

A Systematic Literature Review of Data Mining Algorithms for Drug Sales Analysis

Adhittya Noviarini^{1(✉)}, Hetty Rohayani²

^{1,2} Science and Technology, Muhammadiyah University of Jambi, Jambi, Indonesia
rininovia860@gmail.com

Article Info

Article history:

Received February 14, 2025

Revised May 30, 2025

Accepted June 30, 2025

Keywords:

Data Mining

Pharmaceutical Sales

Drug Sales Analysis

Classification

Prediction

Clustering

Association Rule Mining

Systematic Literature Review

ABSTRACT

The increasing availability of transaction data in the pharmaceutical sector provides opportunities for the application of data mining techniques to support drug sales analysis and decision-making. However, previous studies have applied different data mining algorithms for different analytical objectives, making it difficult to compare their applications, advantages, limitations, and reported performance. This study aims to systematically review and synthesize previous studies on the application of data mining algorithms in pharmaceutical sales analysis. A Systematic Literature Review (SLR) method was employed through problem identification, formulation of research questions, definition of inclusion and exclusion criteria, literature searching, article screening and selection, data extraction, and comparative analysis and synthesis. The literature search was conducted using Google Scholar, Garuda, Scopus, and FEEL, with studies published between 2019 and 2026 considered for the review. A total of 85 records were identified, 18 duplicate records were removed, and 20 studies met the eligibility criteria for the final review. The reviewed studies applied several data mining algorithms, including Naive Bayes, Decision Tree C4.5, K-Nearest Neighbor (KNN), Neural Network, K-Means, Apriori, and FP-Growth, which were categorized into classification, prediction, clustering, and association rule mining objectives. The findings indicate that Naive Bayes, C4.5, and KNN are mainly used for classification and prediction, K-Means for clustering, and Apriori and FP-Growth for identifying purchasing associations. Reported performance varied across studies, and direct comparison was limited by differences in datasets, research objectives, and evaluation methods. The review also identifies research gaps, particularly the limited use of standardized datasets and evaluation metrics and the limited comparative evaluation of multiple algorithms using the same pharmaceutical sales dataset. Future research should consider comparative, hybrid, ensemble, explainable, and automated approaches to improve the applicability of data mining techniques in pharmaceutical sales analysis and decision support.

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



Corresponding Author:

Adhittya Noviarini

Science and Technology, Muhammadiyah University of Jambi, Jambi, Indonesia

Email: rininovia860@gmail.com

1. INTRODUCTION

The development of information technology has led to a rapid increase in the amount of data generated by digital systems, including sales information systems in the healthcare and pharmaceutical sectors. Drug sales transaction data collected from pharmacies, hospitals, and other healthcare facilities contain valuable information related to drug types, sales quantities, prices, purchasing frequency, and transaction patterns. However, the increasing volume and complexity of pharmaceutical sales data can create challenges in identifying useful information for inventory management, demand estimation, product segmentation, and purchasing pattern analysis. Data mining provides techniques and algorithms that can be used to discover patterns, relationships, and valuable information from large datasets to support more effective decision-making [1], [2], [3], [4], [5].

In data mining, various algorithms can be applied according to the characteristics of the data and the objectives of the analysis. These algorithms include classification methods for grouping data into specific categories, prediction methods for estimating future outcomes based on historical data, clustering methods for grouping data with similar characteristics, and association rule mining methods for identifying relationships between items or transactions [6]. Various data mining algorithms have been developed and applied in different fields, including the health and pharmaceutical sectors [7], [8], [9], [10], [11], [12]. In drug sales analysis, these algorithms can be used to identify best-selling and non-best-selling drug categories, predict sales or stock requirements, group drugs based on similar sales characteristics, identify customer purchasing patterns, and support decision-making in drug inventory management. Some frequently used algorithms include Naive Bayes, Decision Tree C4.5, K-Nearest Neighbor (KNN), Neural Network, K-Means, Apriori, and FP-Growth, which have different characteristics, analytical objectives, and performance measures.

Previous studies have demonstrated the application of various data mining algorithms in different analytical contexts. Hetty Rohayani et al [13] applied the Naive Bayes algorithm to predict student graduation and obtained an accuracy rate of 93.88%. Another study conducted by Hetty Rohayani and Muhammad Choirul Umam [14] applied the Neural Network Backpropagation method and reported an accuracy rate of 77.42%. In the context of sales analysis, M. Afdhal et al [15] applied the C4.5 algorithm to predict product purchasing patterns and produced a decision tree model that can support marketing strategies. A. Azis et al [16] utilized the K-Nearest Neighbor (KNN) algorithm to predict sales of best-selling drugs and medical devices, while S. P. Awalina et al [17] applied the C4.5 algorithm to classify drug sales in pharmacies based on their sales levels.

Beyond classification and prediction, other data mining approaches have also been applied to pharmaceutical sales analysis. M. Fajar et al [18], A. Yani et al. [19], and H. I. Muttaqin [20] used the K-Means Clustering algorithm to group drug sales and distribution data based on certain characteristics. Meanwhile, P. A. Lestari et al. [21], Salam Irianto Nadeak and Yusmar Ali [22], and Siti Muntari [23] applied the Apriori and FP-Growth algorithms to identify association patterns in drug purchases that can support promotional strategies and stock management. These studies demonstrate that pharmaceutical sales analysis involves different data mining objectives, including classification, prediction, clustering, and association rule mining. Therefore, a broader review of data mining algorithms is needed to understand how different approaches are applied to different pharmaceutical sales problems.

Although numerous studies have investigated the application of data mining algorithms in pharmaceutical sales analysis, most studies focus on specific algorithms, individual analytical objectives, or particular case studies. As a result, the existing literature provides limited comparative evidence regarding the suitability, strengths, and limitations of different data mining approaches for pharmaceutical sales problems. In addition, differences in datasets, analytical objectives, and evaluation measures make it difficult to directly compare the findings across studies. This situation creates a research gap in understanding how classification, prediction, clustering, and association rule mining approaches have been applied to pharmaceutical sales data and which approaches are most suitable for different analytical objectives. Therefore, a systematic literature review is needed to synthesize existing evidence, compare the applications and evaluation results of different data mining algorithms, identify current research gaps and trends, and provide a basis for selecting appropriate approaches for pharmaceutical sales analysis.

Therefore, this study aims to systematically review and synthesize previous studies on the application of data mining algorithms in pharmaceutical sales analysis. This review focuses on identifying commonly used algorithms, classifying them according to their analytical objectives, examining the characteristics of the datasets and evaluation methods, comparing their reported performance, advantages, and limitations, and identifying research gaps and future research opportunities. The results of this study are expected to provide a structured reference for researchers and practitioners in understanding the suitability of different data mining approaches for specific pharmaceutical sales analysis objectives.

