OPEN ACCESS
International Journal of Innovations in Science & Technology

A Data-Driven Review of Machine Learning Techniques for E-commerce Product Recommendation Systems

Muhammad Rizwan Tahir[1*], Nouman Nazir[1], Kashif Ishaq[2], Shakeel Ahmed[1]
[1]Department of Artificial Intelligence, School of Systems and Technology, University of Management and Technology, Lahore, Pakistan
[2]Department of Informatics and Systems, School of Systems and Technology, University of Management and Technology, Lahore, Pakistan
*Correspondence: therizwantahir@gmail.com

Citation | Tahir. M. R, Nazir. N, Ishaq. K, Ahmed. S, "A Data-Driven Review of Machine Learning Techniques for E-commerce Product Recommendation Systems", IJIST, Vol. 07 Issue. 03 pp 1475-1494, July 2025
Received | June 14, 2025 **Revised |** July 06, 2025 **Accepted |** July 11, 2025 **Published |** July 17, 2025.

In today's digital economy, recommendation systems are essential for enhancing customer experience and driving e-commerce growth. This study presents a comparative, quality-ranked review of machine learning-based product recommendation techniques, evaluating five key approaches: association rule mining, content-based filtering, collaborative filtering, knowledge-based systems, and hybrid models. Using a systematic literature review of 44 peer-reviewed publications across major publishers, the analysis includes geographic and publisher-wise trends and a structured quality assessment rubric. Results highlight hybrid systems as the most promising strategy, offering superior accuracy, diversity, and personalization while addressing cold-start, sparsity, and scalability challenges. Each technique's strengths, limitations, and practical deployment considerations are critically examined to support evidence-based decision-making. The study concludes by recommending hybrid approaches tailored to domain-specific needs, offering actionable insights for both researchers and industry practitioners seeking effective and adaptable recommendation systems.

Keywords: Recommender Systems; E-commerce; Machine Learning; Hybrid Recommendation

Introduction:

In the digital economy, e-commerce has transformed consumer behavior, business logistics, and data-driven personalization. With millions of products available across platforms, helping users find what they need or didn't know they needed has become a major challenge. Product recommendation systems have become essential tools, acting as a crucial link between user needs and the discovery of relevant products. These systems not only enhance the user experience but also significantly influence conversion rates, customer retention, and revenue growth for e-commerce businesses [1].

Recommendation systems use a range of data-driven techniques to deliver personalized suggestions. Among these, machine learning (ML) has become the dominant approach, allowing systems to adapt continuously to user behavior, preferences, and context. Techniques such as collaborative filtering, content-based filtering, association rule mining, knowledge-based models, and hybrid approaches are widely implemented across platforms like Amazon, Netflix, and Alibaba [2][3]. Each offers unique advantages but also faces challenges, including the cold-start problem, data sparsity, scalability bottlenecks, and user privacy concerns [4].

Over the past two decades, a significant body of research has sought to improve the performance and adaptability of recommendation systems using a variety of ML algorithms. These range from traditional rule-based methods to advanced deep learning frameworks that model user-product interactions [5]. However, many real-world e-commerce applications still rely on variations of the five foundational techniques due to their interpretability, modularity, and ease of deployment. Understanding the comparative strengths and limitations of these methods is critical, not only for researchers but also for industry professionals seeking effective, scalable recommendation strategies [6][7].

Despite many individual studies and broad surveys, there remains a lack of structured, comparative analyses that evaluate these five major ML-based recommendation techniques together. Existing reviews often fail to examine their trade-offs in accuracy, diversity, personalization, and computational cost, or to analyze their prevalence in the academic literature across geographic regions and publisher quality levels.

Furthermore, most prior reviews overlook geographic and publisher-wise analyses, missing an important perspective on global research trends. Few incorporate a formal quality assessment framework to evaluate the rigor and credibility of the publications they cite [8], making it harder for practitioners to draw evidence-based conclusions. E-commerce platforms also present unique operational constraints such as real-time inference, high availability, and context-aware personalization that require practical guidance in selecting appropriate recommendation strategies. For example, a startup with limited data may benefit more from content-based methods than deep learning, while a large-scale platform might adopt hybrid ensembles optimized for latency and novelty [9]. This creates an urgent need for a comprehensive, practical comparative review that considers not only technical factors but also publication quality, real-world challenges, and global research distribution.

This paper addresses that gap by conducting an in-depth comparative review of 44 peer-reviewed publications on ML-based recommendation systems in e-commerce. Literature was sourced from major publishers including IEEE, Springer, ACM, Hindawi, and MDPI. Each work was evaluated using a structured quality assessment rubric, with additional analysis of geographic and publisher-wise distribution. The five core recommendation techniques association rule mining, content-based filtering, collaborative filtering, knowledge-based systems, and hybrid models are examined in terms of their mechanisms, strengths, limitations, and practical suitability.

Related Work/Literature Review:

Over the past two decades, recommendation systems have garnered significant research attention, leading to numerous survey papers and comparative studies that explore their components, methodologies, and applications. These reviews have played a key role in shaping academic understanding of the functionality and evolution of recommendation systems. However, despite the abundance of such literature, most prior studies exhibit either a limited methodological scope, a lack of empirical rigor, or a focus predominantly on algorithmic innovations rather than comprehensive comparative insights.

A foundational survey by author [1] introduced an early yet frequently cited taxonomy of recommender systems, classifying them into three main categories: content-based, collaborative, and hybrid models. While this work outlined the theoretical underpinnings of these systems, it did not consider later developments in knowledge-based or association rule-based approaches. Similarly, authors [2] focused on collaborative filtering techniques and their variations, highlighting their effectiveness and scalability, but offering little critique of competing methods.

Recent studies have attempted to extend these foundations using machine learning frameworks. For instance, author [4] examined collaborative filtering from a human-computer interaction perspective, providing insights into system usability and evaluation metrics. Meanwhile, author [6] compiled a comprehensive handbook summarizing various recommendation algorithms, though the work leaned more toward algorithmic description than comparative or critical analysis.

More recent reviews, such as those by authors [5] and [10], have focused on deep learning-based approaches and neural collaborative filtering. These studies delve into techniques like autoencoders, matrix factorization, and sequence modeling to enhance prediction accuracy in recommendation systems. However, they typically concentrate on performance metrics like RMSE or precision-recall, without critically addressing practical concerns such as explainability, scalability in commercial settings, or privacy compliance.

