

Literature Review: Analysis of the Use of Data Mining in Various Recommender Systems

Refia Ningsih¹, Hetty Rohayani²

^{1,2} Sains Dan Teknologi, Universitas Muhammadiyah Jambi, Jambi, Indonesia

¹refiahrefiah@gmail.com, ²rohayanihettytnkk@gmail.com

Article Info

Article history:

Received February 19, 2025

Revised May 15, 2025

Accepted June 30, 2025

Keywords:

Data Mining

Recommender Systems

Systematic Literature Riview

Recommendation algorithms

Collaborative Filtering

Metalearning

ABSTRACT

The rapid development of information technology has led to a significant increase in data generated across various domains, including tourism, e-commerce, education, and entertainment. Data mining has become an essential technique for extracting valuable knowledge from large datasets, with recommender systems representing one of its most important applications for providing personalized recommendations based on user preferences and behavior. This study aims to provide a comprehensive review of the application of data mining techniques in recommender systems. A Systematic Literature Review (SLR) following the PRISMA guidelines was conducted by analyzing 20 peer-reviewed articles published between 2020 and 2025, which were collected from Google Scholar. The selected studies were analyzed using a qualitative descriptive approach based on algorithms, application domains, datasets, evaluation metrics, strengths, limitations, and research trends. The findings indicate that Collaborative Filtering and K-Nearest Neighbor (KNN) are the most frequently applied techniques due to their ability to generate personalized recommendations. Several studies also demonstrate that hybrid approaches improve recommendation quality and help overcome challenges such as the cold-start problem and data sparsity. Furthermore, recent research highlights the growing adoption of machine learning, deep learning, explainable recommender systems, and privacy-aware approaches. This study provides a comprehensive synthesis of current research trends and offers valuable insights for researchers and practitioners in selecting appropriate data mining techniques for developing accurate, efficient, and reliable recommender systems.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Refia Ningsih

Sains Dan Teknologi, Universitas Muhammadiyah Jambi, Jambi, Indonesia

Email: refiahrefiah@gmail.com

1. INTRODUCTION

The rapid development of information technology and the Internet has generated an enormous volume of data from various user activities, including online information searches, electronic transactions, social media interactions, and digital entertainment services. This rapid growth has increased the need for data mining techniques to transform large volumes of raw data into meaningful information that supports decision-making processes. Data mining is the process of discovering hidden patterns, relationships, and valuable knowledge from large datasets through analytical techniques [1],[2]. One of the most significant applications of data mining is the development of recommender systems, which assist users in identifying products, services, or information that best match their needs and preferences based on available data. By analyzing user behavior and historical data, recommender systems help reduce information overload, improve user satisfaction, and support more effective decision-making.

The application of data mining in recommender systems has expanded across various domains, including tourism, music, e-commerce, books, movies, and other digital services. Numerous studies have demonstrated that data mining algorithms can improve recommendation quality according to the characteristics of different application domains. Panjaitan, Pangaribuan, and Cahyadi [3] applied the K-Nearest Neighbor (KNN) algorithm in a tourism recommender system, while Hartono and Sutopo [4] implemented the Upper Confidence Bound (UCB) algorithm in a music recommender system to generate recommendations that better match user preferences. Furthermore, Aisha and Kusumawati [5] combined Collaborative Filtering and K-Nearest Neighbor (KNN) in an e-commerce recommender system, whereas Faurina and Sitanggang [6] integrated Content-Based Filtering and Collaborative Filtering in a tourism recommender system for Bali. These studies demonstrate that different data mining techniques can effectively improve recommendation quality. However, they primarily focus on implementing specific algorithms within particular application domains, making it difficult to generalize the effectiveness of each method across different types of recommender systems.

In addition to implementation studies, several researchers have conducted literature reviews on recommender systems. Putri, Nurakhmadyavi, and Wahyudi [7] reviewed recommender systems in the domains of books and movies, while Ifada and Pratama examined the application of numerical data clustering in Collaborative Filtering-based recommender systems. Meanwhile, Suhailah and Hartatik [8] developed a tourism recommender system using the Triangle Multiplying Jaccard method, whereas Dewi [9] optimized the Pearson Correlation algorithm using the Firefly Algorithm to improve recommendation performance. These studies indicate that each method has distinct characteristics, strengths, and limitations depending on the application domain, dataset characteristics, and evaluation methods employed. However, most existing studies focus on specific algorithms or application domains and do not provide a comprehensive overview of the application of various data mining techniques across different types of recommender systems.

Based on the previous studies, it can be observed that most existing research focuses on the implementation of specific data mining algorithms within particular application domains, providing limited comprehensive synthesis of the effectiveness of different methods across various recommender systems. Furthermore, previous literature reviews generally concentrate on specific domains or recommendation approaches and do not integrate comparisons of algorithms, dataset characteristics, evaluation metrics, implementation challenges, and recent technological trends. Therefore, this study aims to address this research gap by presenting a comprehensive literature review of data mining techniques applied across various recommender system domains. The novelty of this study lies in its comprehensive synthesis of the most widely used algorithms, dataset characteristics, evaluation metrics, strengths and limitations of each method, and key implementation challenges, including the cold-start problem, data sparsity, scalability, privacy, and explainability, based on studies published between 2020 and 2025 [10]. Consequently, this study is expected to provide valuable references for both researchers and practitioners in selecting appropriate data mining techniques according to dataset characteristics and application requirements.