2. METHOD

This study uses the Systematic Literature Review (SLR) method to identify, evaluate, and synthesize previous research related to the application of various data mining algorithms in pharmaceutical sales analysis [24]. The SLR method was selected because it provides a systematic approach for collecting, evaluating, and synthesizing relevant studies and can provide an overview of research trends, methods, and findings in a specific research area [24][25]. The review process consists of several stages, including formulating research questions, defining inclusion and exclusion criteria, developing a literature search strategy, screening and selecting relevant articles, assessing the quality of the selected studies, extracting relevant data, and synthesizing the findings.

2.1 Research Questions

The research questions (RQs) were formulated to guide the systematic literature review process and to ensure that the review addresses the research objectives. The research questions are as follows:

RQ1: What data mining algorithms are most frequently used in pharmaceutical sales analysis?

RQ2: What types of datasets and analytical objectives are commonly addressed in pharmaceutical sales analysis studies?

RQ3: What are the reported performance, advantages, limitations, and applicability of different data mining algorithms in pharmaceutical sales analysis?

RQ4: What research gaps and future research opportunities can be identified from existing studies on data mining for pharmaceutical sales analysis?

2.2 Inclusion Criteria

The inclusion criteria were determined to ensure that the selected studies were relevant to the research objectives and provided sufficient information for the review. The inclusion criteria used in this study are as follows:

1. Articles published between 2019 and 2026.
2. Articles discussing drug sales, pharmaceutical sales, pharmacy sales, or healthcare product sales analysis.
3. Articles applying data mining techniques, including classification, prediction, clustering, or association rule mining.
4. Articles published in national or international journals or conference proceedings.
5. Full-text articles accessible to researchers.

2.3 Exclusion Criteria

The exclusion criteria were established to remove studies that did not meet the research objectives or did not provide sufficient information for the analysis. The exclusion criteria used in this study are as follows:

1. Duplicate articles identified from multiple databases.
2. Articles that are not related to drug, pharmaceutical, pharmacy, or healthcare product sales analysis.
3. Articles for which the full text is not accessible.
4. Articles that do not provide sufficient methodological information to support data extraction and analysis.

2.4 Research Stages

The research stages were carried out systematically, starting from problem identification to drawing conclusions based on the literature review. The research flow used can be seen in Figure 1.



Figure 1. Research Stages Flowchart

The first stage is problem identification, which aims to identify the application of data mining algorithms in drug and pharmaceutical sales analysis. Next, the research objectives are formulated, followed by the determination of literature search keywords. The subsequent stage is the search for relevant scientific articles, followed by article selection based on the predetermined inclusion and exclusion criteria. The selected articles are then analyzed and synthesized to identify the characteristics, applications, performance, advantages, and limitations of the data mining algorithms used in drug and pharmaceutical sales analysis. The identified algorithms are also categorized according to their analytical objectives, including classification, prediction, clustering, and association rule mining. The final stage is the preparation of the research results and conclusions based on the findings of the literature review.

2.5 Article Selection Process Using Prisma 2020

This study adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure a transparent, systematic, and reproducible article selection process. The PRISMA framework was used to identify, screen, assess eligibility, and select relevant studies concerning the application of data mining algorithms in drug and pharmaceutical sales analysis.

The literature search was conducted using several academic databases, including Google Scholar, Garuda, Scopus, and FEEL. The search used keywords related to the research topic, including “drug sales analysis,” “drug sales prediction,” “pharmaceutical sales,” “data mining,” “classification algorithm,” “inventory prediction,” “Naive Bayes,” “Decision Tree,” “K-Nearest Neighbor,” “K-Means,” “Apriori,” and “FP-Growth.” These keywords were selected to identify studies applying data mining algorithms to drug, pharmaceutical, pharmacy, and healthcare product sales analysis.

Initially, 85 records were identified through database searching. After removing 18 duplicate records, 67 studies remained for title and abstract screening. Studies were excluded at this stage if they were not related to drug, pharmaceutical, pharmacy, or healthcare product sales analysis, did not apply relevant data mining techniques, or did not meet the predetermined inclusion criteria.

Subsequently, the remaining studies underwent full-text eligibility assessment. Based on the inclusion and exclusion criteria, 20 articles met all eligibility requirements and were selected for the final systematic literature review and comparative analysis.

To improve the consistency of study selection, the selected articles were assessed based on their relevance to the research objectives, clarity of the research methodology, description of the dataset, data mining algorithm used, and availability of research results. Studies that did not provide sufficient methodological or result information were excluded from the review.

Figure 2 presents the PRISMA 2020 flow diagram illustrating the complete study selection process, from identification through final inclusion. The adoption of the PRISMA 2020 framework enhances the transparency and reproducibility of the review process and provides a structured basis for selecting relevant studies.

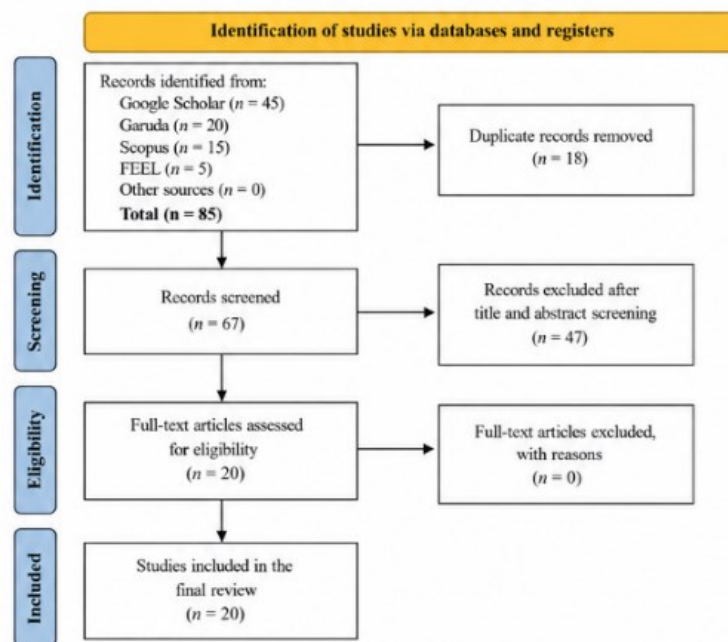


Figure 2. Prisma 2020 Flow Diagram

2.6 Data Sources and Selection

The literature search was conducted using several scientific databases, including Google Scholar, Garuda (Garba Rujukan Digital), Scopus, and FEEL. These databases were selected to identify relevant national and international publications related to drug and pharmaceutical sales analysis and the application of data mining techniques.