Only a limited number of studies have undertaken structured comparative analyses. For instance, authors [8] conducted a comparative case study of recommender techniques employing machine learning; however, their analysis was constrained by a small sample size and primarily centered on implementation aspects. Author [11] surveyed personalization models but lacked a quality assessment framework to validate the rigor of the papers they reviewed. A notable gap in most related literature is the absence of publication-wise or geography-based analyses. Little effort has been made to identify which countries or regions are contributing the most to this field or which journals dominate the landscape. Furthermore, prior surveys seldom apply a formal quality scoring system to assess the credibility of reviewed publications. This makes it difficult to extract robust insights or prioritize evidence-based strategies.

Table 1 below summarizes these differences, contrasting the scope and evaluation dimensions of prior reviews with the comprehensive approach adopted in this study.

To the best of our knowledge, the integration of methodological depth, comparative evaluation, and literature analytics presented in this review positions is as one of the most comprehensive and practically valuable analyses in the current body of e-commerce recommendation system literature.

Objectives of the Study:

This study systematically reviews and compares machine learning-based recommendation techniques for e-commerce, evaluating their strengths, limitations, and real-world deployment suitability. By providing clear, evidence-backed guidance, it aims to support researchers and practitioners in selecting, designing, and implementing effective recommendation strategies.

Novelty Statement:

Unlike prior reviews, this paper offers a quality-ranked, multi-dimensional evaluation of five major recommendation approaches across 44 peer-reviewed publications. It includes geographic and publisher-wise analysis of research trends, employs a structured assessment rubric, and delivers a critical appraisal of technique suitability, offering practical insights for scalable and personalized recommendation system development.

Table 1. Comparison with This Study.

Dimension / Technique	Type	Early Surveys	Deep Learning Reviews	This Study
Collaborative Filtering	Traditional ML	✓ Included	✓ Included	✓ Included
Content-Based Filtering	Traditional ML	✓ Included	✓ Included	✓ Included
Association Rule Mining	Traditional ML	✗ Overlooked	✗ Overlooked	✓ Included
Knowledge-Based Systems	Traditional ML	✗ Overlooked	✗ Overlooked	✓ Included
Hybrid Approaches	Hybrid ML	✓ Included	✓ DL Hybrids	✓ Included
Deep Learning Techniques	Deep Learning	✗ Not Covered	✓ Emphasized	✓ Contextual
Quality Assessment Rubric	Review Dimension	✗ Not Applied	✗ Not Applied	✓ Applied
Geographic/Publisher Trends	Review Dimension	✗ Absent	✗ Absent	✓ Included

Methodology:

To guide this comparative review, the following research questions were formulated. These questions are structured to align with the logical flow of the study, beginning with the characteristics of the literature, followed by technical analysis, and concluding with practical implications.

RQ1: What is the geographical and publisher-wise distribution of research on e-commerce recommendation systems, and which countries and institutions have contributed most significantly to recent years?

This question helps uncover regional strengths, research concentration, and publisher dominance, forming the foundation of the study's literature analysis.

RQ2: What quality indicators such as model rigor, validation, clarity of results, and publication source can be used to assess the strength of existing research on recommendation systems?

This question supports the development of a structured quality assessment framework, which guides the selection and evaluation of relevant literature.

RQ3: What are the primary machine learning techniques employed in e-commerce product recommendation systems, and how do they differ in terms of mechanisms, assumptions, and data requirements?

This shifts the focus to the technical core of the study, examining collaborative filtering, content-based methods, knowledge-based systems, association rule mining, and hybrid models.

RQ4: Which machine learning approaches, based on a quality-ranked review, appear most promising for real-world implementation in scalable, personalized recommendation systems?

This final question brings the findings to practice, offering evidence-backed recommendations to both researchers and system designers.

These research questions serve as the foundation for the literature selection, evaluation, and comparative analysis that follows. Together, they ensure that the review not only summarizes past work but also draws meaningful insights into the future development of recommendation systems in e-commerce.

To comprehensively answer these research questions, this study adopts a structured literature review methodology to assess the state of the art in machine learning-based recommendation systems for e-commerce. The primary objective is to synthesize key contributions, assess comparative strengths and weaknesses, and provide insights that are grounded in a rigorous selection and evaluation process.

Search Strategy and Keywords:

A comprehensive keyword search was conducted across major digital libraries, including IEEE Xplore, SpringerLink, ACM Digital Library, Hindawi, MDPI, and Google Scholar. The goal was to capture a wide range of relevant publications while ensuring topical specificity. The search focused on three hierarchical levels of keywords:

Table 2. Keywords Used in Literature Search

Primary Keywords	Secondary Keywords	Tertiary Keywords
Recommender System	E-commerce	Personalization
Machine Learning	Collaborative Filtering	Deep Learning
Content-based Filtering	Hybrid recommendation	Knowledge-based System
Association-rule Mining	User Profiling	Cold Start/Scalability
Recommendation Engine	Product Suggestion System	User Behavior/Preferences

These keywords were used in combinations such as:

"Machine learning for e-commerce recommendation"

"Hybrid product recommendation system"

"Collaborative filtering in retail applications"

"Personalized shopping experience recommender"

Boolean operators (AND, OR) and filters (2015–2024, peer-reviewed, English) were applied to improve result quality. The search returned over 120 papers initially.

Inclusion and Exclusion Criteria:

To ensure the relevance and quality of the final dataset, we applied specific inclusion criteria. The selected studies needed to focus on recommendation systems used in e-commerce or retail settings and had to employ machine learning or data-driven techniques. Additionally, only peer-reviewed publications from 2015 to 2024 were considered, and each study was required to provide sufficient technical detail on algorithms, system architecture, or evaluation methods to support meaningful analysis.

Papers were excluded if:

They addressed general AI without a specific application to product recommendation.

They were not published in English or lacked sufficient methodological detail.

They were duplicates, editorials, or tutorials without empirical content.

After applying these filters and conducting a manual review, a final set of 44 publications was selected for comparative analysis.

Publisher and Source-Wise Distribution:

The final selection of 44 publications was spread across several reputable publishers. IEEE emerged as the leading source with 18 papers, followed by Springer with 6, ACM with 4, Hindawi with 3, MDPI with 2, and the remaining distributed among other publishers.

