Scientific Journal of Informatics
Vol. 11, No. 3, Aug 2024

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Optimizing Inventory Management: Data-Driven Insights from K-Means Clustering Analysis of Prescription Patterns

Aulia Agung Dermawan[1], Ansarullah Lawi[2], Dimas Akmarul Putera[3], Dwi Ely Kurniawan[4*], Kuntum Khoiro Ummatin[5], Jorvick Steve[6]
[1, 3, 6]Departement of Engineering Management, Institut Teknologi Batam, Indonesia,
[2]Department of Industrial Engineering, Institut Teknologi Batam, Indonesia,
[4]Department of Informatic Engineering, Politeknik Negeri Batam, Indonesia
[5]Department of Engineering Management, Universitas Internasional Semen, Indonesia

Abstract.
Purpose: The goal is to improve how inventory is managed in healthcare by using K-Means clustering to analyze prescription trends. This approach helps ensure better stock availability, streamlines operations, and ultimately increases sales opportunities.
Methods: This research applied the K-Means clustering algorithm to analyze a comprehensive dataset of prescription behaviors from XYZ Clinic. By grouping similar prescriptions into clusters, this method highlighted patterns within the data. These insights led to the identification of unique prescription categories, enabling the creation of tailored recommendations for improving inventory management.
Result: The analysis showed that Cluster 1 should be prioritized for inventory management due to its high sales potential and consistent prescription patterns. It is recommended to increase stock for the medications in Cluster 1 to improve inventory turnover and streamline clinical operations. These findings underscore the value of K-Means clustering in healthcare, especially for enhancing inventory management and operational efficiency.
Novelty: This research presents a novel application of K-Means clustering in healthcare, focusing on prescription patterns and inventory management. While previous studies have primarily used K-Means clustering for areas such as risk assessment and logistics, this study provides valuable data-driven insights to improve inventory management strategies in healthcare. The results highlight how clustering methods can support better decision-making and resource allocation, ultimately leading to greater operational efficiency and improved patient care.

Keywords: Inventory management, K-Means clustering, Prescription patterns, Operational efficiency
Received July 2024 / **Revised** September 2024 / **Accepted** September 2024

INTRODUCTION

The advancement of information systems has had a significant impact on various fields, including education, entertainment, healthcare, and business. In healthcare, substantial investments are being made to enhance these systems, with the goal of improving the quality of patient care and increasing the efficiency of cost management.. [1] Health Information Systems (HIS) and Health Informatics (HI) are essential in harnessing health information technology to enhance the efficiency and quality of healthcare services. They also pave the way for new advancements and innovations in the field. [2] The integration of diverse healthcare systems with business process management systems is increasingly acknowledged for its potential to streamline operations and improve efficiency across various medical institutions. [3] Within this framework, data mining plays a critical role as an analytical approach that explores large datasets to uncover hidden patterns and relationships between variables. This process involves multiple stages, including data selection, cleaning, applying algorithms, and evaluating the results. [4], It helps to summarize information in creative ways, making it more comprehensible and beneficial for end-users [5].

Data mining specifically includes tasks such as frequent pattern mining, clustering, classification, association rule mining, and regression analysis [6]; [7], These tasks are crucial for unlocking the full potential of health information systems to further improve healthcare delivery. Efficient drug requirements planning is especially important for managing medications across various healthcare settings. This process involves drug selection, quantification, procurement, distribution, and usage, all of which directly impact

*Corresponding author.
Email addresses: agung@iteba.ac.id (Dermawan), dwialikhs@polibatam.ac.id (Kurniawan)*
DOI: 10.15294/sji.v11i3.8690

the availability and timely delivery of medications. [8]. Overstocking of medications can interfere with the drug management system, resulting in inefficiencies and possible wastage [9]. By applying data mining techniques like clustering, healthcare institutions can improve drug inventory management, ensuring timely and efficient delivery while minimizing waste. This not only enhances overall healthcare outcomes but also addresses the challenges of fluctuating drug demand. Effective inventory management is a critical element of supply chain operations, particularly in the pharmaceutical industry, where demand can vary significantly. [10]. Poor management of drug inventory can lead to interruptions in drug availability, negatively affecting customer satisfaction and sales performance [11]. Running out of drug supplies during periods of high demand can result in delayed or canceled sales, which directly impacts the organization's revenue [11]. By combining data mining and clustering techniques, healthcare institutions can better predict demand trends, adjust inventory levels accordingly, and maintain a balanced supply chain that meets patient needs while minimizing waste and financial losses. Accurate sales forecasting is especially important for clinics and pharmaceutical distribution companies, where reliable predictions are key to managing inventory efficiently and ensuring financial stability.