The search keywords were formulated based on the research objectives and included “drug sales analysis,” “drug sales prediction,” “pharmaceutical sales,” “data mining,” “classification algorithm,” “Naive Bayes,” “Decision Tree,” “K-Nearest Neighbor,” “K-Means,” “Apriori,” and “FP-Growth.” Articles published between 2019 and 2026 were considered for the review.

The retrieved articles were screened based on their relevance to the research topic and the predetermined inclusion and exclusion criteria. The screening process involved reviewing the article titles and abstracts, followed by full-text assessment. The characteristics and relevant information of the eligible studies were then extracted for further analysis.

2.7 Data Analysis

Data analysis was conducted using a descriptive-comparative approach on the selected studies. Information extracted from each article included the author and publication year, research objective, dataset characteristics, data mining algorithm, analytical category, evaluation method, reported performance, advantages, limitations, and main research findings. The extracted information was organized into comparison tables to facilitate the identification of similarities and differences among the reviewed studies.

The analysis focused on comparing the data mining algorithms according to their analytical objectives, including classification, prediction, clustering, and association rule mining. Where reported by the original studies, algorithm performance was compared using evaluation metrics such as accuracy, precision, recall, F1-score, or other relevant evaluation measures. Performance results were interpreted in the context of the dataset, research objective, and evaluation method used in each study. Direct comparisons between studies were made cautiously because the reviewed studies used different datasets, experimental settings, and evaluation measures.

The analysis also considered the advantages, limitations, and applicability of each algorithm based on the characteristics of the dataset and research objectives. Furthermore, the findings were synthesized to identify commonly used algorithms, research trends, differences in research approaches, factors that may influence algorithm performance, and research gaps. This synthesis was used to identify potential opportunities for future research, including comparative studies using the same pharmaceutical sales dataset, hybrid or ensemble approaches, and the application of data mining techniques to larger and more diverse pharmaceutical sales datasets.

3. RESULTS AND DISCUSSION

3.1 Results of the Selected Studies

This section presents the results of the literature review and the discussion of various data mining algorithms applied to drug and pharmaceutical sales analysis. This study uses a Systematic Literature Review (SLR) approach to identify, examine, evaluate, and synthesize previous studies relevant to the research topic. The reviewed studies cover several data mining approaches, including classification, prediction, clustering, and association rule mining. The algorithms identified in the selected studies include Naive Bayes, C4.5, K-Nearest Neighbor (KNN), K-Means Clustering, Apriori, FP-Growth, and other related algorithms used to analyze drug sales data and support decision-making. The analysis considers the research objectives, datasets, evaluation methods, reported performance, advantages, limitations, and applicability of each approach. To provide a systematic overview of the findings, the results are presented in tables and supported by descriptive and comparative analysis.

3.1.1 References Result Table

Table 1 summarizes the characteristics and findings of the twenty selected studies included in this systematic literature review. The table presents information on the authors and publication year, dataset characteristics, research objectives, data mining algorithms, analytical categories, evaluation methods, key findings, advantages, and limitations reported in the reviewed studies. This information is presented to facilitate a systematic comparison of the different data mining approaches used in drug and pharmaceutical sales analysis.

Table 1. Reference Result Table

Ref	Jurnal Title	Data	Research Objective	Method/ Algorithm	Evaluation Metric	Result	Strength	Limitation
[2]	Data Science and Predictive Analytics: Biomedical and Health Applications using R	Health and pharmaceutical studies from academic and industrial sectors.	To provide theoretical foundations and practical guidance for applying data science, predictive analytics, and machine learning methods to biomedical and health-related problems.	C4.5, KNN, SVM, and K-Means are applied in various biomedical contexts.	Model performance assessment and validation are discussed, but no single evaluation metric is specified as the main metric for the book as a whole.	The book provides practical and theoretical guidance on classification, prediction, and the application of machine learning in medical and pharmaceutical sectors.	Provides both theoretical foundations and practical computational examples covering multiple data science and machine learning approaches in biomedical and health applications.	The book is a broad methodological reference rather than a focused empirical study of pharmaceutical drug sales data.
[13]	Implementation of Data Mining For Predicting the Graduation of Faculty of Science and Technology Students Using the Naive Bayes Method	Data mining techniques with training data (graduates) and testing data (active students).	To predict the graduation status of Faculty of Science and Technology students using the Naive Bayes classification method.	Naive Bayes classification algorithm.	Accuracy and precision.	Accuracy level of 93.88%, with precision of 90% for the "Pass" class and 96.55% for the "Fail" class.	The Naive Bayes method provides a high prediction accuracy and can classify students based on their predicted graduation status.	The study focuses on student graduation prediction rather than pharmaceutical or drug sales data, so its findings cannot be directly generalized to pharmaceutical sales analysis.
[14]	Prediction of Study Program Determination Based on Student Grades with Backpropagation Algorithm	Data mining techniques with training data (graduates) and testing data (active students).	To predict the appropriate study program for students based on their academic grades using the Backpropagation algorithm.	Neural Network Backpropagation, using RapidMiner ; 7 input variables, 2 hidden layers, 50 data samples.	Accuracy: 77.42%.	The model accuracy was 77.42%, showing the potential of applying neural networks to help students choose study programs.	The Backpropagation Neural Network can model relationships between student grades and study program selection.	The study focuses on educational data rather than pharmaceutical sales data, limiting its direct applicability to drug sales analysis.
[15]	Predicting Sales at Hanifah's Store Using the C4.5 Method	Sales transaction dataset obtained from the Hanifah Store sales information system.	To predict consumer purchasing tendencies for products at Hanifah's Store using the C4.5 algorithm.	The C4.5 algorithm is used to build a decision tree to classify products that have a high or low tendency for consumers to buy.	Not reported in the current study summary.	This classification model generates a decision tree that describes purchasing rules. The system improves the effectiveness of marketing and sales strategy development.	C4.5 produces an interpretable decision tree that can describe purchasing rules and support marketing decisions.	The study focuses on a single store, so the findings may have limited generalizability to other sales environments.
[16]	Predicting Sales of Best-Selling Drugs and Medical Devices Using the K-Nearest Neighbor	Sales transaction history from pharmacies and medical device providers.	To predict the sales of best-selling drugs and medical devices based on historical sales transaction data using the K-Nearest	The K-Nearest Neighbor (KNN) algorithm is used for predictive classification of items	Not reported in the current study summary.	The system is able to predict future sales based on historical patterns, helping in distribution and	The method uses historical sales patterns to support the prediction of future product demand.	The study is based on sales data from specific pharmacy and medical-device providers, which may limit