The selection of the 2020–2025 publication period was motivated by the rapid advancement of recommender system research during these years, including the development of conventional data mining techniques and their integration with machine learning, deep learning, big data, and artificial intelligence. The literature published during this period provides valuable insights into algorithm development, dataset characteristics, evaluation metrics, implementation challenges, and emerging research trends in recommender systems [11]. Based on this background, this study addresses the following research questions (RQs): (RQ1) Which data mining techniques are most frequently applied in recommender systems? (RQ2) What are the characteristics of each technique based on application domains and datasets? (RQ3) What are the strengths and limitations of each data mining technique in recommender systems? and (RQ4) What are the recent development trends in data mining-based recommender systems based on studies published between 2020 and 2025?

Based on these research questions, this study aims to conduct a comprehensive literature review of the application of data mining techniques in various recommender systems. Specifically, the study seeks to identify the most frequently used algorithms, analyze their characteristics across different application domains, compare their strengths and limitations, and examine current research trends as well as future development opportunities. The findings of this study are expected to contribute to the academic literature by providing a comprehensive reference for future research while also offering practical implications for recommender system developers in selecting appropriate data mining techniques based on dataset characteristics, user requirements, and application objectives, thereby improving recommendation quality and supporting more effective decision-making.

2. METHOD

2.1 Research Stages

This study employed the Systematic Literature Review (SLR) method to systematically identify, evaluate, and synthesize previous studies on the application of data mining techniques in recommender systems. The SLR approach was selected because it provides a structured, transparent, and reproducible procedure for reviewing scientific literature, thereby improving the reliability and credibility of the research findings. The objective of this review is to identify the most widely used data mining methods, compare their performance, analyze their strengths and limitations, and identify future research opportunities in recommender systems. This method involves collecting, selecting, and reviewing various previous studies relevant to the research topic. The goal is to identify the data mining methods used, their areas of application, and their effectiveness in generating recommendations. The stages of the Systematic Literature Review (SLR) conducted in this study are illustrated in Figure 1.

The review process consisted of several systematic stages, namely problem formulation, literature searching, literature selection, data extraction, quality assessment, data synthesis, and result interpretation. Each stage was conducted systematically to ensure that the selected studies were relevant, credible, and consistent with the objectives of this review.

1. Problem Formulation and Objectives

This stage involved identifying the research problem and defining the objectives of the study. The primary objective was to review the application of data mining techniques in recommender systems, identify the most frequently used algorithms, analyze their strengths and limitations, and explore future research opportunities in this field.

2. Literature Search

The literature search was conducted using Google Scholar as the primary database because it provides broad access to peer-reviewed publications in computer science, information systems, and recommender systems. Only articles published between 2020 and 2025 were considered to ensure the relevance and recency of the reviewed studies.

3. Literature Selection

The retrieved articles were screened using predefined inclusion and exclusion criteria. Only peer-reviewed journal articles published between 2020 and 2025, written in English or Indonesian, and directly related to the application of data mining in recommender systems were included in the review.

4. Data Extraction

Relevant information was extracted systematically from each selected study. The extracted data included the authors, publication year, application domain, datasets, data mining algorithms, evaluation metrics, research findings, strengths, and limitations. This information served as the basis for the comparative analysis.

5. Quality Assessment

Each selected study was evaluated to ensure its quality and relevance. The assessment considered several criteria, including research relevance, methodological clarity, completeness of experimental results, publication quality, and contribution to recommender system research. Only studies that met these criteria were included in the final analysis.

6. Data Analysis and Synthesis

The extracted data were analyzed using a qualitative descriptive approach. The selected studies were compared based on the algorithms used, application domains, datasets, evaluation methods, strengths, limitations, and research trends. The findings were then synthesized to identify similarities, differences, and future research opportunities.

7. Result Interpretation

The final stage involved interpreting the synthesized findings to answer the research questions. The interpretation focused on identifying the most widely used data mining techniques, discussing current trends, highlighting existing challenges, and providing recommendations for future research and practical implementation of recommender systems.

2.2 Literature Sources and Selection

The literature search was conducted using Google Scholar as the primary scientific database. These databases were selected because they provide broad coverage of peer-reviewed publications in computer

science, information systems, and recommender system research. To ensure the relevance and quality of the reviewed studies, only articles published between 2020 and 2025 were considered.

2.2.1 Search Strategy

The literature search was performed using combinations of keywords related to data mining and recommender systems. Boolean operators (AND and OR) were applied to improve the accuracy and relevance of the search results.

Table 1. Search Strings Used in the Literature Search

No.	Search String
1	"Data Mining" AND "Recommender System"
2	"Recommendation System" AND "Collaborative Filtering"
3	"Recommendation System" AND "Content-Based Filtering"
4	"Recommendation System" AND "Machine Learning"
5	"Recommendation System" AND "Data Mining"

To ensure that the literature reviewed was relevant and up to date, this study focused on publications from 2020 to 2025. The literature search was conducted using Google Scholar with the predefined search strings presented in Table 1. The selected period was intended to capture recent developments in data mining techniques and recommendation systems. The retrieved studies were subsequently screened based on their relevance to the research topic and their contribution to the analysis of data mining techniques in recommendation systems.

2.2.2 Inclusion Criteria

Table 2. Inclusion Criteria

Criteria	Description
Publication Year	Articles published between 2020 and 2025.
Language	Articles written in English or Indonesian.
Publication Type	Peer-reviewed journal articles.
Full Text Availability	Only full-text articles were included.
Research Topic	Studies discussing the application of data mining in recommender systems.

The inclusion criteria were applied consistently during the screening process to ensure that the selected studies were relevant to the research objectives. Studies that did not directly discuss the application of data mining techniques in recommendation systems, did not meet the specified publication period, or were not available as complete articles were excluded from the analysis. The screening process was conducted by reviewing the title, abstract, keywords, and full text of the retrieved studies before determining the final studies included in the review.

2.2.3 Exclusion Criteria

Table 3. Exclusion Criteria

Criteria	Description
Duplicate Articles	Duplicate articles retrieved during the literature search were excluded.
Incomplete Articles	Articles without full-text access were excluded.
Irrelevant Topic	Articles not related to recommender systems were excluded.
Irrelevant Method	Articles that did not discuss data mining techniques were excluded.

Through the application of the inclusion and exclusion criteria, the selected studies were carefully evaluated based on their relevance, methodological quality, and contribution to the research objectives. Finally, 20 articles that fulfilled all predefined criteria were included in this systematic literature review.

2.3 Data Analysis and Synthesis

The selected studies were analyzed using a qualitative descriptive approach. The analysis focused on identifying the types of data mining algorithms, application domains, datasets, evaluation metrics, advantages,

limitations, and research findings reported in each study. The extracted information was then synthesized to identify similarities, differences, research trends, and future opportunities in recommender systems.

In addition, a review of the evaluation methods used in each study was conducted. Some evaluation methods commonly found in recommendation systems include measuring accuracy, precision, recall, F1-score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and other evaluation metrics used to assess recommendation quality. Although not all studies used the same metrics, all evaluation results were analyzed proportionally to obtain an overview of the effectiveness of each method.

This study also compared various algorithms and approaches used in recommendation systems, such as Collaborative Filtering, Content-Based Filtering, K-Nearest Neighbor (KNN), Pearson Correlation, Upper Confidence Bound, and other methods. This comparison was conducted to identify the advantages, disadvantages, and trends in the use of each method in various application areas. The results of the analysis were then used as a basis for discussing the development of data mining in recommendation systems, the effectiveness of the methods used, and opportunities for their development in further research.

The analysis was conducted by comparing the characteristics, applications, advantages, limitations, and evaluation approaches of the data mining techniques identified in the selected studies. The studies were also examined according to their application domains to identify differences in the use of recommendation techniques across various contexts. This approach was used to provide a descriptive synthesis of the findings and to identify patterns and trends in the application of data mining techniques in recommendation systems.

2.4 Quality Assessment

To ensure the quality and credibility of the selected studies, a quality assessment was conducted before the data synthesis process. Each article was evaluated using predefined assessment criteria, including relevance to the research topic, clarity of the research methodology, completeness of experimental results, publication quality, and contribution to the field of recommender systems. Only studies that met these quality criteria were included in the final review.

Table 4. Quality Assessment Criteria

Assessment Criteria	Description
Relevance	The study is directly related to the application of data mining in recommender systems.
Research Methodology	The research methodology is clearly described and systematically implemented.
Experimental Results	The study presents evaluation results or experimental findings.
Publication Quality	The article is published in a peer-reviewed journal or reputable conference proceedings.
Research Contribution	The study provides meaningful contributions to recommender system research.

The quality assessment was performed to minimize the inclusion of studies with insufficient methodological rigor and to ensure that the synthesized findings were based on reliable scientific evidence. This process also increased the transparency and consistency of the systematic literature review.

The quality assessment was conducted to evaluate the methodological quality and relevance of each selected study before its findings were included in the synthesis. Each study was assessed based on its relevance to the research objectives, clarity of the research methodology, adequacy of the reported results, quality of the publication, and contribution to the development of recommendation systems. The assessment was used to ensure that the synthesis was based on studies with sufficient methodological and scientific quality. Studies that did not adequately meet the assessment criteria were considered less suitable for supporting the main findings of the review.

The quality assessment was applied to the 20 studies included in the final synthesis. The assessment considered five aspects: relevance to the research topic, clarity of the research methodology, clarity of the reported results, quality of the publication, and contribution to the development or application of recommendation systems. Each study was reviewed according to these criteria before its findings were synthesized. The assessment indicated that the selected studies were sufficiently relevant to the research objectives and provided information that could be used to analyze the application of data mining techniques in recommendation systems.

2.5 Validity and Reliability

The validity of this systematic literature review was maintained by applying predefined inclusion and exclusion criteria, conducting a systematic search through Google Scholar, and using a systematic literature selection process. These procedures ensured that the selected studies were relevant to the research objectives

and minimized the risk of including irrelevant literature. Reliability was ensured by conducting a consistent literature selection process, systematically extracting data from each selected article, and applying the same analysis procedure throughout the review. This approach helped improve the consistency, transparency, and credibility of the research findings.

To maintain the validity and reliability of the review findings, the analysis was based on the predefined inclusion criteria, quality assessment criteria, and systematic literature selection process. The findings were synthesized by comparing information reported across the selected studies, including the data mining techniques, application domains, evaluation approaches, and reported results. This procedure was intended to reduce the influence of subjective interpretation and ensure that the conclusions were supported by the evidence identified in the reviewed literature.

2.6 Prisma Flow Diagram

The article selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework. The framework was used to systematically describe the stages of identification, screening, eligibility assessment, and final inclusion of studies. In the identification stage, 145 records were retrieved from Google Scholar, and 38 duplicate records were removed, leaving 107 records for screening. During the screening stage, 63 records were excluded because they were not relevant to recommendation systems or data mining. A total of 44 reports were then sought for full-text retrieval, of which 4 reports could not be retrieved. The remaining 40 reports were assessed for eligibility, and 20 reports were excluded because they did not meet the predefined criteria. Finally, 20 studies were included in the systematic literature review and qualitative synthesis.

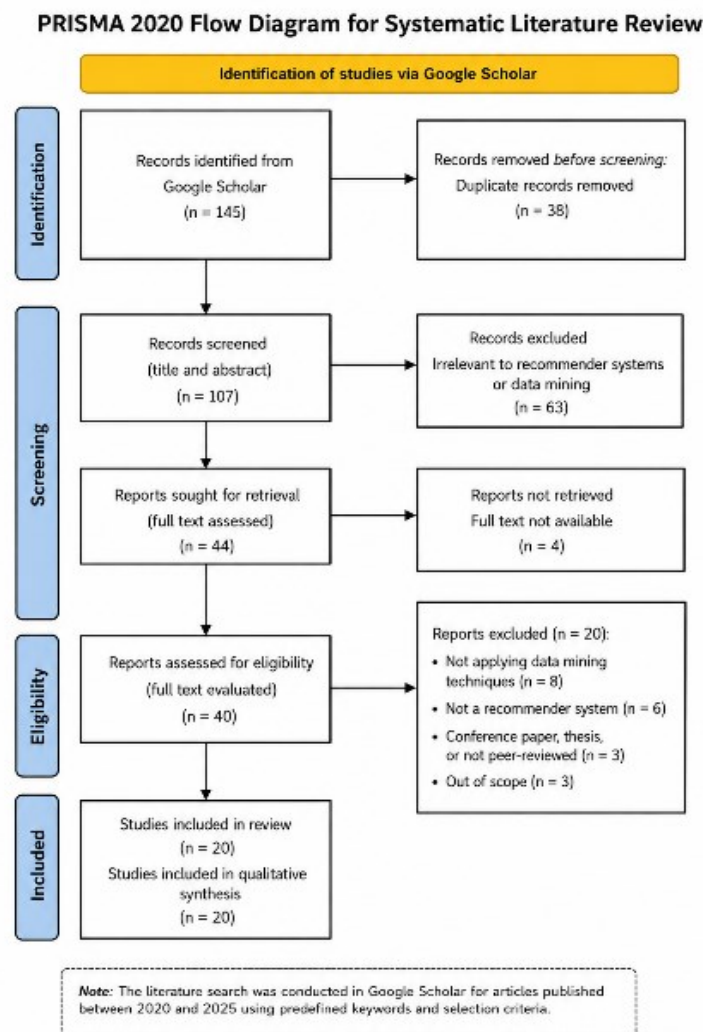


Figure 1. Prisma Flow Diagram

The PRISMA flow diagram illustrates the complete literature selection process, from the initial identification of records to the final inclusion of studies. The selection process consisted of duplicate removal, title and abstract screening, full-text retrieval, eligibility assessment, and final inclusion. Based on these stages, 20 studies met the predefined inclusion criteria and were included in the qualitative synthesis.

3. RESULTS AND DISCUSSION

This section presents the results and discussion of the systematic literature review on the use of data mining techniques in various recommendation systems. The study applies a Systematic Literature Review (SLR) approach to identify, evaluate, and synthesize relevant previous research. The reviewed studies cover various data mining techniques used in recommendation systems, including K-Nearest Neighbor (KNN), Collaborative Filtering, Content-Based Filtering, Pearson Correlation, Upper Confidence Bound, and Triangle Multiplying Jaccard. The discussion focuses on the application domains, methods, evaluation approaches, advantages, limitations, and emerging trends identified in the reviewed studies. The results are presented in tabular form and further discussed to provide an overview of the development and application of data mining techniques in recommendation systems.

3.1 Reference Results Table

Data were extracted from the 20 selected studies, including the datasets, methods, application domains, and main findings reported in each study. The extracted information is presented in the following table to provide a structured overview of the characteristics and results of the reviewed studies.

Table 5. Review Results

No	Researcher	Journal Title	Data	Method	Results
1	C. H. P. Panjaitan et al.(2022)	Analysis of the K-Nearest Neighbor Method Using RapidMiner for a Labuan Bajo Tourism Recommendation System.	Labuan Bajo tourism destination data containing destination characteristics and user preferences	K-Nearest Neighbor (KNN) with RapidMiner	The research results show that the KNN method can group data based on similarity levels and generate tourist destination recommendations that match user needs. The use of RapidMiner also supports data processing and model testing, enabling the recommendation system to assist tourists in selecting suitable destinations.
2	A. D. Hartono & J. Sutopo (2024)	Implementation of the Upper Confidence Bound Algorithm for a Music Recommendation System.	Song data, playback history, and user interactions	Upper Confidence Bound (UCB)	The research shows that the UCB algorithm is able to balance exploration and exploitation in providing music recommendations. The system can recommend songs that align with user preferences while simultaneously discovering new, potentially popular alternatives, resulting in more diverse and relevant recommendations.
3	D. Aisha & R. Kusumawati (2022)	Implementation of Collaborative Filtering and K-Nearest Neighbor Algorithms in an E-Commerce Recommendation System.	Product, user, and rating data	Collaborative Filtering and KNN	The research results show that the combination of Collaborative Filtering and KNN can improve the quality of product recommendations on e-commerce platforms. The system utilizes user rating patterns and similarities between users to generate more personalized recommendations that meet consumer needs.

No	Researcher	Journal Title	Data	Method	Results
4	R. Faurina & E. Sitanggang (2023)	Implementation of Content-Based Filtering and Collaborative Filtering in a Bali Tourism Recommendation System.	Tourist attraction data, attributes, and user preferences	Content-Based Filtering and Collaborative Filtering	The research shows that using the two methods simultaneously can produce more accurate recommendations than using one method alone. The system can consider both the characteristics of tourist attractions and the user's historical preferences, resulting in more relevant and personalized recommendations.
5	N. H. Putri et al.(2022)	Literature Review: Recommendation Systems for Books and Movies	Previous studies on book and movie recommendation systems	Literature Review	The study results indicate that Collaborative Filtering, Content-Based Filtering, and Hybrid Recommendation are the most widely used approaches in book and movie recommendation systems. The study also found that hybrid methods can overcome some limitations of individual methods, thereby improving the quality of recommendations provided to users.
6	E. Suhailah & Hartatik(2023)	Development of a Yogyakarta Tourism Recommendation System Using Triangle Multiplying Jaccard	Yogyakarta destination and user preference data	Triangle Multiplying Jaccard	The research results show that the Triangle Multiplying Jaccard method is effective in calculating similarity between users based on their preferences. The method can generate relevant tourist destination recommendations and help users find tourist locations that match their interests
7	M. M. Dewi(2022)	Optimization of Pearson Correlation for Recommendation Systems Using the Firefly Algorithm	User rating data	Pearson Correlation and Firefly Algorithm	The research shows that applying the Firefly Algorithm can improve the performance of the Pearson Correlation method in recommendation systems. The optimization results in higher accuracy and better recommendation quality than using Pearson Correlation without optimization, making the system more effective in providing recommendations to users.
8	K. R. Sari et al.(2022)	Development of a Movie Recommendation System Using Item-Based	Movie rating data	Item-Based Collaborative Filtering and Apache Mahout	The research results show that the Item-Based Collaborative Filtering method can generate movie recommendations based on the similarity between

No	Researcher	Journal Title	Data	Method	Results
		Collaborative Filtering in Apache Mahout			items. The use of Apache Mahout facilitates the processing of large amounts of data, allowing the system to provide faster and more relevant recommendations based on user preferences.
9	V. M. Hersianty et al.(2025)	Application of TF-IDF and Cosine Similarity in a Job Vacancy Recommendation System	Job vacancy descriptions, qualifications, and skills	TF-IDF and Cosine Similarity	The research shows that the combination of TF-IDF and Cosine Similarity is effective in measuring the level of similarity between user profiles and job vacancy information. The system is able to provide job recommendations that are more suited to users' abilities and interests, thus helping to make the job search process more efficient.
10	H. Rohayani & M. C. Umam (2023)	Prediction of Study Program Selection Based on Student Scores Using Backpropagation	Student academic score data	ANN with Backpropagation	The research results show that the Backpropagation algorithm can learn patterns from student grade data and generate predictions of suitable study programs. The developed model can be used as a decision-making tool to provide study program recommendations based on students' academic abilities.
11	H. Rohayani et al.(2022)	Data Mining Implementation to Predict Student Graduation Using Naive Bayes	Student academic data	Naive Bayes	The research results show that the Naive Bayes method can be used to predict student graduation rates based on available academic data. The developed model can help universities identify students who have the potential to graduate on time and those who require more attention, thus supporting academic decision-making.
12	M. Maulidah et al.(2022)	Decision Tree Classification Algorithm for Book Recommendations Based on Categories	Book category and collection data	Decision Tree	The research results show that the Decision Tree algorithm can classify books based on their characteristics and categories, generating recommendations that match user interests. In addition to being easy to implement, the method produces understandable decision rules that support the recommendation process.

No	Researcher	Journal Title	Data	Method	Results
13	A. Widyan & A. F. Rozi(2022)	Product Recommendation Analysis Using the ECLAT Algorithm Based on Sales History	Sales transactions and purchase history	ECLAT	The research results show that the ECLAT algorithm can identify relationships between products that are frequently purchased together. This information can be used to generate product recommendations, support marketing strategies, and increase sales opportunities through more targeted product offerings.
14	P. Brazdil et al.(2022)	Metalearning: Applications to Automated Machine Learning and Data Mining	Various datasets for metalearning research	Metalearning and AutoML	This book explains how a metalearning approach can be used to select the most appropriate algorithm based on the characteristics of the analyzed data. The discussion shows that metalearning can improve the efficiency of the data mining model selection process and support the development of more adaptive intelligent systems in various application fields.
15	J. Han & M. Kamber(2022)	Data Mining: Concepts and Techniques	Various datasets for explaining data mining concepts	Various data mining methods	This book is a primary reference in the field of data mining because it comprehensively explains the concepts, processes, and techniques of data mining. The discussion shows that data mining methods can be applied to various problems, including recommendation systems, to discover hidden patterns and generate useful information for decision making.
16	Vernanda Sam Saputra et al.(2023)	Design and Development of a Book Recommendation System Based on Item-Based Collaborative Filtering Using KNN	Book rating data	Item-Based Collaborative Filtering and KNN	The study successfully developed a book recommendation system that provides recommendations based on user preferences. Cross-validation testing produced an RMSE value of 0.8191 and an MAE value of 0.6235, indicating that the model achieved good accuracy in predicting book recommendations.
17	Bagas Ilham Rabbani et al.(2022)	Recommendation System Using Mobile iOS-Based Item-Based	User rating data	Item-Based Collaborative Filtering and KNN	The study results indicate that the Item-Based Collaborative Filtering method can be successfully implemented in an iOS mobile application. The system can calculate

No	Researcher	Journal Title	Data	Method	Results
		Collaborative Filtering			similarities between items and generate relevant recommendations based on user rating patterns.
18	Venny Meida Hersianty et al.(2025)	Application of TF-IDF and Cosine Similarity in a Job Vacancy Recommendation System	Job seeker profiles and vacancy descriptions	TF-IDF and Cosine Similarity	The system can match user profiles with available job openings by calculating the similarity between documents. The evaluation results show an accuracy rate of approximately 80%, indicating that the method is effective in providing relevant job recommendations.
19	Donny Susianto & Ziad Rusdi(2023)	Recommendation System for Electronic Product Sales Using Collaborative Filtering	Transaction, rating, and purchase history data	Collaborative Filtering	The study shows that Collaborative Filtering can identify customer purchasing patterns and generate electronic product recommendations that match user needs. The system helps customers find relevant products based on the preferences and purchasing behavior of other users.
20	Kirana Nuryunita & Yani Nurhadryani (2023)	Development of a Recommendation Module in OpenCart Using Item-Based Collaborative Filtering	Book ratings and OpenCart activity data	Item-Based Collaborative Filtering, Adjusted Cosine Similarity, and Weighted Sum	The study results indicate that the Item-Based Collaborative Filtering method can effectively generate product recommendations. Adjusted Cosine Similarity can calculate the similarity between items, while Weighted Sum is used to generate rating predictions that help improve the quality of recommendations.

Based on the application domains, the reviewed studies can be grouped into several categories, including tourism, e-commerce and product recommendation, entertainment, education, employment, and general recommendation systems. Tourism-related studies focus on recommending destinations based on user preferences and destination characteristics, while e-commerce and product recommendation studies mainly utilize user ratings, transaction history, and purchasing patterns. Entertainment-related studies include music, movie, and book recommendation systems. In the education domain, data mining techniques are applied to predict study program selection and student graduation. Meanwhile, employment-related research applies text similarity techniques to match user profiles with job vacancy information. This categorization indicates that the selection of data mining techniques is influenced by the characteristics of the data and the requirements of each application domain.

3.2 Analysis

Based on the 20 studies summarized in Table 5, the use of data mining in recommendation systems has developed significantly and plays an important role in supporting users' decision-making processes across various fields. Recommendation systems have been applied in tourism, music, e-commerce, books, films, marketplaces, education, and job vacancies, demonstrating the flexibility of data mining in processing different types of data and user needs. The review found that Collaborative Filtering is one of the most widely used approaches because it can utilize user interaction patterns and preferences to generate relevant recommendations.

In addition, various methods, including K-Nearest Neighbor (KNN), Content-Based Filtering, Pearson Correlation, Triangle Multiplying Jaccard, Upper Confidence Bound (UCB), TF-IDF, and Cosine Similarity, have been applied to improve recommendation quality. The analysis shows that Collaborative Filtering and KNN can identify similarities between users and items, thereby supporting more personalized recommendations. In tourism recommendation systems, these methods can help users find destinations that match their interests, while in e-commerce, marketplace, book, and movie recommendation systems, they can generate recommendations based on user ratings and activity history.

However, several studies have identified limitations in certain methods, including the cold-start problem, limited rating data, and increasing computational complexity as data volume grows. To address these issues, some studies have applied hybrid approaches by combining methods such as Collaborative Filtering with KNN, Content-Based Filtering with Collaborative Filtering, and Pearson Correlation optimized using the Firefly Algorithm. The results indicate improvements in recommendation quality in terms of accuracy, relevance, and the ability to capture user preferences. In addition, clustering techniques and similarity measures such as Cosine Similarity can improve the efficiency of processing large amounts of data.

Overall, the analysis indicates that data mining is an effective technology for developing recommendation systems because it can generate accurate, personalized, and user-oriented recommendations. The selection of an appropriate method is strongly influenced by the characteristics of the data and the objectives of the system. Therefore, future development of recommendation systems should consider hybrid methods and optimization techniques to improve system performance and overcome the limitations of individual methods.

The suitability of data mining techniques in recommendation systems is influenced by the characteristics of the available data and the requirements of the application domain. In tourism recommendation systems, K-Nearest Neighbor (KNN), Content-Based Filtering, Collaborative Filtering, and Triangle Multiplying Jaccard can utilize destination characteristics and user preferences to identify relevant recommendations. In e-commerce and product recommendation systems, Collaborative Filtering and ECLAT can utilize user ratings, transaction histories, and purchasing patterns to identify products that may be relevant to users. In entertainment domains such as music, movies, and books, recommendation techniques can utilize user ratings, playback history, item characteristics, and user preferences. Meanwhile, TF-IDF and Cosine Similarity are suitable for job vacancy recommendation because they can measure the similarity between textual information in user profiles and job descriptions. These findings indicate that the selection of a recommendation technique should consider the type, availability, and characteristics of the data as well as the requirements of the application domain.

The reviewed studies also show that each data mining technique has specific advantages and limitations. Collaborative Filtering is effective for utilizing user-item interaction patterns and generating personalized recommendations, but its performance can be affected by data sparsity and the cold-start problem involving new users or items. K-Nearest Neighbor (KNN) can identify similar users or items based on their characteristics, but its computational cost may increase when processing large or high-dimensional datasets. Content-Based Filtering can provide recommendations based on item characteristics and user preferences, but its performance depends on the availability and quality of item attributes. TF-IDF and Cosine Similarity are useful for measuring similarity between textual information, but their effectiveness depends on the quality and representation of the text data. Association-based methods such as ECLAT can identify relationships among frequently purchased items, although their usefulness depends on the availability of sufficient transaction data. Therefore, no single data mining technique can be considered universally superior, and the choice of method should be adapted to the characteristics of the data and the objectives of the recommendation system.

3.3 Comparison of Evaluation Metrics

The reviewed studies employed various evaluation metrics to measure the performance and effectiveness of recommendation systems. The selection of evaluation metrics depended on the objectives of each study, the characteristics of the datasets, and the recommendation methods applied. The most frequently used metrics included Accuracy, Precision, Recall, F1-score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). These metrics provide different perspectives for assessing recommendation quality, ranging from prediction accuracy to the relevance of recommended items.

Table 6. summarizes the evaluation metrics commonly used in the reviewed studies.

Evaluation Metric	Purpose	Commonly Used In
Accuracy	Measures the overall correctness of recommendations	Classification-based recommendation systems

Evaluation Metric	Purpose	Commonly Used In
Precision	Measures the proportion of relevant recommended items	Collaborative Filtering
Recall	Measures the proportion of relevant items successfully retrieved	Recommendation systems
F1-score	Balances Precision and Recall	Machine Learning-based recommendation
MAE	Measures average prediction error	Rating prediction
RMSE	Measures prediction error with higher penalty for large errors	Rating prediction

The analysis indicates that no single evaluation metric is sufficient to comprehensively measure the performance of a recommendation system. Accuracy is useful for measuring prediction correctness; however, it does not fully reflect the relevance of recommendations. Precision and Recall are more appropriate for evaluating recommendation quality because they measure the relevance of recommended items. Meanwhile, MAE and RMSE are commonly used to evaluate prediction errors in rating-based recommendation systems. Therefore, combining multiple evaluation metrics can provide a more comprehensive assessment of system performance.

The evaluation results reported in the reviewed studies show that different metrics were used according to the objectives of each recommendation system. For example, the book recommendation study using Item-Based Collaborative Filtering and KNN reported an RMSE of 0.8191 and an MAE of 0.6235, while the job vacancy recommendation study using TF-IDF and Cosine Similarity reported an accuracy of approximately 80%. These results provide evidence of the performance of the respective models within their individual experimental settings. However, the reported values cannot be directly compared to determine which method is universally superior because the studies used different datasets, domains, evaluation metrics, and experimental conditions. Therefore, the performance comparison in this review is presented descriptively rather than as a statistical ranking of the algorithms.

3.4 Challenges in Recommendation Systems

Recommendation systems have experienced significant advancements through the application of various data mining techniques. However, the reviewed studies reveal several challenges that continue to affect the effectiveness and performance of recommendation systems. One of the most common challenges is the cold-start problem, which occurs when new users or new items have insufficient interaction history, making it difficult for the system to generate accurate recommendations. This issue is particularly evident in Collaborative Filtering methods that rely heavily on historical user-item interactions.

Another major challenge is data sparsity, where user rating matrices contain a large number of missing values. Sparse data reduce the ability of similarity-based algorithms such as Collaborative Filtering and K-Nearest Neighbor to identify meaningful relationships between users or items. Several studies attempted to overcome this limitation by combining recommendation techniques with hybrid approaches or similarity optimization methods.

As recommendation systems are increasingly implemented in large-scale applications, scalability has become an important issue. Processing large datasets may require greater computational resources and longer processing times, which can affect system efficiency. Furthermore, data privacy has become increasingly important because recommendation systems often rely on users' personal information and behavioral data. Maintaining recommendation accuracy while protecting user privacy remains an important challenge for future research. Another emerging issue is explainability, as users increasingly expect recommendation systems to provide understandable reasons for recommended items. Improving the transparency of recommendation systems can increase user trust and acceptance. Therefore, future studies should focus not only on improving recommendation accuracy but also on addressing these practical challenges to develop more reliable and user-centered recommendation systems.

Another emerging issue is explainability, where users increasingly expect recommendation systems to provide understandable explanations for why certain items are recommended. Improving the transparency of recommendation systems is expected to increase user trust and acceptance. Therefore, future studies should focus not only on improving recommendation accuracy but also on addressing these practical challenges to develop more reliable and user-centered recommendation systems.

In addition to these challenges, the implementation of recommendation systems in real-world environments also needs to consider algorithmic bias, information security, and computational requirements. Algorithmic bias may occur when historical user data do not adequately represent different user groups, which can affect the fairness of the recommendations. Information security is also important because recommendation

systems may process user profiles, preferences, transaction histories, and other behavioral information. Furthermore, the increasing volume and complexity of data can require greater computational resources to maintain system performance and scalability. Therefore, the development of recommendation systems should consider not only recommendation accuracy but also fairness, security, scalability, and computational efficiency.

3.5 Future Research Opportunities

Future research on recommendation systems is expected to focus on integrating advanced artificial intelligence techniques with traditional data mining methods. Based on the reviewed studies, hybrid recommendation systems show considerable potential for overcoming the limitations of single-method approaches, particularly in addressing cold-start problems and improving recommendation accuracy. Therefore, developing hybrid models that combine Collaborative Filtering, Content-Based Filtering, clustering techniques, and machine learning algorithms is expected to become an important research direction.

Recent developments in Deep Learning have also created opportunities to improve recommendation quality by learning complex patterns in user behavior and item characteristics. In addition, AutoML (Automated Machine Learning) offers the potential to automate model selection and parameter optimization, reducing the need for manual intervention while improving model performance. Another promising area is Meta-Learning, which enables recommendation systems to adapt more effectively to different datasets and application domains.

Furthermore, future studies should pay greater attention to Context-Aware Recommendation Systems, which consider contextual information such as location, time, user activity, and environmental conditions to generate more personalized recommendations. Researchers are also encouraged to investigate approaches that improve explainability, privacy protection, and computational efficiency to ensure that recommendation systems remain reliable, transparent, and scalable in real-world applications.

Overall, future developments should not only emphasize improving recommendation accuracy but should also focus on developing intelligent, efficient, explainable, and user-centered recommendation systems that can address the growing challenges of modern digital environments. The implementation of these emerging approaches also presents several challenges that require further investigation. Deep Learning-based recommendation systems generally require sufficient training data and computational resources, while AutoML requires reliable data and appropriate evaluation procedures to select and optimize models effectively. Meta-Learning provides opportunities to adapt recommendation models across different datasets and application domains, but its effectiveness depends on the availability of suitable learning tasks. Context-Aware Recommendation Systems can improve personalization by incorporating contextual information, but collecting and processing such information may raise additional privacy and computational concerns. Therefore, future research should not only investigate improvements in recommendation performance but also examine practical implementation, computational requirements, privacy, scalability, and the applicability of these emerging techniques in different recommendation domains.

4. CONCLUSION

Based on the results of this systematic literature review on the use of data mining in various recommendation systems, it can be concluded that data mining plays an important role in improving the quality and effectiveness of recommendation systems across various fields, including tourism, music, e-commerce, books, movies, marketplaces, education, and job vacancies. Various methods, such as Collaborative Filtering, Content-Based Filtering, K-Nearest Neighbor (KNN), Pearson Correlation, Upper Confidence Bound (UCB), Triangle Multiplying Jaccard, TF-IDF, and Cosine Similarity, have demonstrated their ability to support the identification of user preferences and the generation of relevant recommendations. Collaborative Filtering and KNN are among the frequently applied methods in the reviewed studies because they can utilize similarities between users or items to support personalized recommendations. Furthermore, several studies indicate that hybrid methods or combinations of algorithms can improve recommendation quality compared with individual methods. However, recommendation systems still face limitations, including cold-start problems, data sparsity, and increasing computational complexity when processing large datasets. The findings of this review provide practical implications for researchers and practitioners. Researchers can use the results as a reference for selecting appropriate data mining techniques and identifying research trends in recommendation systems. Meanwhile, practitioners and system developers can use these findings to support the design and implementation of recommendation systems that are accurate, personalized, and efficient according to user needs and application domains. Future research is recommended to explore advanced hybrid recommendation approaches by integrating traditional data mining techniques with emerging artificial intelligence technologies, such as Deep Learning, AutoML, Meta-Learning, and Context-Aware Recommendation Systems. Future

studies should also evaluate recommendation systems using larger and more diverse datasets, investigate explainable recommendation models, strengthen privacy protection mechanisms, and assess system performance in real-time environments to improve scalability and practical applicability.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2022.
- [2] C. Huang, T. Yu, K. Xie, S. Zhang, L. Yao, and J. McAuley, "Foundation Models for Recommender Systems: A Survey and New Perspectives," *arXiv*, 2024.
- [3] C. H. P. Panjaitan, L. J. Pangaribuan, and C. I. Cahyadi, "Analisis Metode K-Nearest Neighbor Menggunakan RapidMiner untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo," 2022.
- [4] A. D. Hartono and J. Sutopo, "Implementasi Algoritma Upper Confidence Bound Untuk Sistem Rekomendasi Musik," 2024.
- [5] D. Aisha and R. Kusumawati, "Implementasi Metode Algoritma Collaborative Filtering dan K-Nearest Neighbor pada Sistem Rekomendasi E-Commerce," 2022.
- [6] R. Faurina and E. Sitanggang, "Implementasi Metode Content-Based Filtering dan Collaborative Filtering pada Sistem Rekomendasi Wisata di Bali," 2023.
- [7] N. H. Putri, "Literature Review: Sistem Rekomendasi untuk Buku dan Film," 2022.
- [8] E. Suhailah and Hartatik, "Pembuatan Sistem Rekomendasi Pariwisata Yogyakarta Menggunakan Triangle Multiplying Jaccard," 2023.
- [9] M. M. Dewi, "Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly," 2022.
- [10] A. Klimashevskaya, D. Jannach, M. Elahi, and C. Trattner, "A Survey on Popularity Bias in Recommender Systems," *User Model. User-adapt. Interact.*, vol. 34, pp. 1777–1834, 2024.
- [11] T. M. Al-Hasan, A. N. Sayed, F. Bensaali, Y. Himeur, and G. Dimitrakopoulos, "From Traditional Recommender Systems to GPT-Based Chatbots: A Survey of Recent Developments and Future Directions," *Big Data Cogn. Comput.*, vol. 8, no. 4, 2024.
- [12] C. H. P. Panjaitan, L. J. Pangaribuan, and C. I. Cahyadi, "Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo," 2022.
- [13] K. R. Sari, W. Suharso, and Y. Azhar, "Pembuatan Sistem Rekomendasi Film dengan Menggunakan Metode Item-Based Collaborative Filtering pada Apache Mahout," 2020.
- [14] V. M. Hersianty, "Penerapan Algoritma TF-IDF dan Cosine Similarity dalam Sistem Rekomendasi Lowongan Pekerjaan," 2025.
- [15] H. Rohayani and M. C. Umam, "Prediksi Penentuan Program Studi Berdasarkan Nilai Siswa dengan Algoritma Backpropagation," 2023.
- [16] H. Rohayani, "Implementasi Data Mining untuk Memprediksi Kelulusan Mahasiswa Fakultas Sains dan Teknologi Menggunakan Metode Naive Bayes," 2022.
- [17] M. Maulidah, "Algoritma Klasifikasi Decision Tree untuk Rekomendasi Buku Berdasarkan Kategori Buku," 2020.
- [18] A. Widyan and A. F. Rozi, "Analisis Rekomendasi Produk Menggunakan Algoritma ECLAT Berdasarkan Riwayat Data Penjualan PT XYZ," 2021.
- [19] P. Brazdil, *Metalearning: Applications to Automated Machine Learning and Data Mining*. 2022.
- [20] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. 2023.
- [21] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender Systems Survey: Evolution, Challenges and Recent Trends," *IEEE Access*, vol. 9, pp. 138634–138658, 2021.
- [22] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep Learning Based Recommender System: A Survey and New Perspectives," *ACM Computing Surveys*, vol. 52, no. 1, 2020.
- [23] M. Elahi, D. Jannach, M. Ge, and F. Ricci, "Explainable Food Recommendation: A Survey and New Perspectives," *Frontiers in Artificial Intelligence*, vol. 4, 2021.
- [24] A. Da'u and N. Salim, "Recommendation System Based on Deep Learning Methods: A Systematic Review," *Artificial Intelligence Review*, vol. 55, pp. 1077–1128, 2022.
- [25] M. Dacrema, P. Cremonesi, and D. Jannach, "Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches," *ACM Transactions on Information Systems*, vol. 39, no. 4, 2021.