The main challenge clinics encounter is inaccurate sales forecasting, which can result in excessive inventory buildup and financial losses from increased storage costs. [12]; [13]). This issue goes beyond clinics, as accurate sales forecasting is essential in many industries to avoid both overstocking and understocking, which can result in significant financial losses. [14] ;[15]. By using advanced data mining techniques, organizations can greatly improve the accuracy of their sales forecasts, leading to better inventory management and stronger financial performance. To further streamline drug inventory, the K-means clustering method can be applied to identify and group frequently prescribed medications. K-means is a powerful tool for analyzing large datasets and efficiently categorizing similar items. [16] By using clustering techniques like K-means, hospitals can categorize drugs based on how frequently they are prescribed, making it easier to manage drug inventories more efficiently and maintain optimal stock levels. [17]. K-means clustering is commonly applied because of its simplicity and rapid convergence, making it an ideal choice for efficiently and quickly grouping data [18]. This approach allows for the grouping of similar items, and when applied to drug inventories, it can reveal patterns in prescription habits, leading to more efficient inventory management. Moreover, K-means clustering can identify groups of drugs that are often prescribed together, helping to optimize stock levels for these specific categories and ensuring that essential medications are readily available. [17]. In this context, the study evaluated the effectiveness of the K-means algorithm for planning drug requirements in health clinics.

Previous research has laid the groundwork for conducting drug clustering tests using the K-means method, as evidenced by several studies. For example, a study from China assessed the risk of road collapse during and after tunnel construction by using an enhanced entropy weight method and K-means clustering. This research combined entropy weight and analytic hierarchy process (AHP) techniques to reduce bias, while the K-means algorithm improved the classification of evaluation outcomes, demonstrating the versatility of K-means in various fields, including healthcare. [19]. In another study from India, researchers introduced an innovative method that combines spectral clustering algorithms with a quadratic support vector machine (SVM) to predict learning styles on an e-learning platform. The research involved gathering data from log files, extracting features using web mining, clustering these features, and predicting learning styles through the quadratic SVM. The proposed method was tested on real-time datasets and demonstrated improved performance compared to existing techniques, highlighting its potential for more accurate learning style predictions. [20]. In a collaborative study between institutions in Malaysia and China, researchers explored the integration of smart city information technology with data mining algorithms to detect accounting fraud. By applying K-means clustering, the study was able to identify abnormal financial patterns and fraudulent activities in accounting data, leading to a significant reduction in misjudgment rates compared to traditional methods. This highlights the effectiveness of combining advanced technology with data analysis techniques in fraud detection. [21]. Additionally, a study from Taiwan examined the use of simulation optimization to reduce medication inventory costs in an outpatient pharmacy. The research combined a two-stage clustering method with the (s, S) inventory policy to create a simulation optimization model. By using Arena OptQuest software, the study identified the optimal minimum and maximum stock levels for each medication. This approach led to a remarkable 55% reduction in average inventory costs and a 68% decrease in inventory volume, all while maintaining continuous patient care. [17]. Moreover, a study from China focused on improving order allocation efficiency in the logistics service supply chain and optimizing customer satisfaction. The researchers developed a cloud-based logistics service supply chain, analyzed the order allocation process, and introduced a hybrid approach that combines K-means clustering with Quality of

Service (QoS) matching. Through simulation experiments, they evaluated the algorithm's recall ratio and precision, demonstrating its effectiveness in preventing service regressions while meeting customer needs. [22].

Our study presents a novel use of K-means clustering to analyze prescription patterns in a healthcare setting. While previous research has mainly focused on applying K-means clustering in areas like infrastructure risk assessment, educational technology, financial fraud detection, pharmacy inventory optimization, and logistics supply chain management, this study emphasizes its innovative application within healthcare. Specifically, it utilizes data mining techniques to identify and classify prescription patterns in clinical environments, offering new insights for improving healthcare operations.