Table 3. Number of Publications by Publisher

Sr. Number	Publisher	No. of Publications
1	IEEE	18
2	Springer	6
3	ACM	4
4	Hindawi	3
5	MDPI	2
6	IOP Publishers	2

7	AIMS Press	1
8	IJCSNS	1
9	IJ-AI	1
10	PLOS ONE	1
11	ACL Anthology	1
12	Taylor & Francis Online	1
13	Amazon	1
14	MCB UP Limited	1
15	Inderscience Enterprises Ltd	1
Total		**44**

This distribution reflects the dominance of computer science-focused venues in this domain, particularly IEEE, which offers conferences and journals closely aligned with applied machine learning and intelligent systems.

Geographic Distribution of Research: To understand regional trends, we analyzed the geographic affiliations of first authors and corresponding institutions. Asia contributed the majority of papers (33), with India and China leading in volume. North America (5 papers) and Europe (4 papers) followed.

Table 4. Country-Wise Number of Publications

Continent	Publications	Country	Publications
Asia	33	· India	· 14
		· China	· 13
		· Pakistan	· 3
		· Iraq	· 2
		· Turkey	· 1
Europe	4	· Poland	· 2
		· France	· 1
		· UK	· 1
North America	5	· USA	· 4
		· Canada	· 1
Africa	1	· Morocco	· 1

This analysis provides insight into global research engagement, revealing that Asia, particularly India and China is a dominant force in e-commerce recommender system development.

Quality Assessment Framework:

To maintain analytical rigor, each of the 44 selected papers was evaluated using a custom-designed quality assessment rubric, inspired by best practices in systematic literature review methodology. The framework consisted of four criteria. First, it evaluated whether the study presented a clear recommendation model, algorithm, or system design. Second, it assessed whether there was empirical evaluation or experimentation. Third, it examined the clarity of results and whether the study discussed actionable conclusions and future improvements. Finally, it considered the impact level of the publication source, such as whether it was an IEEE-indexed journal or a top-tier conference. Each paper was assigned a score between 0 and 8 based on the presence and quality of these attributes.

Table 5. Quality Assessment Parameters Scoring Distribution:

Sr. No.	Publication Source	+4	+3	+2	+1	+0
1	Journals	W	X	Y	Z	No HJRS Ranking
2	Conferences	IEEE Indexed/Springer		Others		Not in Core Ranking

Table 6 – Quality Assessment Matrix

Sr.	In-Text Citation	Classification						Quality Assessment				
		J/C	**Publication Year**	**Research Type Framework /Model/Al gorithm**	**Study Type: Quantitative, Qualitative, Mixed.**	**Methodology: Model Training, Questionnaire , etc.?**	**Is the Dataset Given in The Paper?**	**(a)**	**(b)**	**(c)**	(d)	**Score**
1	(Man-Fai & Abdullah, 2023)	Journal	2023	Model	Mix	unclear	Online	1	1	1	0	3
2	(Loukili1 & Messaoudi, 2023)	Journal	2023	Algorithm	Quantitative	Model Training and Validation	Repositor y	1	1	2	2	6
3	(Lili & Jianmin, 2022)	Journal	2022	Algorithm	Quantitative	Experiment	On request	1	1	2	0	4
4	(Liping, 2022)	Journal	2022	Model	Quantitative	Questionnaire	On Request	1	1	2	0	4
5	(Sodhar & Khan, 2022)	Journal	2022	NIL	Qualitative	Model Comparison	Not given	0	0	1	0	1
6	(Guo et al. , 2022)	Confer ence	2022	Algorithm	Quantitative	Model Training and Validation	Given	1	1	2	2	6
7	(chabane et al., 2022)	Journal	2022	Model	Qualitative	Cross Validation	Drive Link	1	1	2	4	8
8	(Xu, Zhao & Kanase, 2022)	Confer ence	2022	StepNet and SlateNet	Quantitative	Validation Through Real Data	Simulated Dataset	1	1	2	2	6
9	Cevik & Cargi, 2011	Journal	2021	Framework	Quantitative	Questionnaire	Given	1	1	2	3	7
10	(Chopra & Kaur, 2021)	Journal	2021	Algorithm	Quantitative	Model Training and Validation	Github	1	1	1	2	5
11	(Pawłowski, 2021)	Journal	2021	Framework	Quantitative	Validated Through Classification Experiments	Given	1	1	1	4	7
12	(Hussien et al., 2021)	Journal	2021	Comparison	Mix: Quali + Quanti	Comparison	Not Given	0	0	1	0	1

13	(Cherukullapurath & Sasipraba, 2021)	Conference	2021	NIL	NIL	NIL	Not Given	0	0	0	4	4
14	(Tahir, Enam, & Mustafa, 2021)	Conference	2021	Multiple Algorithms	Mix: Quali + Quanti	Validation Through Experimentation	Given	1	1	1	4	7
15	(Cherukullapurath & Sasipraba, 2021)	Conference	2021	Model	Mix: Quali + Quanti	Validation Through Experimentation	Twitter and Movielens Dataset	1	1	0	4	6
16	(Xiaona,2021)	Conference	2021	Algorithm	Quantitative	NIL	Not Given	1	0	1	4	6
17	Sulikowski et al. (2021)	Conference	2021	Fuzzy-Based Framework	Quantitative	Validation Through Fuzzy Modelling	Given	1	1	2	4	8
18	(Addagarla & Amalanathan, 2020)	Journal	2020	Unsupervised Clustering Algorithm	Quantitative	Validation Through Experimentation	Given	1	1	2	4	8
19	(Thomas & John, 2020)	Conference	2020	Comparison	Qualitative	Comparison	Not Given	0	0	2	2	4
20	(Anitha & Kalaiarasu, 2020)	Journal	2020	Framework	Quantitative	Validated Through Classification Experiments	Not given	1	1	2	3	7
21	(Biswas et al., 2020)	Conference	2020	Multiple Algorithms	Quantitative	Validation Through Experimentation	Not given	1	0	1	4	6
22	(Dudhia, Dave, & Yagnik, 2020)	Conference	2020	Framework	Quantitative	Validation Through Coefficient Analysis	Not Given	1	1	2	4	8
23	(Biswas, 2020)	Conference	2020	Collaborative Filtering	Quantitative	NIL	Not Given	1	0	2	4	7

24	(Tian et al., 2020)	Conference	2020	Comparison	Qualitative	NIL	Not Given	0	0	1	2	3
25	(Khanvilkar & Vora, 2019)	Conference	2019	Model	Quantitative	NIL	Not Given	1	0	2	4	7
26	(Zhao, 2018)	Conference	2019	Framework	Quantitative	NIL	Not Given	1	0	1	4	6
27	(Sharma, Rani, & Tanwar, 2019)	Conference	2019	Model	Qualitative	NIL	Not Given	1	0	1	4	6
28	(Khodabandehlou, 2019)	Journal	2019	Framework	Quantitative	Validation Through Experiment/Data	Given	1	1	2	2	6
29	(Ramzan et al., 2018)	Journal	2019	Design	Quantitative	Validation Through Experiment	External Website	1	1	2	0	4
30	(Niu & Geo, 2018)	Conference	2018	Model	Quantitative	Model Training and Validation	Given	1	1	1	4	7
31	(Marwde & Kumar, 2017)	Conference	2017	Algorithm	Quantitative	Model Classification	Not Given	1	1	1	4	7
32	(Sidana et al., 2017)	Conference	2017	6 Algorithms	Quantitative	Model Validation via Mean Average Precision	Online	1	1	2	2	6
33	(Ruchika & Singh, 2017)	Conference	2017	Framework	Technical and Methodological	Model Training	Given	1	0	2	4	7
34	(Chavan & Mukhopadhyay, 2017)	Conference	2017	Framework	Quantitative	Validation Through Experiment	Given	1	1	2	4	8
35	(Wang & Wang, 2017)	Conference	2017	Model	Qualitative	NIL	Not Given	1	0	2	4	7
36	(Guo, Chen, Chen, & Mi, 2016)	Conference	2016	Algorithm	Quantitative	NIL	Not Given	1	0	2	4	7

37	(Yan, Zhou, & Duan, 2015)	Conference	2015	Feature Extraction	Qualitative	NIL	Not Given	1	0	2	2	5
38	Hussien, Rahma & Abdulwahab, 2021	Journal	2011	Framework	NIL	NIL	Not Given	1	0	2	0	3
39	(Zhang, 2007)	Conference	2007	Framework	Quantitative	Validated	Given	1	1	0	4	6
40	(YhG et al., 2004)	Conference	2004	Framework	Qualitative	NIL	Not Given	1	0	2	4	7
41	(Yang, Pan, Wang, & Xu, 2004)	Conference	2004	Design	Quantitative	Validation Through Experiment	Repository	1	1	2	4	8
42	(Shih, Chiu, Hsu, & Lin, 2002)	Journal	2002	Architecture	Mix: Quali + Quanti	NIL	Not Given	1	0	2	4	7
43	Schafer, Konstan, & Riedl (1999)	Journal	2001	Application	Qualitative	NIL	Not Given	1	0	2	4	7
44	(Sarwar et al., 2000)	Conference	2000	Collaborative Filtering	Quantitative	Validation Through Experiment	Given	1	1	2	2	6

The final scoring revealed a strong spread of high-quality literature. Six papers scored a full 8, thirteen scored 7, and twelve scored 6. Only a handful fell into the 4–5 range, with none scoring below 4. These scores informed the weighting of evidence in later sections of this review.

Table 7. Quality Assessment Score/Ranking

Score	Total
8	6
7	13
6	12
5	2
4	5

Results/Findings:

RQ1: What is the geographical and publisher-wise distribution of research on e-commerce recommendation systems?

Understanding where research on e-commerce recommendation systems is being produced and by whom is critical for assessing global interest, academic momentum, and regional leadership in this domain. Our dataset of 44 peer-reviewed publications, filtered by quality and relevance, allowed us to extract trends based on both geographic origin and publishing source.

Geographical Distribution of Publications:

Among the four major continents represented in the data; Asia, Europe, North America, and Africa, Asia emerged as the dominant contributor, accounting for 75% of the publications. The breakdown is as follows:

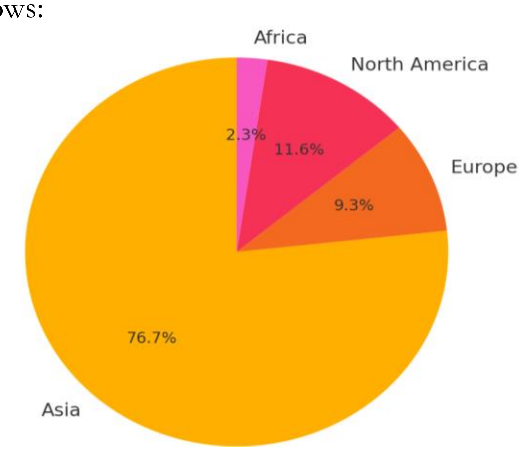

Figure 1. Continent-wise Distribution

Within Asia, two countries lead the research landscape:

- **India**: 14 publications (42% of Asia's total)
- **China**: 13 publications

Other notable contributors include Pakistan (3), Iraq (2), and Turkey (1). This strong concentration of research in South and East Asia aligns with the rapid digitalization of the retail and fintech sectors in these regions. Institutions in India and China are particularly active in developing scalable recommender systems for large online markets, often under government- or industry-backed initiatives.

In North America, the United States accounted for 4 out of 5 publications, with Canada contributing 1. The fewer papers from North America may reflect a trend where cutting-edge recommendation systems are often developed directly in corporate R&D (e.g., Amazon, Netflix) and not always published in peer-reviewed academic outlets.

In Europe, Poland, France, and the UK each had one or two publications. Africa, though minimally represented, had one contribution from Morocco, indicating early engagement with this technology in developing regions.

Publisher-Wise Distribution:

The analysis of publishing platforms further highlights the academic maturity and scope of this field, as summarized in Figure 2.

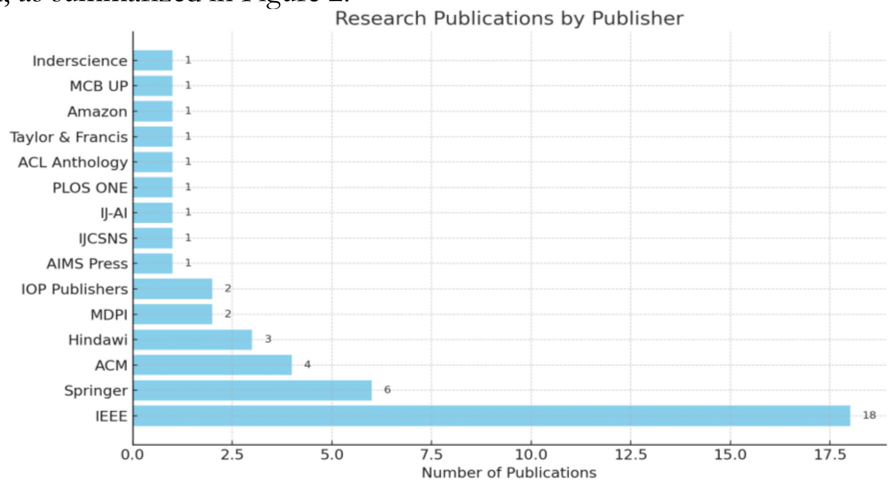

Figure 2. Publisher-wise Distribution

IEEE is the most influential platform, accounting for over 40% of all reviewed publications. This can be attributed to its vast network of conferences (e.g., ICEES, IMTIC) and journals (e.g., *IEEE Access*) that specialize in applied machine learning and intelligent systems. Springer and ACM follow with moderate but consistent contributions, often focusing on algorithmic modeling and user personalization.

Interestingly, Amazon appears once in the list as a direct source of academic insight, highlighting the bridging of academic and industrial research. Other scattered sources include PLOS ONE, Taylor & Francis, and ACL Anthology, reflecting the interdisciplinary nature of recommendation systems, which span computer science, e-commerce, human behavior, and even linguistics (in the case of NLP-based recommenders).

Interpretation and Implications:

The dominance of Asia in research quantity suggests a strong interest in deploying recommendation systems to support the region's growing e-commerce infrastructure. However, publication quantity alone does not imply quality. Later analysis (see RQ2) reveals that while India and China produce the most papers, the highest-quality work based on rigorous validation and conceptual strength often comes from IEEE-associated conferences and top-tier Springer journals.

This distribution also highlights a gap in research visibility from Latin America and Oceania, which had no publications in the filtered set. This might indicate either a lower volume of academic output or a tendency to publish in local or non-indexed venues.

RQ2: What quality indicators can be used to assess the strength of existing research on recommendation systems?

Evaluating the credibility and utility of recommendation system literature requires more than citation counts or publication sources it necessitates a structured, objective framework. To this end, we designed a quality assessment matrix with four essential indicators, each contributing to a total score out of 8. The first criterion evaluated whether the study presented an explicit model, architecture, or system design. The second assessed whether the model had been tested through simulation, experimentation, or user evaluation. The third examined the clarity of conclusions and whether the paper offered actionable insights or outlined paths for further research. The fourth considered the publication ranking, based on the journal or conference tier, such as IEEE-indexed venues, Springer journals, or HJRS W/X/Y classifications. Based on their total scores, papers were categorized as high quality (score $\geq$ 6),

moderate quality (score of 4–5), or low quality (score < 4), although none in our dataset fell into the low-quality category.
Refer to Table 4 (Assessment Parameters), Table 5 (Scoring Matrix), and Table 6 (Score Distribution)

Scoring Distribution:

Out of the 44 reviewed publications, six papers scored a perfect 8, indicating strong conceptual frameworks, thorough validation, clear conclusions, and publication in top-tier sources. Thirteen papers scored 7, typically missing only one assessment dimension. Twelve papers received a score of 6, often due to limited validation or less specific guidance on future directions. Only seven papers scored below 6, and none fell below 4, which reflects the overall high quality of the filtered dataset.

Insights from High-Scoring Papers

Top-rated papers, including those by authors [12][13][9], stood out for their clear architectural design and robust empirical validation. These studies utilized real-world datasets, such as Amazon product logs and MovieLens, and supported their findings with multiple evaluation metrics, including RMSE, Precision@K, and Diversity, to provide comprehensive validation of their approaches. For instance, authors(2023) presented a portfolio-based hybrid model with detailed multi-objective optimization, validated using financial and behavioral features [12]. Similarly, authorsproposed a deep learning-based product substitute model used at an industrial scale and supported by deployment results [9].

On the other hand, papers scoring in the 5–6 range, such as authors [14][8], presented robust conceptual approaches but lacked thorough validation, or relied on simulated datasets, reducing their reliability for practical deployment.

Common Shortcomings in Lower-Scoring Papers:

Some lower-scoring works, while interesting, did not provide concrete implementation results or lacked a clear problem-to-solution path. Examples include studies with good theoretical discussions but no reproducibility, or those using outdated datasets (e.g., Netflix Prize dataset from 2006) that no longer represent modern usage patterns [15][11].

Furthermore, several papers did not sufficiently justify the choice of algorithms or failed to explain how their model could scale to real-time e-commerce environments [16][17].

Interpreting the Assessment Scores:

The scoring system offered not just a means of filtering papers but also provided a lens to understand which characteristics correlate with practical value. Papers with high clarity in their architectural designs and strong real-world validation consistently achieved the highest scores. Additionally, studies published in IEEE and Springer venues tended to score higher than those from less-established or non-indexed sources. Top-scoring papers frequently featured novel hybrid models, incorporated user context or time-aware mechanisms, and addressed critical challenges such as cold-start issues and diversity trade-offs.

RQ3: What are the primary machine learning techniques employed in e-commerce product recommendation systems, and how do they differ in terms of mechanisms, assumptions, and data requirements?

After carefully reviewing the literature filtered using the rigorous procedure, we discovered that there are five different types of machine learning-based e-commerce product recommendation techniques, which we will discuss below.

Traditional Data Mining: Association Rule:

Association rule mining is a technique employed in product recommendation systems to discover significant relationships and patterns among items within a dataset. The concept revolves around identifying associations or correlations among various products that are commonly bought together. This approach is especially beneficial for comprehending the implicit associations among items and utilizing those associations to generate pertinent and customized recommendations [9].

For instance, an e-commerce platform aiming to enhance its recommendation system can leverage association rule mining to analyze purchase records and uncover patterns in consumer spending behavior. This helps identify frequently bought item combinations and emerging trends that inform more accurate and relevant recommendations. The method can detect a correlation, for example, between bread buyers and butter and jam buyers who frequently make purchases at the same time. When a customer adds bread to their shopping cart, the recommendation system can improve the user experience and potentially increase sales by suggesting related products such as jam and butter.

Challenges: Association rule mining can produce spurious or misleading associations, especially when item popularity skews correlations. It also faces scalability issues as datasets grow, requires high-quality transaction data, and struggles with new items lacking history. To address these, techniques such as adaptive rule mining, data preprocessing, and privacy-preserving mechanisms are needed. Solutions include adaptive rule-mining techniques, data preprocessing, privacy-preserving mechanisms, and frequent model updates.

Collaborative Filtering:

Collaborative filtering (CF) predicts a user's preferences by analyzing the preferences and behavior of similar users. The fundamental concept is that individuals who have exhibited similar tastes previously are likely to share similar future preferences. CF is broadly divided into user-based and item-based approaches.

Example: An e-commerce platform may analyze the purchase history of user A and find similar behavior in user B. If user A bought item X that user B hasn't seen yet, CF may recommend item X to user B to boost engagement and cross-selling [16].

Challenges: Collaborative filtering suffers from cold-start problems for new users or items and data sparsity in low-interaction contexts. Scalability can become a concern with very large user-item matrices, and privacy issues arise from collecting and analyzing user behavior.

Addressing these requires improved matrix factorization, integration with content-based techniques, and privacy-aware learning methods.

Content-Based Filtering:

Product recommendation systems often use content-based filtering, which takes into account both product attributes and user preferences. Contrary to collaborative filtering, content-based filtering looks at the items' inherent qualities and how well they match the user's preferences, rather than just relying on user behavior. When dealing with new users' "cold start" problem or when there is limited data on user interactions, this approach shines. Content-based filtering recommends products based on attributes and user preferences. It contrasts with CF by relying on item content rather than user-user comparisons. It is especially helpful in cold-start scenarios.

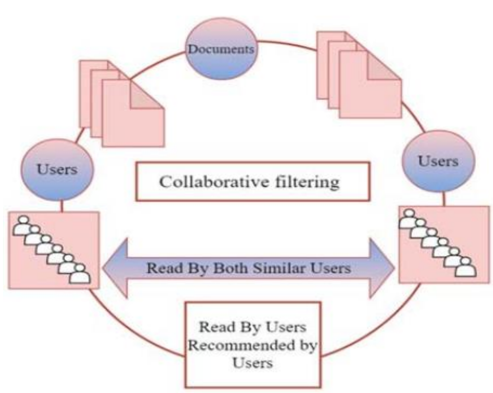

Figure 3. Collaborative Filtering

Example: A streaming platform suggests sci-fi movies to a user based on previously liked genres, directors, or themes [18].

Challenges: Content-based filtering often suffers from limited serendipity and over-specialization, making recommendations too narrow. It depends heavily on accurate content representation and can face scalability challenges when handling large and diverse catalogs.

To overcome these obstacles, we need to use sophisticated feature extraction methods, combine content-based and collaborative filtering into one hybrid approach, and constantly tweak our algorithms to make content-based filtering systems more random and diverse in their recommendations.

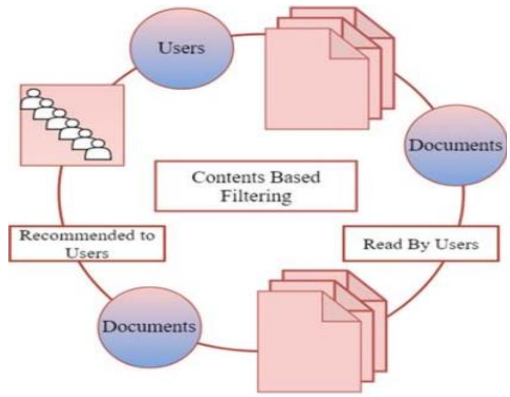

Figure 4. Content-Based filtering

Knowledge-Based Recommendation Systems:

These systems leverage domain-specific knowledge, such as expert-defined rules or explicit user constraints. They excel when product characteristics are complex or infrequently purchased (e.g., laptops, electronics).

Example: If a user specifies, they need a laptop with a dedicated GPU and 1TB of storage, the system filters offerings that match this requirement.

Challenges: These systems require continuous knowledge acquisition and rule maintenance, which limits scalability. They may struggle with personalization for unpredictable user needs and face cold-start issues when explicit data is insufficient.

To overcome these obstacles, we need to build efficient systems for acquiring and maintaining knowledge, create scalable algorithms, combine knowledge-based techniques with other recommendation methods to create hybrid approaches, and tweak the system often to accommodate new product information and user tastes.

Hybrid Recommendation Systems:

Hybrid models combine two or more techniques to exploit their strengths and mitigate their weaknesses. The most common combinations are collaborative + content-based or collaborative + knowledge-based. These systems aim to improve both accuracy and diversity, handle cold-start scenarios more effectively, and integrate multiple signals such as ratings, tags, and click-through data. For example, one approach might use association rule mining for initial pattern discovery before applying collaborative filtering for personalized ranking. Another design could include a knowledge-based layer that filters options according to user constraints before collaborative filtering refines the recommendations.

Challenges: Hybrid systems also introduce additional challenges. Their complexity means that incorporating more components leads to higher training and maintenance costs. Explainability becomes an issue, as these models are harder to debug and interpret. Additionally, tuning the system specifically determining the optimal weighting of each component is a non-trivial task that requires careful experimentation and validation.

Despite the complexity, hybrid models are now standard in large-scale platforms like Netflix and Amazon [18][11][17].

Shared Challenges Across Techniques: Despite their differences, all of these recommendation techniques share common challenges, including managing cold-start scenarios, handling data sparsity, ensuring scalability for large-scale applications, protecting user privacy, and adapting to dynamic user preferences. Recognizing these shared obstacles is essential for selecting or designing effective recommendation systems in real-world e-commerce settings.

RQ4: Which machine learning approaches, based on a quality-ranked review, appear most promising for real-world implementation in scalable, personalized recommendation systems?

To identify the most promising techniques for real-world deployment, we cross-analyzed each method's theoretical value (as described in RQ3) with its quality assessment scores (RQ2) and practical challenges (RQ3). Our evaluation considers both academic rigor (model validation, clear conclusions, strong publication venue) and applied viability (scalability, personalization, adaptability).

Hybrid Recommendation Systems – The Leading Strategy:

Hybrid models emerged as the most robust and versatile approach, scoring consistently high across both quality and functionality criteria. Studies such as authors (2023), (2022), (2022) not only proposed advanced hybrid architectures but also validated them using real-world e-commerce datasets, achieving improvements in both accuracy and diversity.

These systems effectively mitigate the individual weaknesses of standalone techniques by combining collaborative filtering's strength in personalization with content-based filtering's resilience to cold-start problems. They also leverage association rule mining to generate interpretable patterns and integrate knowledge-based constraints to ensure controlled, rule-driven filtering. As a result, hybrid systems offer high adaptability across different industries and user types, flexibility in handling diverse data inputs such as behavioral signals and metadata, and are easily extendable to include contextual, temporal, and user feedback components.

Collaborative Filtering – Popular but Fragile:

Collaborative filtering remains the most implemented technique in e-commerce platforms (e.g., Amazon, Netflix), but its effectiveness is heavily data-dependent. Papers like authors (2020)) demonstrated strong results when user-item interactions were dense. However, in sparse datasets or new-user scenarios, CF underperformed without hybrid augmentation.

While model-based CF (e.g., using matrix factorization or autoencoders) offers some relief, these techniques often come with higher computational costs and privacy risks.

Content-Based Filtering – Strong for Cold Start, Weak for Variety:

Content-based models, such as those in Pawłowski (2021) and Lili & Jianmin (2022), are promising for smaller or startup-scale platforms where item metadata is rich but user history is limited. They performed well in cold-start tests and were easier to explain to users (e.g., "Recommended because you liked X").

However, due to over-specialization, they may reduce long-term engagement unless combined with diversity-aware algorithms or collaborative inputs.

Knowledge-Based Systems – Niche Power:

Knowledge-based systems showed high value in domain-specific contexts, particularly in electronics, real estate, and medical applications. Studies like by authors (2004) and (2002) scored well in quality assessment but are generally harder to scale due to manual rule creation and limited adaptability.

They are best suited for high-stakes recommendations requiring user-defined constraints (e.g., "Only show laptops with SSD and over 8GB RAM").

Association Rule Mining – Interpretability Without Personalization:

While ARM was frequently included in papers like by authors (2020) mentioned its effectiveness is narrow in scope. ARM excels in basket-based suggestions ("customers who bought X also bought Y") but lacks personalization and adaptability. High-quality ARM studies were often hybridized with other techniques, and few relied on ARM alone for recommendations.

Table 8. Comparative Analysis of Techniques

Technique	Cold-Start Handling	Scalability	Personalization	Real-Time Adaptation	Best Fit Context
Hybrid Models	Strong (via CBF/KBS)	High (modular and scalable)	Strong (behavior + content)	Moderate to High	Enterprise e-commerce, multi-domain platforms
Collaborative Filtering	Weak (needs historical data)	Moderate to High (model-based)	Strong (deep personalization)	Low (requires retraining)	Large user-item matrix systems
Content-Based Filtering	Strong (uses item features)	Moderate (depends on metadata)	Moderate (based on similarity)	Moderate (needs frequent updates)	Niche apps, media streaming, and startups
Knowledge-Based Systems	Moderate (explicit rules)	Low (manual rule maintenance)	Moderate (domain-aligned)	Low (static logic)	Regulated/high-trust domains (e.g., electronics, healthcare)
Association Rule Mining	Weak (needs frequent patterns)	Low (rule explosion risk)	Weak (population-level patterns)	Low (not adaptive)	Product bundling, upselling, and offline analytics

Table 8 highlights the relative strengths and weaknesses of each approach across key deployment dimensions, helping identify their best-fit contexts in real-world e-commerce.

To better illustrate the relative strengths and trade-offs of these techniques across key evaluation dimensions, Figure X presents a visual summary comparing their performance in cold-start handling, scalability, personalization, adaptability, and explainability.

Comparison with Existing Reviews:

Existing literature reviews on recommendation systems have often focused on categorizing approaches such as collaborative filtering, content-based filtering, and hybrid models, typically summarizing their algorithmic structures and theoretical foundations without providing a structured, quality-ranked comparison. Classic surveys tend to describe these methods individually but rarely assess their practical challenges or deployment suitability in real-world e-commerce contexts [8].

More recent reviews have emphasized emerging methods like deep learning and neural collaborative filtering, often showcasing improved accuracy on benchmark datasets. However, they frequently overlook practical concerns such as scalability constraints, cold-start mitigation strategies, explainability requirements, and data sparsity challenges. Many also lack a systematic framework for evaluating the methodological rigor and publication quality of the literature they cite [15].

While Table 8 in Section 4.4 provided a detailed numeric comparison of these techniques across key operational dimensions, prior reviews rarely offer such structured or multi-dimensional evaluations. In contrast, this study adopts a quality-ranked, multi-dimensional approach that assesses relative strengths and trade-offs across critical factors such as cold-start handling, scalability, personalization, adaptability, and explainability. To highlight our study's added value and enable a clear, side-by-side contrast with existing literature, Figure 5 below presents a visual summary of these comparative strengths using a heatmap.

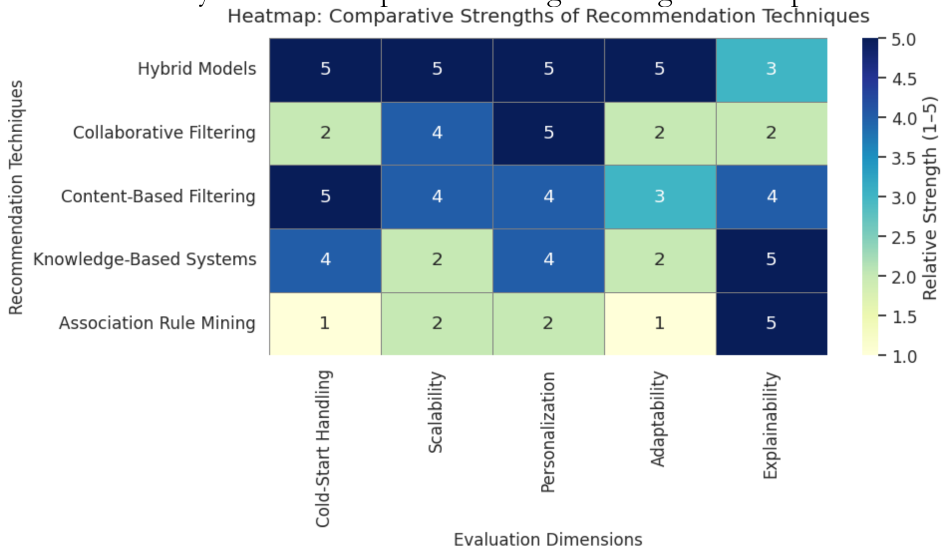

Figure 5. Heatmap summarizing recommendation techniques across key deployment dimensions.

As shown in Figure 5, hybrid recommendation systems demonstrate consistently strong performance across all dimensions, making them the most promising choice for scalable and personalized deployment. While collaborative filtering excels in personalization, it is limited in cold-start handling and adaptability. Content-based and knowledge-based methods offer better explainability and cold-start resilience but face challenges in scalability and diversity. Association rule mining, while highly interpretable, is less effective for dynamic, personalized recommendations.

By combining quality scoring, publication trends analysis, and a critical evaluation of deployment suitability, this review offers practical guidance that goes beyond prior surveys. It enables researchers and practitioners to make more informed, context-aware decisions when selecting or designing recommendation systems for real-world applications.

Conclusion:

This study provided a structured and quality-ranked review of machine learning-based recommendation systems for e-commerce, examining 44 peer-reviewed publications across five primary techniques: content-based filtering, collaborative filtering, association rule mining, knowledge-based systems, and hybrid models. By applying a systematic assessment framework, the review offered a nuanced understanding of each approach's strengths, limitations, and deployment suitability in diverse e-commerce contexts.

The analysis emphasizes that while each technique offers specific advantages, hybrid recommendation systems stand out as the most flexible and scalable option, capable of integrating user behavior, item metadata, and rule-based constraints to deliver personalized and context-aware recommendations. This highlights the importance of designing systems that can balance accuracy, scalability, interpretability, and user trust in real-world applications [18][19].

Future Work:

Looking ahead, future research should focus on optimizing hybrid architectures, addressing their design and tuning complexities, and developing evaluation frameworks that consider not only accuracy but also diversity, novelty, and user satisfaction. Advancements in deep learning, reinforcement learning, and explainable AI also hold strong potential for enhancing recommendation precision and transparency [11][17]. By tackling these challenges, the field can move towards building intelligent, adaptive, and user-centered recommendation systems that better serve the evolving needs of digital commerce.

References:

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013, doi: https://doi.org/10.1016/j.knosys.2013.03.012.

[2] T. M. K. Xiaoyuan Su, "A Survey of Collaborative Filtering Techniques," *Adv. Artif. Intell.*, 2009, doi: https://doi.org/10.1155/2009/421425.

[3] D. C. Gema Bello-Orgaz, Jason J. Jung, "Social big data: Recent achievements and new challenges," *Inf. Fusion*, vol. 28, pp. 45–59, 2016, doi: https://doi.org/10.1016/j.inffus.2015.08.005.

[4] J. A. K. Michael D. Ekstrand, John T. Riedl, "Collaborative Filtering Recommender Systems," *Found. Trends® Human–Computer Interact.*, vol. 4, no. 2, pp. 81–173, 2011, doi: http://dx.doi.org/10.1561/1100000009.

[5] S. Sedhain, A. K. Menony, S. Sannery, and L. Xie, "AutoRec: Autoencoders meet collaborative filtering," *WWW 2015 Companion - Proc. 24th Int. Conf. World Wide Web*, pp. 111–112, May 2015, doi: 10.1145/2740908.2742726;SUBPAGE:STRING:ABSTRACT;CSUBTYPE:STRING:CONFERENCE.

[6] F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems Handbook: Third Edition," *Recomm. Syst. Handb. Third Ed.*, pp. 1–1060, Jan. 2022, doi: 10.1007/978-1-0716-2197-4.

[7] J. J. K. Qi An, Saifur Rahman, Jingwen Zhou, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," *Sensors*, vol. 23, no. 9, p. 4178, 2023, doi: https://doi.org/10.3390/s23094178.

[8] B. Thomas and A. K. John, "Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1085, no. 1, p. 012011, Feb. 2021, doi: 10.1088/1757-899X/1085/1/012011.

[9] K. A. J. Mingming Guo, Nian Yan, Xiquan Cui, San He Wu, Unaiza Ahsan, Rebecca West, "Deep Learning-based Online Alternative Product Recommendations at Scale," *Assoc. Comput. Linguist.*, pp. 19–23, 2020, doi: 10.18653/v1/2020.ecnlp-1.3.

[10] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, Jan. 2019, doi: 10.1145/3285029;REQUESTEDJOURNAL:JOURNAL:CSUR;TAXONOMY:TAXONOMY:ACM-PUBTYPE;PAGEGROUP:STRING:PUBLICATION.

[11] L. Tian, B. Yang, X. Yin, and Y. Su, "A Survey of Personalized Recommendation Based on Machine Learning Algorithms," *ACM Int. Conf. Proceeding Ser.*, pp. 602–610, Nov. 2020, doi: 10.1145/3443467.3444711;SUBPAGE:STRING:ABSTRACT;CSUBTYPE:STRING:CONFERENCE.

[12] M.-F. Leung *et al.*, "A portfolio recommendation system based on machine learning and big data analytics," *Data Sci. Financ. Econ. 2023 2152*, vol. 3, no. 2, pp. 152–165, 2023, doi: 10.3934/DSFE.2023009.

[13] M. E. G. Manal Loukili, Fayçal Messaoudi, "Machine learning based recommender system for e-commerce," *Inst. Adv. Eng. Sci.*, vol. 12, no. 4, 2023, doi: http://doi.org/10.11591/ijai.v12.i4.pp1803-1811.

[14] M. Tahir, R. N. Enam, and S. M. N. Mustafa, "E-commerce platform based on Machine Learning Recommendation System," *IMTIC 2021 - 6th Int. Multi-Topic ICT Conf. AI Meets IoT Towar. Next Gener. Digit. Transform.*, 2021, doi: 10.1109/IMTIC53841.2021.9719822.

[15] A. A. Ssvr Kumar Addagarla, "Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce," *Symmetry (Basel).*, vol. 12, no. 11, p. 1783, 2020, doi: https://doi.org/10.3390/sym12111783.

[16] J. Anitha and M. Kalaiarasu, "Optimized machine learning based collaborative filtering (OMLCF) recommendation system in e-commerce," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 6, pp. 6387–6398, Jun. 2021, doi: 10.1007/S12652-020-02234-1/METRICS.

[17] R. Sharma, S. Rani, and S. Tanwar, "Machine learning algorithms for building recommender systems," *2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, pp. 785–790, May 2019, doi: 10.1109/ICCS45141.2019.9065538.

[18] D. J. Dudhia, S. R. Dave, and S. Yagnik, "Self attentive product recommender - A hybrid approach with machine learning and neural network," *2020 Int. Conf. Emerg. Technol. INCET 2020*, Jun. 2020, doi: 10.1109/INCET49848.2020.9154034.

[19] A. Biswas, K. S. Vineeth, A. Jain, and Mohana, "Development of Product Recommendation Engine by Collaborative Filtering and Association Rule Mining Using Machine Learning Algorithms," *Proc. 4th Int. Conf. Inven. Syst. Control. ICISC 2020*, pp. 272–277, Jan. 2020, doi: 10.1109/ICISC47916.2020.9171210.