Ref	Jurnal Title	Data	Research Objective	Method/ Algorithm	Evaluation Metric	Result	Strength	Limitation
			Neighbor algorithm.	and best sellers.		promotional strategies.		generalization to other pharmaceutical sales environments.
[17]	Application of the C4.5 Algorithm for Drug Sales Prediction at Waruixeng Pharmacy	Transaction records and stock database from Waruixeng Pharmacy.	To classify drug sales levels at Waruixeng Pharmacy using the C4.5 algorithm.	The C4.5 algorithm is applied for drug classification based on sales level.	Not reported in the current study summary.	The system divides the data into two groups: 172 products that sold and 163 that did not. This result makes it easier to control distribution and predict drug demand trends.	C4.5 provides clear classification results that can help distinguish drugs according to their sales levels.	The study is limited to data from one pharmacy, and the reported result does not provide sufficient quantitative performance measures for comparison with other algorithms.
[18]	Analysis of Drug Sales Patterns at An-Naafi Pharmacy Using the K-Means Clustering Method	Sales, profit, and revenue data for various types of available drugs. Pharmacy in An-Naafi.	To group drug sales data based on similar profit and revenue characteristics using K-Means Clustering.	K-Means Clustering is used to group data based on similar characteristics, in this case profit and revenue.	Not reported in the current study summary.	Two clusters were found: one for low-income drugs and another for high-income drugs. The results of this clustering are used as evaluation material in developing stock management strategies and drug procurement, as well as in determining the promotion of certain drugs.	K-Means can identify groups of drugs with similar sales characteristics, which can support inventory and promotional decisions.	The analysis produces clusters based mainly on sales-related characteristics and does not directly measure predictive performance.
[19]	Data Mining Implementation: Analyzing Sales Data, Using the K-Means Clustering Algorithm	Sales of medical devices, in various distribution areas.	To group medical devices according to their sales characteristics using the K-Means Clustering algorithm.	K-Means Clustering.	Not reported in the current study summary.	Three clusters were obtained: very popular, medium, and less popular health tools. The information was used for distribution strategies.	The clustering results provide clear product groupings that can support distribution planning.	The study focuses on medical-device sales rather than drug sales, and the clustering results are not directly comparable with classification or prediction accuracy.

Ref	Jurnal Title	Data	Research Objective	Method/ Algorithm	Evaluation Metric	Result	Strength	Limitation
[20]	Application of the K-Means Clustering Algorithm for Drug Data Grouping at the Community Health Center, Unter Iwes Sumbawa	Distribution data and demand for GBAT in the work area of the Community Health Center.	To group pharmaceutical distribution and demand data into clusters based on demand characteristics using K-Means Clustering.	K-Means Clustering.	Not reported in the current study summary.	High- and low-demand area clusters are used to optimize pharmaceutical supply and distribution.	The clustering approach can identify areas with different levels of pharmaceutical demand and support supply and distribution planning.	The study focuses on a specific community health center work area, which may limit the generalizability of the results.
[21]	Sales Data Analysis at Ritonga Farma Pharmacy Using Apriori Data Mining	Sales transaction data from Bitonga Farma Pharmacy	To identify frequently purchased drug combinations from pharmacy sales transactions using the Apriori algorithm.	Apriori is used to find frequent itemsets.	Frequent itemsets and association patterns; specific support and confidence values are not reported in the current summary.	Repeated purchasing patterns were identified, such as frequently purchased drug combinations, which is very useful for cross-selling strategies.	Apriori can reveal frequently purchased drug combinations that can support product placement and cross-selling strategies. The published study specifically discusses using frequent-purchase patterns to improve product placement.	The method identifies associations between products but does not directly predict future drug sales or measure classification accuracy.
[22]	Analysis Of Data Mining Assosiations On Drug Sales at Pharmacies With Apriori	Drug sales transaction data at pharmacies.	To identify associations and frequently purchased combinations of drugs from pharmacy sales transactions using the Apriori algorithm.	Apriori	Minimum support of 25% and minimum confidence of 60%.	5 association rules were found with minimum support of 25% and confidence of 60%, it is known that the combination of drugs is often purchased together to support marketing and promotional strategies.	The method successfully identifies purchasing relationships that can be used to support pharmacy marketing and promotional strategies.	The results depend on the selected support and confidence thresholds, and the study focuses on association patterns rather than predictive performance.
[23]	Application of Data Mining Using the FP-Growth Algorithm to Analyze Drug Sales Patterns	Drug sales transaction data for one year.	To identify drug purchasing patterns and frequently purchased drug combinations using the FP-Growth algorithm.	FP-Growth	Minimum confidence of 40%.	36 itemset combinations were found that met the confidence value, thus helping stock and sales management strategies.	FP-Growth can identify frequent item combinations efficiently without requiring the same candidate-generation process as Apriori.	The study focuses on identifying association patterns and does not provide a direct comparison with classification or prediction algorithms.

Ref	Jurnal Title	Data	Research Objective	Method/ Algorithm	Evaluation Metric	Result	Strength	Limitation
[27]	Analysis of Herbal Product Sales Data Using the C4.5 Algorithm read E-Commerce Azka Jaisy Store	E-commerce based herbal product sales from Azka Jaisy Store.	To identify purchasing patterns of herbal products based on price and purchase quantity using the C4.5 algorithm.	C4.5 is implemented, via RapidMiner barga software with volume variables purchase, and product type.	Not reported in the available study summary	The decision tree model is able to classify purchasing tendencies. This information is used to improve promotional personalization.	C4.5 produces an interpretable decision tree that can identify purchasing patterns based on price and purchase quantity.	The study focuses on herbal products in a single e-commerce store, so its findings may not be directly generalizable to pharmaceutical drug sales.
[28]	Algorithm Optimizing Product Sales Prediction with the KNN Algorithm	Combined dataset from e-commerce and conventional retail sectors.	To develop a sales prediction model using the K-Nearest Neighbor algorithm and improve its accuracy through cross-validation.	The KNN algorithm is applied along with cross-validation to obtain an accurate prediction model.	Accuracy: 85%.	The accuracy rate reaches 85%. The classification results are useful for customer segmentation and stock optimization.	The use of cross-validation helps improve the reliability of the prediction model.	The dataset combines e-commerce and conventional retail data, which may contain different purchasing characteristics and can affect model consistency.
[29]	Implementation of FP-Growth in Analyzing Sales, Drugs	Purchase history of the system, information hospital management.	To identify frequent drug purchasing patterns from hospital management system data using the FP-Growth algorithm.	FP-Growth	Not reported in the current study summary.	Generates a frequent itemset model used for clustering and product bundling offers.	FP-Growth can identify frequent itemsets that can be used to support product bundling and sales strategies.	The study focuses on association patterns and does not provide a direct predictive performance comparison with other algorithms.
[30]	Application of Bayes Classifier for Classification of Sales, Drugs at Nugraha Pharmacy	Drug sales transaction data at Nugraha Pharmacy.	To analyze drug sales transaction data and support drug stock management by predicting demand trends using the Naive Bayes Classifier.	Naïve Bayes Classifier	Not reported in the current study summary.	The algorithm is able to identify frequently purchased drug categories and helps predict stock requirements, thereby supporting drug inventory management.	Naive Bayes provides a relatively simple and fast classification approach and can support stock-related decision-making. The article describes its use for identifying frequently purchased drug categories and predicting demand trends.	The study is based on transaction data from a single pharmacy, which may limit the generalizability of the findings.

Ref	Jurnal Title	Data	Research Objective	Method/ Algorithm	Evaluation Metric	Result	Strength	Limitation
[31]	Implementati on of Data Mining with Naive Algorithm Bayes on Drug Sales	Data on sales of vitamins/m edicines at pharmacies.	To classify drugs according to their sales status using the Naive Bayes algorithm to support drug stock decision-making.	Naïve Bayes	Not reported in the current study summary.	Used to classify drugs that sell and don't sell so it helps drug stock decision making.	The method can classify drugs into selling and non-selling categories, providing information that can support stock management.	The analysis focuses on classification of sales status and does not provide a comparison with other classification algorithms.
[32]	Analysis of Drug Sales Clustering Using the K-Means Algorithm for Drug Stock Optimization in OLE Pharmacies	Data on stock, price, sales, and quantity of drugs sold.	To group drugs according to their sales characteristics using K-Means Clustering to support drug stock optimization.	K-Means	Not reported in the current study summary.	Generating two drug sales groups that help optimize inventory and stock control.	K-Means provides clear groups of drugs that can help pharmacy managers organize inventory and control stock.	The clustering approach does not directly provide predictive performance measures such as accuracy, precision, recall, or F1-score.
[33]	Implementati on of Data Mining with the C4.5 Algorithm in Drug Sales	Data on sales of vitamins and medicines.	To classify drug and vitamin sales into selling and non-selling categories using the C4.5 algorithm to support cross-selling and stock management.	Decision Tree C4.5	Accuracy: 99.83%.	Able to predict which drug categories are selling and which are not selling, thus supporting cross-selling strategies and stock management.	The C4.5 decision tree produces highly accurate classification results and provides information about products that are likely to sell or not sell.	The study uses data related to vitamins and medicines from a specific sales environment, so the very high accuracy may not necessarily generalize to other pharmaceutical datasets.
[34]	Implementati on of the K-Nearest Neighbor (KNN) Algorithm for Classification of Drug Sales Data in Pharmacies	Drug sales transaction data, including drug type, sales quantity, price, and sales category.	To classify drug sales levels using the K-Nearest Neighbor algorithm based on historical pharmacy sales data.	K-Nearest Neighbor (KNN) algorithm.	Not reported in the current study summary.	The KNN algorithm is able to classify drug sales levels based on historical data with a fairly good level of accuracy, so it can be used as decision support in stock management and sales planning.	KNN can classify drug sales levels based on historical transaction characteristics and provide decision support for stock management.	The study does not provide sufficient quantitative performance information in the current summary to directly compare KNN with other algorithms.

3.1.2 Algorithm Distribution Analysis

Table 2. Distribution of Algorithms in Drug Sales Analysis Studies

Algorithm	Number of Studies
C4.5	5
K-Means	5
Naïve Bayes	3

Algorithm	Number of Studies
KNN	4
Apriori	2
FP-Growth	2
Neural Network	1
SVM	1

Note: One study may apply more than one data mining algorithm; therefore, the total number of algorithm applications may exceed the total number of selected studies.

Based on the reviewed studies, various data mining algorithms have been applied in drug sales analysis according to different analytical objectives. C4.5 and K-Means are among the most frequently used algorithms in the reviewed studies, with each appearing in five studies[6], [7]. C4.5 is commonly applied to classification tasks because it can generate decision trees that are relatively easy to understand and interpret, while K-Means is used to group drug sales data based on sales characteristics, demand levels, and distribution patterns [35]. K-Nearest Neighbor (KNN) and Naive Bayes are also utilized in classification and prediction tasks [36], [37], [38]. The reported results indicate that these algorithms can provide useful performance for predicting drug sales trends and identifying product categories, although their performance may vary depending on the dataset, analytical objective, and evaluation method used in each study.

In addition, Apriori and FP-Growth algorithms are used to discover association patterns among drug products, particularly to identify frequently purchased drug combinations that can support promotional strategies and inventory management [39]. FP-Growth reduces the need for repeated candidate itemset generation during frequent itemset mining, which can make it advantageous for processing large transaction datasets [40]. Although Neural Network appears less frequently in the reviewed studies, its application indicates potential for modeling complex and nonlinear relationships in pharmaceutical sales data [8], [41]. SVM also appears in one reviewed study and is applied alongside other algorithms in a biomedical context. Overall, these findings indicate that the selection of a data mining algorithm depends on the analytical objective, characteristics of the dataset, evaluation method, and expected outcomes in supporting decision-making within the pharmaceutical sector.

3.1.3 Advantages and Limitations of Algorithms

Table 3. Advantages and Limitations of Algorithms Used in Drug Sales Analysis

Algorithm	Advantages	Limitations
Naive Bayes	Fast processing, simple implementation, suitable for large datasets.	Assumes independence among attributes, which may not always reflect real conditions.
C4.5	Produces decision trees that are easy to understand and interpret.	Sensitive to noisy data and overfitting in certain cases.
K-Nearest Neighbor (KNN)	Good prediction accuracy and simple concept.	Computationally expensive for large datasets and sensitive to the value of k.
K-Means Clustering	Easy to implement and effective for grouping similar data.	Requires the number of clusters to be determined beforehand and sensitive to initial centroid selection.
Apriori	Effective in discovering association rules and purchasing patterns.	Requires high computational resources when handling large datasets.
FP-Growth	Faster than Apriori in generating frequent itemsets.	More complex implementation and interpretation.
Neural Network	Capable of learning complex patterns and relationships within data.	Requires more training data, longer training time, and higher computational resources.

Based on the reviewed studies, each data mining algorithm has different advantages and limitations depending on the research objectives and characteristics of the dataset. Naive Bayes is advantageous because of its relatively simple implementation and fast processing, but its assumption of independence among attributes may not always reflect the relationships present in pharmaceutical sales data. C4.5 provides interpretable decision trees that can support decision-making and product classification; however, its performance may be affected by noisy data and overfitting in certain conditions [42], [43]. K-Nearest Neighbor (KNN) can be applied to classification and prediction tasks, but its computational requirements may increase

when processing larger datasets [44]. K-Means is useful for grouping drug sales data according to similar characteristics, although the number of clusters must be determined in advance and the results may depend on the initial centroid selection [45]. Apriori and FP-Growth are useful for identifying purchasing patterns and associations among pharmaceutical products. However, Apriori may require substantial computational resources when processing large transaction datasets, whereas FP-Growth can improve the efficiency of frequent itemset generation but has a relatively more complex implementation [46], [47]. Neural Network can model complex patterns and relationships in data, but its application generally requires sufficient training data, longer training processes, and greater computational resources [48], [49]. Therefore, no single algorithm can be considered universally superior for drug sales analysis. The selection of an appropriate algorithm should be based on the research objective, dataset characteristics, evaluation method, and available computational resources [50].

3.2. Comparative Analysis of Data Mining Algorithms in Drug Sales Analysis

Based on the results of the literature review, data mining algorithms have been applied to drug sales analysis for different analytical objectives, including classification, prediction, clustering, and association rule mining. The reviewed studies show that different algorithms are selected according to the characteristics of the data and the objectives of the analysis. The algorithms identified in the reviewed studies include Naive Bayes, Decision Tree C4.5, K-Nearest Neighbor (KNN), K-Means Clustering, Apriori, FP-Growth, and Neural Network. Therefore, the comparative analysis in this section focuses on the strengths, limitations, and suitable applications of each algorithm in drug sales analysis.

The Naive Bayes algorithm is commonly applied as a classification method because of its relatively simple calculation process and fast processing. Research conducted by Hetty Rohayani (2022) reported a classification accuracy of 93.88%, demonstrating the potential of Naive Bayes to provide effective classification results in its respective research context. Furthermore, research by Rifa Maulida et al (2026) demonstrated that the Naive Bayes algorithm can be applied to drug sales classification and can support the identification of drug sales categories and stock requirements. These findings indicate that Naive Bayes can be considered when a simple and computationally efficient classification approach is required. However, its performance depends on the assumption of feature independence, which may not always reflect the relationships among variables in pharmaceutical sales data. Therefore, its suitability should be evaluated according to the characteristics of the dataset and the analytical objective.

The Decision Tree C4.5 algorithm is applied in several studies related to drug sales analysis and has the advantage of producing decision rules that are relatively easy to understand and interpret. Research conducted by M. Afdhal et al. (2022) [15] shows that the C4.5 algorithm can classify product purchasing tendencies based on sales transaction data. Similarly, S. P. Awalina et al. (2023) [17] applied C4.5 to classify drug data into sold and unsold categories. These findings indicate that C4.5 is particularly useful when interpretability is an important consideration in the decision-making process. However, Decision Tree models may be susceptible to overfitting, particularly when trained on complex datasets with numerous attributes. Therefore, appropriate model construction and pruning are important to improve the generalization of C4.5 models.

The K-Nearest Neighbor (KNN) algorithm is applied in several studies to predict drug and medical device sales. Research by A. Azis et al. (2024) [16] shows that KNN can be used to identify best-selling products based on previous sales patterns. Another study by A. Pratiwi (2023) [28] reported an accuracy of 85% in predicting product sales. These findings indicate that KNN can provide useful prediction results; however, its performance is influenced by the characteristics of the dataset and the value of the k parameter used. In addition, the computational cost of KNN may increase as the dataset becomes larger because distance calculations are required for the training instances. Therefore, KNN can be considered for sales prediction and classification, particularly when the dataset size and parameter selection can be appropriately managed.

In the clustering category, K-Means Clustering is frequently applied to group drug sales data based on similar characteristics. Research conducted by M. Fajar et al. (2024) [18], A. Yani et al. (2023) [19], and H. I. Muttaqin (2023) [20] shows that K-Means can group drug sales data according to characteristics such as sales levels, distribution areas, and drug requirements. The resulting clusters can support inventory management, distribution planning, and promotional strategies by providing information about groups of drugs with similar sales characteristics. However, the quality of K-Means clustering depends on the appropriate selection of the number of clusters (k) and the characteristics of the data being analyzed. Therefore, the resulting clusters should be evaluated based on their relevance to the analytical objective rather than solely on the formation of distinct groups.

In addition to classification and clustering, several studies have applied the Apriori and FP-Growth algorithms to identify purchasing patterns among pharmaceutical products. Research by P. A. Lestari et al. (2024) [21] and Salam Irianto Nadelak and Yusmar Ali (2021) [22] identified drug combinations that were

frequently purchased together. Meanwhile, research by A. M. Hutabarat et al (2020) [29] applied FP-Growth to generate frequent itemsets that can support product bundling and sales promotion strategies. These findings indicate that association rule mining can provide useful information about relationships among pharmaceutical products and can support marketing strategies and inventory management. From a methodological perspective, FP-Growth can reduce the need for repeated candidate itemset generation compared with the Apriori approach, which may provide computational advantages when processing large transactional datasets. However, the performance of both algorithms depends on the characteristics of the transaction data and the parameters used in the mining process.

Table 4. Comparative Summary of Data Mining Algorithms Applied in Pharmaceutical Sales Analysis

Algorithm	Strengths	Limitations	Suitable Applications
Naive Bayes	Fast computation and simple implementation	Assumes feature independence; performance may decrease when features are highly correlated	Drug sales classification, stock prediction
Decision Tree C4.5	Easy to interpret and provides understandable decision rules	Susceptible to overfitting on complex datasets	Product classification, decision support
K-Nearest Neighbor (KNN)	Simple concept and can provide useful prediction results	Sensitive to the value of k and computationally costly for large datasets	Sales prediction, product classification
Neural Network	Capable of modeling complex and nonlinear relationships	Requires sufficient training data, longer training time, and greater computational resources	Complex pharmaceutical sales prediction and demand forecasting
K-Means Clustering	Effective for grouping similar sales characteristics and inventory segmentation	Requires the number of clusters (k) to be determined in advance	Sales segmentation, inventory grouping
Apriori	Identifies association rules and frequently purchased products	May generate many candidate itemsets and require substantial computational resources for large datasets	Market basket analysis, purchasing pattern analysis
FP-Growth	Efficient frequent itemset mining without repeated candidate generation	More complex implementation and interpretation	Large transaction database analysis, purchasing pattern analysis

When compared according to their analytical objectives, the reviewed algorithms show different areas of applicability. Naive Bayes, Decision Tree C4.5, and KNN are mainly applied to classification and prediction tasks, although the reported performance varies across datasets and evaluation settings. K-Means is primarily used for clustering drug sales data and can support the identification of groups with similar sales or inventory characteristics. Meanwhile, Apriori and FP-Growth are applied to association rule mining to identify relationships and frequently purchased combinations among pharmaceutical products. Neural Network is less frequently represented in the reviewed studies but provides an alternative for modeling complex relationships in sales-related data. These findings suggest that algorithm selection should be determined by the analytical objective rather than by the algorithm name or a single reported performance value.

The comparative analysis indicates that no single algorithm consistently outperforms others across all pharmaceutical sales scenarios. Although some studies reported relatively high accuracy values, such as 93.88% for Naive Bayes and 85% for KNN, these values cannot be interpreted as evidence that one algorithm is superior to another because the studies used different datasets, objectives, and evaluation settings. Therefore, the selection of a data mining algorithm should consider the analytical objective, dataset characteristics, evaluation method, computational requirements, and expected decision-making outcomes. This comparison highlights the importance of conducting future comparative studies in which multiple algorithms are evaluated using the same pharmaceutical sales dataset and standardized evaluation measures.

3.3 Factors Affecting Algorithm Performance

The performance of data mining algorithms in pharmaceutical sales analysis is influenced not only by the selected algorithm but also by several technical and domain-specific factors, including dataset characteristics, preprocessing techniques, feature selection, and business conditions.

One of the most influential factors is the size and complexity of the dataset. Algorithms such as Neural Network can be applied to large and complex datasets because they are capable of modeling nonlinear relationships among variables. In contrast, algorithms such as Naive Bayes and K-Nearest Neighbor (KNN) can be applied to relatively small and medium-sized datasets and may provide computational advantages due to their simpler model structures. Therefore, selecting an appropriate algorithm should consider the volume and characteristics of the available pharmaceutical sales data.

Data preprocessing also plays an essential role in preparing data for analysis. Preprocessing activities, including data cleaning, normalization, duplicate removal, and handling missing values, can reduce data inconsistencies and noise that may affect the results of data mining. Appropriate preprocessing is therefore important for obtaining more reliable classification, prediction, and clustering results. Consequently, preprocessing should be regarded as an important stage before implementing any data mining algorithm.

Another important consideration is feature selection. Selecting relevant variables from pharmaceutical transaction data enables algorithms to focus on meaningful information while reducing computational complexity. Irrelevant or redundant attributes may decrease prediction accuracy and increase processing time. Therefore, appropriate feature selection can contribute to more reliable classification and prediction results.

Furthermore, class imbalance may occur in pharmaceutical sales datasets when certain categories of medicines dominate transaction records. This condition may affect classification performance because prediction models can tend to favor majority classes. Therefore, balancing techniques, resampling methods, and appropriate evaluation metrics should be considered when working with imbalanced datasets.

Overall, the reviewed studies indicate that achieving high predictive performance requires not only selecting an appropriate algorithm but also considering dataset characteristics, preprocessing quality, feature engineering, and evaluation strategies. These factors collectively determine the effectiveness of data mining techniques in supporting pharmaceutical sales analysis and decision-making.

Moreover, algorithm performance may also be influenced by domain-specific characteristics of pharmaceutical sales data. Seasonal demand fluctuations, promotional campaigns, product expiration periods, and changes in customer purchasing behavior may affect sales patterns and, consequently, the results of data mining analysis. Therefore, integrating domain knowledge with data mining techniques can provide more contextually relevant analysis and support decision-making for pharmaceutical inventory management.

3.4 Research Gaps and Future Research Directions

Although numerous studies have investigated the application of data mining algorithms in pharmaceutical sales analysis, several important research gaps remain to be addressed. Most existing studies primarily focus on traditional algorithms such as Naive Bayes, C4.5, K-Nearest Neighbor (KNN), K-Means, Apriori, and FP-Growth, while relatively limited attention has been given to more advanced approaches, including ensemble learning methods, hybrid models, and deep learning techniques.

Another limitation observed in the reviewed studies is the lack of standardized datasets and evaluation metrics. Many researchers employ different datasets, preprocessing techniques, and performance indicators, making direct comparison among algorithms difficult. Consequently, future studies should consider using benchmark datasets and standardized evaluation criteria to improve the reliability and comparability of research findings. The adoption of publicly available benchmark datasets would facilitate more objective evaluations of algorithm performance while improving the reproducibility and transparency of future research.

In addition, only a limited number of studies discuss practical implementation challenges, such as data privacy, computational efficiency, real-time data processing, and scalability in pharmaceutical information systems. Addressing these challenges is important because pharmaceutical sales systems may generate increasing volumes of transaction data that require efficient, scalable, and secure analytical approaches.

Future studies are encouraged to conduct comparative evaluations of multiple data mining algorithms using the same pharmaceutical sales dataset and standardized evaluation metrics. In addition, emerging approaches such as ensemble learning, hybrid machine learning models, Explainable Artificial Intelligence (XAI), and Automated Machine Learning (AutoML) can be explored to improve analytical performance and model interpretability. The integration of machine learning and data mining techniques with pharmaceutical information systems and real-time business intelligence systems may also provide opportunities for more timely and practical decision support in pharmaceutical inventory management and drug sales forecasting.

Overall, these research opportunities highlight the need for more comprehensive, scalable, and practically applicable data mining solutions in pharmaceutical sales analysis. Future investigations should not only focus on improving predictive accuracy but also emphasize interpretability, computational efficiency,

scalability, and real-world implementation to maximize the benefits of data-driven decision-making in the pharmaceutical sector. Therefore, addressing these research gaps may improve the performance, interpretability, computational efficiency, and practical applicability of data mining techniques in pharmaceutical sales management and decision support systems.

4. CONCLUSION

This systematic literature review synthesized evidence from twenty selected studies published between 2019 and 2026 to examine the application of various data mining algorithms in pharmaceutical sales analysis. The findings indicate that data mining techniques have become valuable tools for supporting sales prediction, inventory management, customer purchasing pattern analysis, and decision-making within the pharmaceutical sector. Among the reviewed studies, classification algorithms such as Naive Bayes, C4.5, and K-Nearest Neighbor (KNN) were frequently employed for prediction and classification tasks, while K-Means, Apriori, and FP-Growth were commonly applied for clustering and association rule mining. Each algorithm demonstrated specific strengths depending on the analytical objectives and dataset characteristics.

The comparative analysis further revealed that no single algorithm consistently outperforms others across all pharmaceutical sales analysis tasks. Naive Bayes, C4.5, and KNN are suitable for classification and prediction tasks, while K-Means is more appropriate for grouping pharmaceutical products according to similar sales characteristics. Apriori and FP-Growth are particularly useful for identifying associations and frequently purchased combinations among pharmaceutical products. However, these findings should not be interpreted as evidence that one algorithm is universally superior because the reviewed studies used different datasets, objectives, and evaluation methods. Therefore, algorithm selection should be aligned with the specific analytical objective, dataset characteristics, evaluation method, and operational requirements of pharmaceutical organizations.

This study contributes to the existing literature by providing a systematic synthesis of commonly applied data mining algorithms in pharmaceutical sales analysis and identifying their strengths, limitations, and practical applicability. The findings provide a basis for researchers and practitioners to select data mining approaches according to specific analytical objectives and data characteristics. The reviewed applications also indicate that data mining techniques can support pharmaceutical inventory management, demand analysis, marketing strategies, and evidence-based decision-making when appropriately implemented.

Despite these contributions, this review has several limitations. First, it only includes studies published between 2019 and 2026. Second, the review was limited to articles retrieved from selected academic databases, which may have excluded other relevant publications. Third, the number of selected studies was limited to twenty articles, and the relatively small number of studies may affect the breadth of the synthesis. Fourth, the reviewed studies used different datasets, research objectives, preprocessing techniques, algorithms, and evaluation metrics, which limited the possibility of making direct performance comparisons among algorithms. These methodological differences should therefore be considered when interpreting the findings of this review.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all individuals and institutions who provided valuable support, guidance, and encouragement throughout the completion of this research.

REFERENCES

- [1] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*. Cambridge University Press, 2020.
- [2] I. D. Dinov, *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*. Springer, 2023.
- [3] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2021.
- [4] V. Kotu and B. Deshpande, *Data Science: Concepts and Practice*. Morgan Kaufmann, 2019.
- [5] C. C. Aggarwal, *Data Mining: The Textbook*. Springer, 2021.
- [6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2021.
- [7] E. Alpaydin, *Introduction to Machine Learning*, 4th ed. MIT Press, 2020.
- [8] G. Bonaccorso, *Machine Learning Algorithms*. Packt Publishing, 2020.
- [9] S. Raschka and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn*. Packt, 2022.
- [10] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2022.
- [11] S. Chakrabarti, *Data Mining for Business Analytics*. Wiley, 2021.
- [12] J. D. Kelleher, B. Mac Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics*. MIT Press, 2020.
- [13] H. Rohayani, Rico, K. Kurniawansyah, and A. S. Harahap, "Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Sains dan Teknologi Menggunakan Metode Naive Bayes," *J. Vis. Technol.*, vol. 5, no. 2, pp. 35–41, 2022.

- [14] H. Rohayani and M. C. Umam, "Prediksi Penentuan Program Studi Berdasarkan Nilai Siswa dengan Algoritma Backpropagation," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 651–657, 2022.
- [15] M. Afdhal, V. Ariandi, and Rita, "Memprediksi Penjualan pada Toko Hanifah Menggunakan Metode C4.5," 2022.
- [16] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," 2024.
- [17] S. P. Awalina, R. Helilintar, and A. B. Setiawan, "Penerapan Algoritma C4.5 untuk Prediksi Penjualan Obat di Apotek Warujayeng," 2023.
- [18] M. Fajar, N. Rahaningsih, and R. Danar Dana, "Analisis Pola Penjualan Obat Di Apotek an-Naafi Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 486–492, 2024, doi: 10.36040/jati.v8i1.8395.
- [19] A. Yani, Z. Azmi, and D. Suherdi, "Implementasi Data Mining Menganalisa Data Penjualan Menggunakan Algoritma K-Means Clustering," 2023.
- [20] H. I. Muttaqin, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Data Obat di Puskesmas Unter Iwes Sumbawa," 2023.
- [21] P. A. Lestari, M. Nasution, and S. Z. Harahap, "Analisa Data Penjualan pada Apotek Ritonga Farma Menggunakan Data Mining Apriori," 2024.
- [22] S. I. Nadeak and Y. Ali, "Analysis of Data Mining Associations on Drug Sales at Pharmacies with Apriori," 2021.
- [23] S. Muntari, "Penerapan Data Mining Menggunakan Algoritma FP-Growth untuk Menganalisa Pola Penjualan Obat," 2024.
- [24] B. Kitchenham and S. Charters, "Guidelines for Performing Systematic Literature Reviews in Software Engineering," 2019.
- [25] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*. Wiley, 2020.
- [26] M. J. Page, "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews," *BMJ*, vol. 372, 2021.
- [27] A. J. Kahfi, "Analisis Data Penjualan Produk Herbal Menggunakan Algoritma C4.5 pada E-Commerce Azka Jaisy Store," 2025.
- [28] A. Pratiwi, "Optimasi Prediksi Penjualan Produk dengan Algoritma K-Nearest Neighbor," 2023.
- [29] A. M. Hutabarat, D. Saripurna, and Azlan, "Implementasi FP-Growth dalam Menganalisa Penjualan Obat," 2020.
- [30] R. Maulida, N. G. Setyoningrum, and Alam, "Penerapan Algoritma Naive Bayes Classifier untuk Klasifikasi Penjualan Obat di Apotek Nugraha," 2026.
- [31] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [32] M. Aisyah and W. Prihartono, "Analisis Pengelompokan Penjualan Obat Menggunakan Algoritma K-Means untuk Optimalisasi Stok Obat pada Apotek OLE," 2025.
- [33] R. Avrizar and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma C4.5 pada Penjualan Obat," 2019.
- [34] Nurhayati and M. Iqbal, "Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Data Penjualan Obat pada Apotek," 2022.
- [35] J. Prasetyana and F. Charlos, "Implementation of K-Means Algorithm in Data Mining for Drug Market Segmentation," 2026.
- [36] P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," *Mach. Learn.*, vol. 29, no. 2–3, pp. 103–130, 1997.
- [37] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [38] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2023.
- [39] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994.
- [40] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," in *ACM SIGMOD International Conference on Management of Data*, 2000.
- [41] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2022.
- [42] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, 1986.
- [43] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. MIT Press, 2022.
- [44] R. M. Febrialdo and Y. Wendra, "Penerapan Data Mining Memprediksi Penjualan Obat Menggunakan Metode K-Nearest Neighbor (Studi Kasus: Apotek Difana)," 2025.
- [45] K. Handoko, P. Simanjuntak, and E. Elisa, "Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori Pada Kimia Farma Batam," 2025.
- [46] N. T. Ayu, Jasmir, and I. S. Wijaya, "Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Persediaan Stok Obat Pada Apotek Safa," 2024.
- [47] C. Adam and K. Handoko, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan," 2023.
- [48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2023.
- [49] A. Saputra and B. Nugroho, "Prediksi Penjualan Obat Menggunakan Algoritma Neural Network pada Sistem Data Mining," 2024.
- [50] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer, 2021.