METHODS

Knowledge discovery in databases (KDD)

KDD (Knowledge Discovery in Databases) has emerged as an essential field in response to the challenges created by the rapid accumulation of data, which surpasses both technical and human abilities to process and interpret it effectively ("Knowledge Discovery in Database Systems", 2017). The goal of KDD is to extract meaningful and novel insights from large datasets, providing valuable knowledge that can inform decision-making and improve various processes. [23]. A key component of KDD is data mining, which involves identifying patterns within structured, semi-structured, and unstructured data, contributing to the broader process of knowledge discovery [24]. The KDD process is described as tedious and repetitive, highlighting the complexity involved in uncovering meaningful insights from data [25]. KDD is a systematic and iterative process aimed at extracting valuable and novel patterns from large datasets [26]. This methodology involves multiple interactive stages that require the use of data mining techniques to identify meaningful patterns [27]. Data mining, as a crucial component of KDD, facilitates the transformation of data into actionable [27]. The KDD process is user-centric, emphasizing interactivity and iteration to reveal new insights [28]. The evolution of KDD has led to the recognition of data mining as a subprocess within the broader framework of knowledge discovery, underscoring the significance of deriving high-level insights through systematic processes [29]. The KDD process involves several key activities, including data selection, cleaning, transformation, data mining, and analysis/assimilation. These steps work together to ensure the effective extraction of valuable knowledge from large datasets, helping to uncover patterns and insights that can be applied to real-world challenges [30].

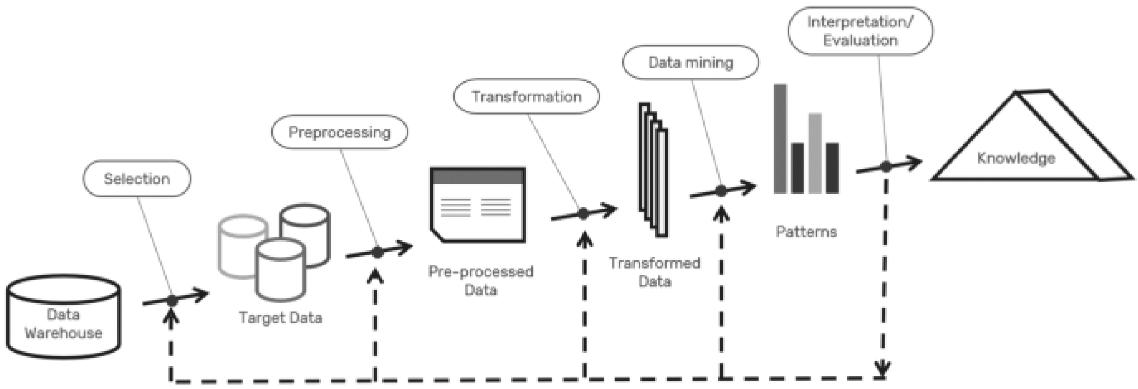

Figure 1. Stages of knowledge discovery in databases process

To initiate the knowledge discovery process, several essential steps must be carefully carried out [31]. The first step, Data Selection, involves identifying and extracting relevant information from existing databases. This foundational phase focuses on choosing data that holds potential value for future analyses. Following this, Data Cleaning is essential to ensure the quality of the data by addressing issues like completeness, accuracy, and integrity, which are critical for producing reliable results in later stages of analysis. [32]. This phase removes inconsistencies and non-standard elements that could skew analysis results.

Once data selection and cleaning are complete, Data Integration brings together information from various sources, ensuring coherence and consistency throughout the combined dataset. Following this, Data Transformation adjusts the data format to meet the needs of the analytical algorithms that will be applied next [33]. This step prepares the data effectively for advanced data mining techniques.

Data Mining serves as the heart of the KDD process, using advanced techniques to discover meaningful patterns and insights from the prepared dataset. Once extracted, these patterns undergo Patterns Evaluation, where they are rigorously assessed against predefined criteria to highlight the most valuable findings. Finally, Knowledge Presentation ensures that these insights are communicated clearly, often through visual tools like graphs, diagrams, or detailed reports. This systematic approach ensures that valuable knowledge is not only extracted and evaluated but also presented in a way that supports informed decision-making and actionable insights.

These steps are vital components of the knowledge discovery process, facilitating the extraction of meaningful insights and patterns from large datasets. Researchers highlight the significance of these stages in ensuring the validity [34], novelty [35], and utility [36] of the discovered knowledge [37] [38] [39] [40]. By systematically adhering to these stages, KDD ensures effective data processing that reveals meaningful and actionable insights, addressing the challenges associated with the growing volume of data.

The k-means (KM) clustering algorithm

The K-means (KM) clustering technique divides a multidimensional data space into several clusters, with each cluster centered around a randomly assigned centroid. The data points are grouped based on their proximity to the closest centroid, ensuring that each cluster contains data points most similar to its centroid. This process helps to identify natural groupings within the data, simplifying complex datasets for further analysis. [41]. The KM algorithm assigns each observation to the nearest cluster in a dataset of d-dimensional vectors X = {x1,x2,…,xd}. It uses the Euclidean distance that exists between a finding and a cluster centroid as a metric to minimise the sum of squares generated by the objective function. [42]:

1. Calculate k as the number of clusters to be generated using the elbow criteria approach specified by the following formula.

$$SSE = \sum(k = 1)^K \sum(x_i = S_k)\left||N_i - C_k|\right|^2 \tag{1}$$

2. Establish the first cluster centre point (centroid) by randomly selecting k. An initial centroid is randomly determined from a pool of items accessible in up to k clusters. The next ith centroid cluster is calculated using the following formula.

$$v = (\sum(i = 1)^n x_i)/n, i = 1,2,3, \dots, n \tag{2}$$

3. To get the distance between each item and every centroid of each cluster, use Euclidean Distance and use the following formula.

$$d(x, y) = \left||x - y|\right| = \sqrt{(\sum(i = 1)^n (x_i - y_i)^2)}, i = 1,2,3, \dots, n \tag{3}$$

4. Assign each item to the closest centroid. The assignment of objects to clusters during iteration is often done using hard k-means. In this method, each object is explicitly designated as a member of a cluster based on its closeness to the cluster's centre point.
5. Execute an iteration and then ascertain the location of the newly formed centroid by reference to the equation.
6. Proceed to repeat the third step if the altered centroid is not positioned identically.

The davies-bouldin index (DBI)

The Davies-Bouldin Index (DBI), introduced by David L. Davies and Donald W. Bouldin in 1979, is a metric used to evaluate the quality of clustering. It is calculated as the average ratio of intra-cluster distances (how compact each cluster is) to inter-cluster distances (how separate clusters are) for each cluster relative to its nearest neighbor. A lower DBI value indicates better clustering, with tighter groupings and more distinct separation between clusters. The DBI shows a positive correlation in "within-cluster" scenarios and a negative correlation in "between-cluster" scenarios. It is highly regarded for internal validation, one of the two primary methods of cluster validation (the other being external validation), both of which are crucial for accurately assessing the outcomes of clustering analyses. [43].

RapidMiner
RapidMiner is an interactive platform designed for machine learning and data mining tasks. As a free, open-source project developed in Java, it uses a modular approach to handle complex problems, enabling users to build advanced chains of operators for various computational tasks. The platform uses XML to define operator trees, which represent knowledge discovery (KD) processes. RapidMiner offers flexible operators for importing and exporting data in multiple file formats and includes over 100 algorithms tailored for tasks such as classification, regression, and clustering [44].

RESULT AND DISCUSSION

This research focuses on analyzing the sales data from XYZ Clinic between January and April 2024, as shown in Table 1. The data is sourced directly from the clinic's sales tracking system, ensuring its accuracy and reliability. It includes essential attributes like product names and quantities sold during this period. The dataset, consisting of 44 different products, has been carefully reviewed to eliminate any errors or inconsistencies, providing a solid foundation for analysis. Each entry includes details such as the sale date, product name, quantity sold, and relevant patient information. The purpose of this study is to identify sales trends and consumption patterns among patients, with the results expected to support better decision-making in inventory management and marketing strategies at XYZ Clinic.

Table 1. Product sales data set

No	Product Name	Number of Sales			
		Jan	Feb	Mar	April
1	Oxy blue cream	71	69	102	124
2	Oxy blue cream blue	50	96	126	90
3	Cream skin barrier	96	70	11	123
4	Skin care	108	40	61	13
5	Acne cream night	34	94	111	84
	…	…	…	…	…
43	Baby skin-3	82	61	11	102
44	Placental serum 10ml	80	66	99	18

Use of k-means algorithm
The K-Means algorithm is used in this study to tackle the challenge of segmenting or categorizing medications into different clusters. The authors tested five different configurations of the K-Means operator, varying the number of clusters between 2, 3, 4, 5, and 6. The most optimal configuration was selected based on an evaluation using the Davies-Bouldin Index. This approach helps address the known sensitivity of K-Means to the number of clusters, as changes in cluster numbers can significantly impact segmentation results. Determining the right number of clusters is crucial for achieving accurate segmentation outcomes.

To further enhance the analysis, the authors also considered the Elbow Method, which calculates and compares WCSS (Within-Cluster Sum of Squares) values to determine the optimal number of clusters. Additionally, the study explored the use of the k-Medoids technique as an alternative to improve the accuracy of the segmentation process.

Table 2. Davies bouldin index evaluation results

Cluster	Davies Bouldin
2	-1.639
3	-1.538
4	-1.458
5	-1.220
6	-1.158

From the table above, it can be seen that K-Means with 6 clusters has a smaller value compared to the number of other clusters. This indicates that configurations with 6 clusters provide the best performance. Thus, the number of such clusters can be considered the most optimal and recommended for use in the process of clustering customer data.

Clustering results based on product sales

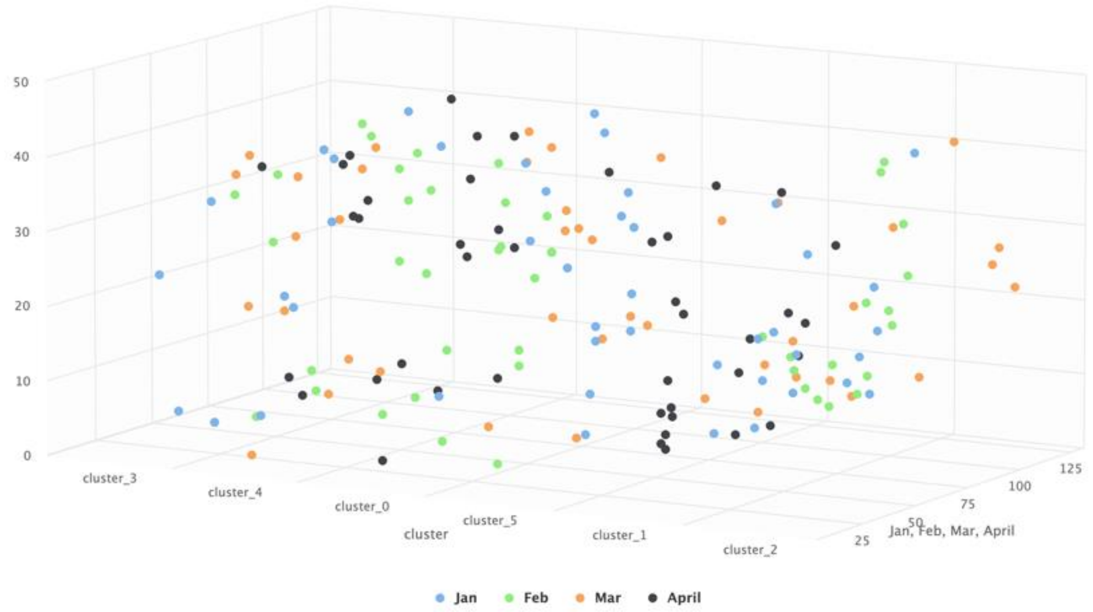

Figure 2. The different between gaussian and salt & papper noise

Based on the figure above it can be seen that the clusters with the highest expenditure are:

Cluster 0: There are 9 Product items.
Cluster 1: There are 10 Product items.
Cluster 2: There are 6 Product items.
Cluster 3: There are 6 Product items.
Cluster 4: There are 7 Product items.
Cluster 5: There are 6 Product items.

The total number of items from all clusters is 44. Based on this data, cluster 1 has the highest number of items, which probably shows the highest expenditure among other clusters.

Figure 3. Matrix scatter plot

The figure above illustrates the distribution of drug sales data at XYZ Clinic from January to April, corresponding to the clustering analysis we conducted earlier. This graph offers a detailed perspective on the fluctuations in daily or weekly sales within each month. The matrix uses points plotted on two-dimensional coordinates, where the X-axis and Y-axis may represent variables like sales volume and transaction frequency or other relevant metrics. This visual representation helps to highlight patterns and trends in the clinic's sales data.

1) Distribution in Months: Within each panel from January to April, there is a wide distribution of points, showing variations in sales data. Points that are more concentrated in a particular area may indicate a more stable sales pattern or a high frequency of transactions at a certain sales value.
2) Comparison Between Months: The comparison between panels shows a difference in the distribution of points from one month to another. This indicates a change in sales patterns, which could be due to seasonal factors, promotions, or changes in sales policy.
3) Outliers: Some points that are far from the main concentration can be considered as outliers, which may indicate unusual transactions or extraordinary events affecting sales.

CONCLUSION

The cluster visualization results show that Cluster 1 contains the highest number of product items (10 items), potentially indicating the highest expenditure compared to other clusters. Across the 44 product items, they are distributed among 6 clusters: Cluster 0 (9 items), Cluster 2 (6 items), Cluster 3 (6 items), Cluster 4 (7 items), and Cluster 5 (6 items). Given its larger number of items, Cluster 1 is identified as the most significant cluster. It is recommended that clinics focus on increasing stock for products within this cluster to improve inventory management, sales planning, and overall clinical operational efficiency.

REFERENCES

[1] T. M. Lee, A. H. Ghapanchi, A. Talaei-Khoei, and P. Ray, "Strategic Information System Planning in Healthcare Organizations," *J. Organ. End User Comput.*, vol. 27, no. 2, pp. 1–31, 2015, doi: 10.4018/joeuc.2015040101.

[2] H. Jabareen, Y. Khader, and A. Taweel, "Health Information Systems in Jordan and Palestine: The

Need for Health Informatics Training," *East. Mediterr. Heal. J.*, vol. 26, no. 11, pp. 1323–1330, 2020, doi: 10.26719/emhj.20.036.
[3] G.-W. Kim, K.-W. Park, H. Hong, and D.-H. Lee, "Process Model Verifier for Integrated Medical Healthcare Systems Using Business Process Management System," 2014, doi: 10.1109/isce.2014.6884298.
[4] M. Sasikala, M. Deepika, and M. S. S. Shankar, "Pattern Identification and Predictions in Data Analysis," *Int. J. Eng. Comput. Sci.*, vol. 7, no. 03, pp. 23686–23691, 2018, doi: 10.18535/ijecs/v7i3.05.
[5] L. Zhang, K. Liu, I. Ilham, and J. Fan, "Application of Data Mining Technology Based on Data Center," *J. Phys. Conf. Ser.*, vol. 2146, no. 1, p. 12017, 2022, doi: 10.1088/1742-6596/2146/1/012017.
[6] Q. Shen and L. Gao, "A Novel Selection and Matching of Patterns Model for Traffic Prediction in Wireless Network," *Destech Trans. Comput. Sci. Eng.*, no. wicom, 2018, doi: 10.12783/dtcse/wicom2018/26301.
[7] P. Patil and S. Ratnoo, "Gravitational Search Algorithms in Data Mining: A Survey," *Ijarcce*, vol. 6, no. 6, pp. 168–173, 2017, doi: 10.17148/ijarcce.2017.6631.
[8] A. R. Fahriati, D. S. Suryatiningrum, and T. J. Saragih, "Inventory Control of Drugs Listed in Private Health Insurance at Pharmacies in South Tangerang Using ABC Analysis," *Pharmacol. Clin. Pharm. Res.*, vol. 6, no. 1, p. 18, 2021, doi: 10.15416/pcpr.v6i1.31541.
[9] S. K. Sahu *et al.*, "Effectiveness of Supply Chain Planning in Ensuring Availability of CD/NCD Drugs in Non-Metropolitan and Rural Public Health System," *J. Health Manag.*, vol. 24, no. 1, pp. 132–145, 2022, doi: 10.1177/09720634221078064.
[10] J.-C. B. Munyaka and S. V Yadavalli, "Inventory Management Concepts and Implementations: A Systematic Review," *South African J. Ind. Eng.*, vol. 32, no. 2, 2022, doi: 10.7166/33-2-2527.
[11] T. A. Zwaida, C. Pham, and Y. