mapping based on density (degree of development) and centrality (degree of relevance) to categorize key areas in the scientific literature on Machine Learning applied to e-commerce. The topics are grouped into four quadrants: driving themes, basic themes, specialized themes, and marginal themes.

For the RQ5 results, it is evident that most of the identified topics fall within the marginal themes category, including "Machine Learning and Security," "ML Neural Networks," and "ML Fraud Detection," indicating low centrality and limited development compared to other areas. Only "Intelligent Detection" appears as a specialized theme, with high density but low centrality,

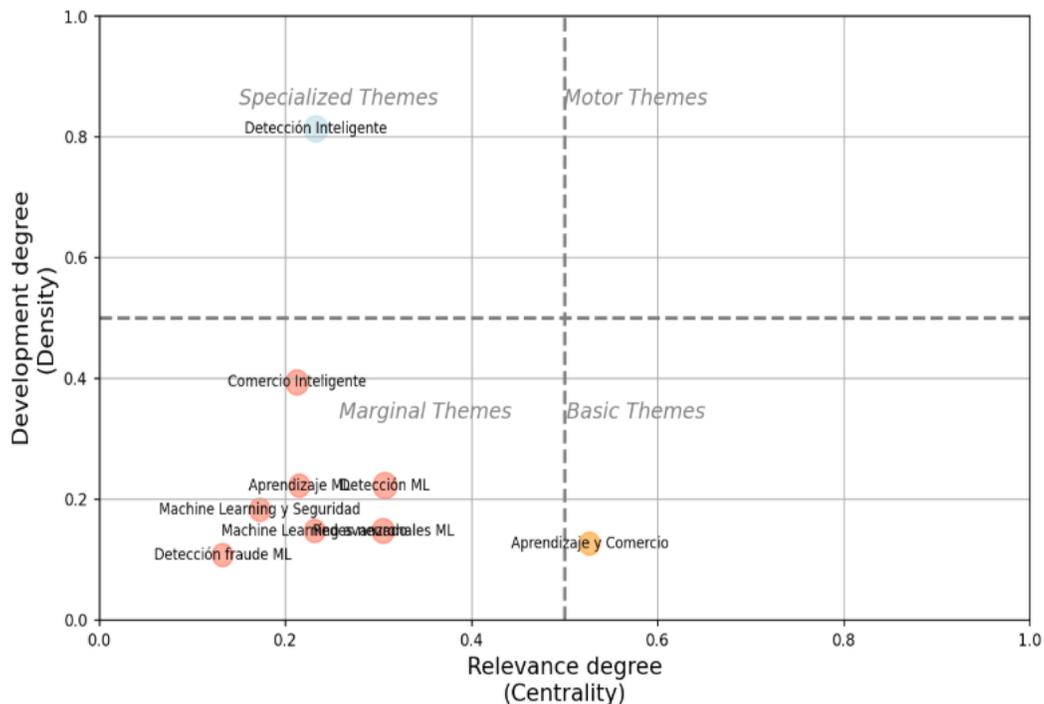


Fig. 12. Thematic categories

Table 8. Thematic Categories

Topic	Density	Centrality	Total Citations	Total Documents	Category
Intelligent Detection	0.81	0.23	152	13	Specialized
Smart Commerce	0.39	0.21	138	18	Marginal
ML Learning	0.22	0.22	118	14	Marginal
ML Detection	0.22	0.31	154	18	Marginal
Machine Learning and Security	0.18	0.17	118	14	Marginal
Advanced Machine Learning	0.15	0.23	118	14	Marginal
ML Neural Networks	0.15	0.31	136	16	Marginal
Learning and Commerce	0.13	0.53	116	10	Basic
ML Fraud Detection	0.11	0.13	118	14	Marginal

highlighting a highly developed but weakly connected field. "Learning and Commerce" stands out as the only basic theme, with high centrality but low development, suggesting its transversal importance and potential for future research.

According to Sabina-Cristiana Necula and Vasile-Daniel Păvăloaia [86], the main cluster—

holding the highest scores—is e-commerce, followed by the decision trees and collaborative filtering algorithms clusters, which also show high scores, indicating frequent citations by other network clusters.

Finally, they present clusters related to feature extraction and embeddings.

These findings can guide strategic efforts toward consolidating basic themes in less-studied regions or emerging sectors such as online education or digital tourism. Additionally, future research could focus on transforming marginal themes into driving themes through theoretical and empirical strengthening. Temporal analysis could also reveal how these categories evolve over time across different geographical contexts.

## 5 Conclusions and Future Research

The findings of this systematic literature review allow us to assert that the use of Machine Learning algorithms in e-commerce (RQ1) has been consolidated through the predominant application of neural networks (CNN, RNN), followed by clustering techniques and natural language processing, demonstrating a clear orientation toward service personalization, sentiment analysis, and fraud detection. This methodological diversity reflects the field's maturation and its applicability in predictive, classificatory, and operational tasks. Regarding the quartile levels of the journals addressing this topic (RQ2), there is a significant concentration of publications in Q1 and Q2, ensuring a high level of editorial and methodological quality.

This trend not only validates the scientific rigor of the studies but also suggests a growing interest in publishing in higher-impact journals. Concerning authors and collaboration networks (RQ4), consolidated nuclei of researchers with recurring co-authorship relationships were identified, indicating the existence of active research communities and opportunities for academic cooperation. Finally, regarding the thematic categories addressed (RQ5), the predominance of marginal topics stands out, although they present high development potential, such as "ML Detection" and "Machine Learning and Security." Only one theme was classified as basic: "Learning and Commerce," suggesting a clear opportunity for future studies that better articulate theoretical foundations with practical ML applications in e-commerce.

Future studies should explore the temporal evolution of marginal categories and experimentally validate their applicability in sectors

such as health, tourism, or digital education. It is also recommended to further investigate objectivity and polarity metrics to enhance the argumentative quality of published conclusions.

## References

1. **Abbas, K., Dong, S., Khan, A. (2023).** Point process based time sensitive personalised recommendation. *Procedia Computer Science*, Vol. 218, pp. 1791–1804. DOI: 10.1016/j.procs.2023.01.157.
2. **Aburbeian, A.M., Fernández-Veiga, M. (2024).** Secure internet financial transactions: A framework integrating multi-factor authentication and machine learning. *AI*, Vol. 5, No. 1, pp. 177–194. DOI: 10.3390/ai5010010.
3. **Adnan, M., Maboud, Y.E., Mahajan, D., Nair, P.J. (2021).** Accelerating recommendation system training by leveraging popular choices. *Proceedings of the VLDB Endowment*, Vol. 15, No. 1, pp. 127–140. DOI: 10.14778/3485450.3485462.
4. **Aljohani, K. (2024).** The role of last-mile delivery quality and satisfaction in online retail experience: An empirical analysis. *Sustainability*, Vol. 16, No. 11, Article 4743. DOI: 10.3390/su16114743.
5. **Alotaibi, F.M. (2023).** A machine-learning-inspired opinion extraction mechanism for classifying customer reviews on social media. *Applied Sciences*, Vol. 13, No. 12, Article 7266. DOI: 10.3390/app13127266.
6. **Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., Seaman, B. (2019).** Sales demand forecast in e-commerce using a long short-term memory neural network methodology. *Lecture Notes in Computer Science*, Vol. 11955, pp. 462–474. DOI: 10.1007/978-3-030-36718-3\_39.
7. **Bathla, G., Singh, P., Singh, R. K., Cambria, E., Tiwari, R. (2022).** Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. *Neural Computing and Applications*, Vol. 34, No. 22, pp. 20213–20229. DOI: 10.1007/s00521-022-07531-8.

8. **Beltzung, L., Lindley, A., Dinica, O., Hermann, N., & Lindner, R. (2020).** Real-time detection of fake-shops through machine learning. *Proceedings of the IEEE International Conference on Big Data*, pp. 2254–2263. DOI: 10.1109/BigData50022.2020.937 8204.
9. **Brackmann, C., Hütsch, M., Wulfert, T. (2023).** Identifying application areas for machine learning in the retail sector. *SN Computer Science*, Vol. 4, No. 5, Article 426. DOI: 10.1007/s42979-023-01888-w.
10. **Chanaa, A., El Faddouli, N. (2024).** Prerequisites-based course recommendation: Recommending learning objects using concept prerequisites and metadata matching. *Smart Learning Environments*, Vol. 11, No. 1, Article 16. DOI: 10.1186/s40561-024-00301-0.
11. **Ciubotariu, M., Socoliuc, M., Mihaila, S., Savchuk, D. (2019).** Marketing and management of innovations. *Marketing and Management of Innovations*, Vol. 3, No. 1, pp. 223–241. DOI: 10.21272/mmi.2019.3-17.
12. **Cotacallapa, H., Saboya, N., Canas-Rodrigues, P., a Iot Salas, R., Lopez-Gonzales, J. (2024).** A flat-hierarchical approach based on machine learning model for e-commerce product classification. *IEEE Access*, Vol. 12, pp. 72730–72745. DOI: 10.1109/ACCESS.2024.3400693.
13. **Cui, C., Saboya, N., Canas, P., Salas, R., López-Gonzales, J. (2022).** Intention adaptive graph neural network for category-aware session-based recommendation. *Lecture Notes in Computer Science*, Vol. 13246, pp. 150–165. DOI: 10.1007/ 978-3-031-00126-0\_10.
14. **Dewey, J., Ingram, C., Sanchez-Arias, R. (2023).** A supervised learning approach to assessing accounts receivable risk in small-to-medium enterprises. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 1805–1815. DOI: 10.46254/NA07.20220419.
15. **Fu, M., et al. (2020).** ICS-Assist: Intelligent customer inquiry resolution recommendation in online customer service for large e-commerce businesses. *Lecture Notes in Computer Science*, Vol. 12571, pp. 370–385. DOI: 10.1007/978-3-030-65310-1\_26.
16. **Govindaraj, M., Gnanasekaran, C., Sivakulanthay, T., Gnanamanickam, S.V. (2023).** Role of artificial intelligence across various media platforms: A quantitative investigation of media expert's opinion. *Journal of Law and Sustainable Development*, Vol. 11, No. 5, Article e1175. DOI: 10.55908/sdgs.v11i5.1175.
17. **Grob, G.L., Cardoso, Â., Liu, C.H.B., Little, D.A., Chamberlain, B.P. (2019).** A recurrent neural network survival model: Predicting web user return time. *Lecture Notes in Computer Science*, Vol. 11053, pp. 152–168. DOI: 10.1007/978-3-030-10997-4\_10.
18. **Grote, T. (2024).** Fairness as adequacy: A sociotechnical view on model evaluation in machine learning. *AI and Ethics*, Vol. 4, No. 2, pp. 427–440. DOI: 10.1007/s43681-023-00280-x.
19. **Gurumoorthy, K.S., Sanyal, S., Chaoji, V. (2021).** Think out of the package: Recommending package types for e-commerce shipments. *Lecture Notes in Computer Science*, Vol. 12461, pp. 290–305. DOI: 10.1007/978-3-030-67670-4\_18.
20. **Huang, R. (2023).** Improved content recommendation algorithm integrating semantic information. *Journal of Big Data*, Vol. 10, No. 1, Article 84. DOI: 10.1186/s40537-023-00776-7.
21. **Hwang, R., Kim, T., Kwon, Y., Rhu, M. (2020).** Centaur: A Chiplet-based, Hybrid Sparse-Dense Accelerator for Personalized Recommendations. *Proceedings of the 2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*, pp. 968–981. DOI: 10.1109/ISCA45697.2020.00083.
22. **In, Y., Yoon, K., Kim, K., Shin, K., Park, C. (2024).** Self-Guided Robust Graph Structure Refinement. *Proceedings of the ACM Web Conference*, pp. 697–708. DOI: 10.1145/3589334.3645522.
23. **Janavičiūtė, A., Liutkevičius, A., Dabužinskas, G., Morkevičius, N. (2024).** Experimental evaluation of possible feature

- combinations for the detection of fraudulent online shops. *Applied Sciences*, Vol. 14, No. 2, Article 919. DOI: 10.3390/app14020919.
24. **Ji, G.-P., Zhuge, M., Gao, D., Fan, D.-P., Sakaridis, C., Van Gool, L. (2023).** Masked Vision-language Transformer in Fashion. *Machine Intelligence Research*, Vol. 20, No. 3, pp. 421–434. DOI: 10.1007/s11633-022-1394-4.
  25. **Jiang, Y., et al. (2022).** Zoomer: Boosting Retrieval on Web-scale Graphs by Regions of Interest. *Proceedings of the IEEE 38th International Conference on Data Engineering (ICDE)*, pp. 2224–2236. DOI: 10.1109/ICDE53745.2022.00212.
  26. **Khmelnitsky, I., et al. (2023).** Analysis of Recurrent Neural Networks via Property-directed Verification of Surrogate Models. *International Journal on Software Tools for Technology Transfer*, Vol. 25, No. 3, pp. 341–354. DOI: 10.1007/s10009-022-00684-w.
  27. **Kumar, B., Roy, S., Singh, K., Pandey, S., Kumar, A., Sinha, A. (2023).** A static machine learning based evaluation method for usability and security analysis in e-commerce website. *IEEE Access*, Vol. 11, pp. 40488–40510. DOI: 10.1109/ACCESS.2023.3247003.
  28. **Li, S., Luo, H., Zhao, G., Tang, M., Liu, X. (2022).** Bi-directional Bayesian probabilistic model based hybrid grained semantic matchmaking for web service discovery. *World Wide Web*, Vol. 25, No. 2, pp. 445–470. DOI: 10.1007/s11280-022-01004-7.
  29. **Liu, J., Wu, T., Wu, J., Chen, Z., Gong, J., Chi, H. (2024).** Forecasting analysis of demand for agricultural products in e-commerce based on single forecasting model. *Frontiers in Artificial Intelligence and Applications*, Vol. 382, pp. 479–489. DOI: 10.3233/FAIA231332.
  30. **Liu, T., Yu, Z. (2022).** The analysis of financial market risk based on machine learning and particle swarm optimization algorithm. *EURASIP Journal on Wireless Communications and Networking*, Vol. 2022, No. 1, Article 31. DOI: 10.1186/s13638-022-02117-3.
  31. **Liu, Y., Hou, J., Zhao, W. (2024).** Deep learning and user consumption trends classification and analysis based on shopping behavior. *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp. 1–23. DOI: 10.4018/JOEUC.340038.
  32. **Liu, Z., Yeh, W.-C., Lin, K.-Y., Lin, C.-S., Chang, C.-Y. (2024).** Machine learning based approach for exploring online shopping behavior and preferences with eye tracking. *Computer Science and Information Systems*, Vol. 21, No. 2, pp. 593–623. DOI: 10.2298/CSIS230807077L.
  33. **Ma, M., Lei, Y., Li, Y., Wang, C. (2020).** Nowhere to hide: Anonymous user recognition based on multidimensional trajectory set. *Journal of Physics: Conference Series*, Vol. 1550, No. 3, Article 032100. DOI: 10.1088/1742-6596/1550/3/032100.
  34. **Maashi, M., Alabduallah, B., Kouki, F. (2023).** Sustainable financial fraud detection using Garra Rufa fish optimization algorithm with ensemble deep learning. *Sustainability*, Vol. 15, No. 18, Article 13301. DOI: 10.3390/su151813301.
  35. **Mahendra, Y., Bolla, B. (2024).** Unveiling the power of knowledge graph embedding in knowledge aware deep recommender systems for e-commerce: A comparative study. *Procedia Computer Science*, Vol. 235, pp. 1364–1375. DOI: 10.1016/j.procs.2024.04.128.
  36. **Mingxun, Z., Jiewu, Y., Zhigang, M., & Yanping, W. (2023).** Deep forest-based e-commerce recommendation attack detection model. *Security and Communication Networks*, Vol. 2023, Article 8413247. DOI: 10.1155/2023/8413247.
  37. **Mohamed, A., Abuoda, G., Ghanem, A., Kaoudi, Z., & Abounaga, A. (2022).** RDFFrames: Knowledge graph access for machine learning tools. *VLDB Journal*, Vol. 31, No. 2, pp. 321–346. DOI: 10.1007/s00778-021-00690-5.
  38. **N, N., & J, C. (2024).** Sentence classification using attention model for e-commerce product review. *Journal of Computer Science*, Vol. 20, No. 5, pp. 535–547. DOI: 10.3844/jcssp.2024.535.547.

- 39. Navaneethan, M., & Janakiraman, S. (2023).** An optimized deep learning model to ensure data integrity and security in IoT-based e-commerce blockchain application. *Journal of Intelligent and Fuzzy Systems*, Vol. 44, No. 5, pp. 8697–8709. DOI: 10.3233/JIFS-220743.
- 40. Nishimura, N., Sukegawa, N., Takano, Y., & Iwanaga, J. (2023).** Predicting online item-choice behavior: A shape-restricted regression approach. *Algorithms*, Vol. 16, No. 9, Article 415. DOI: 10.3390/a16090415.
- 41. Niu, C., et al. (2020).** Billion-scale federated learning on mobile clients. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 1–14. DOI: 10.1145/3372224.3419188.
- 42. Nkongolo, M., Tokmak, M. (2023).** Zero-day threats detection for critical infrastructures. In *Communications in Computer and Information Science*, Vol. 1878, pp. 32–47. DOI: 10.1007/978-3-031-39652-6\_3.
- 43. Pun, N.S., Sonbhadra, S.K. (2021).** Recommending best course of treatment based on similarities of prognostic markers. *arXiv Computer Science*. Retrieved from <http://arxiv.org/abs/2107.07500>.
- 44. Quang, L.N.T., Thao, U.N.T., Nhi, A.D.V., Phuong, H.N.T. (2022).** E-commerce price suggestion algorithm – A machine learning application. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 2598–2609. DOI: 10.46254/IN02.20220609.
- 45. Rodrigues, D., da Silva, A. (2021).** A study on machine learning techniques for the schema matching network problem. *Journal of the Brazilian Computer Society*, Vol. 27, No. 1, Article 14. DOI: 10.1186/s13173-021-00119-5.
- 46. Samonte, P.M.J., Britanico, E., Antonio, K.E.M., De la Vega, J.E.J., Espejo, T.J.P., Samonte, D.C. (2022).** Applying deep learning for the prediction of retail store sales. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 1–11. DOI: 10.46254/AF03.20220028.
- 47. Singh, M.K., Rishi, O.P., Singh, A.K., Singh, P., Choudhary, P. (2021).** Implementation of knowledge-based collaborative filtering and machine learning for e-commerce recommendation system. *Journal of Physics: Conference Series*, Vol. 2007, No. 1, Article 012032. DOI: 10.1088/1742-6596/2007/1/012032.
- 48. Špicas, R., Neifaltas, A., Kanapickienė, R., Keliuotytė-Staniulėnienė, G., Vasiliauskaitė, D. (2023).** Estimating the acceptance probabilities of consumer loan offers in an online loan comparison and brokerage platform. *Risks*, Vol. 11, No. 7, Article 138. DOI: 10.3390/risks11070138.
- 49. Tax, N., et al. (2021).** Machine learning for fraud detection in e-commerce: A research agenda. *Communications in Computer and Information Science*, Vol. 1482, pp. 30–54. DOI: 10.1007/978-3-030-87839-9\_2.
- 50. Tran, D.T., Huh, J.-H. (2023).** New machine learning model based on the time factor for e-commerce recommendation systems. *The Journal of Supercomputing*, Vol. 79, No. 6, pp. 6756–6801. DOI: 10.1007/s11227-022-04909-2.
- 51. Tsuboi, Y., Sakai, Y., Shimizu, R., Goto, M. (2024).** Multiple treatment effect estimation for business analytics using observational data. *Cogent Engineering*, Vol. 11, No. 1. DOI: 10.1080/23311916.2023.2300557.
- 52. Umer, S., Mohanta, P.P., Rout, R.K., Pandey, H. M. (2021).** Machine learning method for cosmetic product recognition: A visual searching approach. *Multimedia Tools and Applications*, Vol. 80, No. 28–29, pp. 34997–35023. DOI: 10.1007/s11042-020-09079-y.
- 53. Upreti, R., Lind, P.G., Elmokashfi, A., Yazidi, A. (2024).** Trustworthy machine learning in the context of security and privacy. *International Journal of Information Security*, Vol. 23, No. 3, pp. 2287–2314. DOI: 10.1007/s10207-024-00813-3.
- 54. Walter, Y. (2024).** The rapid competitive economy of machine learning development: A discussion on the social risks and benefits. *AI and Ethics*, Vol. 4, No. 2, pp. 635–648. DOI: 10.1007/s43681-023-00276-7.

55. **Wang, W. (2024).** A IoT-based framework for cross-border e-commerce supply chain using machine learning and optimization. *IEEE Access*, Vol. 12, pp. 1852–1864. DOI: 10.1109/ACCESS.2023.3347452.
56. **Wasilewski, A. (2024).** Functional framework for multivariate e-commerce user interfaces. *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 19, No. 1, pp. 412–430. DOI: 10.3390/jtaer19010022.
57. **Xie, L., Liu, J., Wang, W. (2024).** Predicting sales and cross-border e-commerce supply chain management using artificial neural networks and the Capuchin search algorithm. *Scientific Reports*, Vol. 14, No. 1, Article 13297. DOI: 10.1038/s41598-024-62368-6.
58. **Xu, J., Mu, S. (2022).** Research on the construction of cross-border e-commerce logistics service system based on machine learning algorithms. *Discrete Dynamics in Nature and Society*, Vol. 2022, Article 3943869. DOI: 10.1155/2022/3943869.
59. **Xu, Z., Chu, C., Song, S. (2024).** An effective federated recommendation framework with differential privacy. *Electronics*, Vol. 13, No. 8, Article 1589. DOI: 10.3390/electronics13081589.
60. **Zhang, G., Li, Z., Huang, J., Wu, J., Zhou, C. (2022).** eFraudCom: An e-commerce fraud detection system via competitive graph neural networks. *ACM Transactions on Information Systems*, Vol. 40, No. 3, pp. 1–29. DOI: 10.1145/3474379.
61. **Zhang, L., Wang, M., Zhou, X., Wu, X., Cao, Y., Xu, Y. (2023).** Dual graph multitask framework for imbalanced delivery time estimation. *Lecture Notes in Computer Science*, Vol. 13946, pp. 606–618. DOI: 10.1007/978-3-031-30678-5\_46.
62. **Zhang, S., Zhang, X., Lau, J. H., Chan, J., Paris, C. (2021).** Less is more: Rejecting unreliable reviews for product question answering. *Lecture Notes in Computer Science*, Vol. 12459, pp. 567–583. DOI: 10.1007/978-3-030-67664-3\_34.
63. **Zhang, W., et al. (2021).** GRAIN. *Proceedings of the VLDB Endowment*, Vol. 14, No. 11, pp. 2473–2482. DOI: 10.14778/3476249.3476295.
64. **Zhang, Y., et al. (2022).** PICASSO: Unleashing the potential of GPU-centric training for wide-and-deep recommender systems. *Proceedings of the 38th International Conference on Data Engineering (ICDE)*, pp. 3453–3466. DOI: 10.1109/ICDE53745.2022.00324.
65. **Zhao, S., Yin, Z., Xie, P. (2024).** Multi-angle perception and convolutional neural network for service quality evaluation of cross-border e-commerce logistics enterprise. *Peer J Computer Science*, Vol. 10, Article e1911. DOI: 10.7717/peerj-cs.1911.
66. **Zhou, S., Hudin, N.S. (2024).** Advancing e-commerce user purchase prediction: Integration of time-series attention with event-based timestamp encoding and graph neural network-enhanced user profiling. *PLOS ONE*, Vol. 19, No. 4, Article e0299087. DOI: 10.1371/journal.pone.0299087.
67. **Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S.G. (2009).** Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology*, Vol. 51, No. 1, pp. 7–15. DOI: 10.1016/j.infsof.2008.09.009.
68. **Cárdenas-Quispe, A., Vergaray-Mezarina, R., Gamboa-Cruzado, J. (2021).** Machine learning para la detección de malware en Android: Revisión sistemática de la literatura. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, No. E45, pp. 318–331.
69. **Giuffrida, N., Fajardo-Calderin, J., Masegosa, A. D., Werner, F., Steudter, M., Pilla, F. (2022).** Optimization and machine learning applied to last-mile logistics: A review. *Sustainability*, Vol. 14, No. 9. DOI: 10.3390/su14095329.
70. **Alegre-Veliz, R., et al. (2022).** Machine learning for feeling analysis in Twitter communications: A case study in HEYDRUI, Perú. *International Journal of Interactive Mobile Technologies*, Vol. 16, No. 24, pp. 126–142. DOI: 10.3991/ijim.v16i24.35493.

- 71. Aparcana-Tasayco, A.J., Gamboa-Cruzado, J. (2022).** Machine learning for management in software-defined networks: A systematic literature review. *IEIE Transactions on Smart Processing and Computing*, Vol. 11, No. 6, pp. 400–411. DOI: 10.5573/IEIESPC.2022.11.6.400.
- 72. Juro-Barrios, J., Gamboa-Cruzado, J., Baylon, A. R., Jurado, C.V. (2023).** Practical validation in a neutrosophic environment of the NEBS methodology for the optimization of SME financing through machine learning. *International Journal of Neutrosophic Science*, Vol. 20, No. 3, pp. 137–149. DOI: 10.54216/IJNS.200313.
- 73. Monti, D., Rizzo, G., Morisio, M. (2021).** A systematic literature review of multicriteria recommender systems. *Artificial Intelligence Review*, Vol. 54, No. 1, pp. 59–120. DOI: 10.1007/s 10462-020-09851-4.
- 74. Valdivia, J.R., Gamboa-Cruzado, J., de la Cruz Vélez de Villa, P. (2023).** Systematic literature review on machine learning and its impact on APIs deployment. *Computación y Sistemas*, Vol. 27, No. 4, pp. 1107–1124. DOI: 10.13053/CyS-27-4-4371.
- 75. Gamboa-Cruzado, J., Crisostomo-Castro, R., Vilabuleje, J., López-Goycochea, J., Valenzuela, J. N. (2024).** Heart attack prediction using machine learning: A comprehensive systematic review and bibliometric analysis. *Journal of Theoretical and Applied Information Technology*, Vol. 102, No. 5, pp. 1930–1944.
- 76. Veeramanju, K. T. (2023).** A systematic review on machine learning algorithms for customer satisfaction classification in various fields. *International Journal of Management, Technology, and Social Sciences*, Vol. 8, No. 3, pp. 326–339. DOI: 10.47992/ijmts.2581.6012.0305.
- 77. Fedorko, R., Král, Š., Fedorko, I. (2022).** Artificial intelligence and machine learning in the context of e-commerce: A literature review. *Lecture Notes in Networks and Systems*, Vol. 461, pp. 1067–1082. DOI: 10.1007/978-981-19-2130-8\_82.
- 78. Demircan, M., Seller, A., Abut, F., Akay, M.F. (2021).** Developing Turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering*, Vol. 2, pp. 202–207. DOI: 10.1016/j.ijcce.2021.11.003.
- 79. Mutemi, A., Bacao, F. (2024).** E-commerce fraud detection based on machine learning techniques: Systematic literature review. *Big Data Mining and Analytics*, Vol. 7, No. 2, pp. 419–444. DOI: 10.26599/BDMA.2023.9020023.
- 80. Rodrigues, V.F., Policarpo, L., Silveira, D., Righi, R., Costa, C. (2022).** Fraud detection and prevention in e-commerce: A systematic literature review. *Electronic Commerce Research and Applications*, Vol. 56, Article 101207. DOI: 10.1016/ j.elerap.2022.101207.
- 81. Liu, H., Zhao, J., Zhou, L., Yang, J., Liang, K. (2024).** Intelligent performance evaluation of e-commerce express services using machine learning: A case study with quantitative analysis. *Expert Systems with Applications*, Vol. 240, Article 122511. DOI: 10.1016/j.eswa.2023.122511.
- 82. Mustajab, S. (2020).** Machine learning techniques: A systematic review. *Invertis Journal of Science and Technology*, Vol. 13, No. 3, p. 159. DOI: 10.5958/ 2454-762x.2020.00016.5.
- 83. Micol-Policarpo, L., et al. (2021).** Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, Vol. 41, Article 100414. DOI: 10.1016/j.cosrev.2021.100414.
- 84. Huang, H., Zavareh, A.A., Mustafa, M.B. (2023).** Sentiment analysis in e-commerce platforms: A review of current techniques and future directions. *IEEE Access*, Vol. 11, pp. 90367–90382. DOI: 10.1109/ACCESS.2023 .3307308.
- 85. Daza, A., González-Rueda, N.D., Aguilar-Sánchez, M.S., Robles-Espíritu, W.F., Chauca-Quiñones, M.E. (2024).** Sentiment analysis on e-commerce product reviews using machine learning and deep learning algorithms: A bibliometric analysis and systematic literature review, challenges and future works. *International Journal of*

- Information Management Data Insights, Vol. 4, No. 2, Article 100267. DOI: 10.1016/j.jjime.2024.100267.
- 86. Necula, S.C., Păvăloaia, V.D. (2023).** AI-driven recommendations: A systematic review of the state of the art in e-commerce. *Applied Sciences*, Vol. 13, No. 9. DOI: 10.3390/app13095531.
- 87. Erdoğan, U. (2023).** A systematic review on the use of artificial intelligence in e-commerce. *Toplum, Ekonomi ve Yönetim Dergisi*, Vol. 4 (Özel), pp. 184–197. DOI: 10.58702/teyd.1357551.
- 88. Cano, J.A., Londoño-Pineda, A., Rodas, C. (2022).** Sustainable logistics for e-commerce: A literature review and bibliometric analysis. *Sustainability*, Vol. 14, No. 19. DOI: 10.3390/su141912247.
- 89. Jain, P.K., Pamula, R., Srivastava, G. (2021).** A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, Vol. 41, Article 100413. DOI: 10.1016/j.cosrev.2021.100413.
- 90. Duarte, V., Zuniga-Jara, S., Contreras, S. (2022).** Machine learning and marketing: A systematic literature review. *IEEE Access*, Vol. 10, pp. 93273–93288. DOI: 10.1109/ACCESS.2022.3202896.

*Article received on 19/01/2025; accepted on 04/05/2025.*

*\*Corresponding author is Blanca Cecilia López-Ramírez.*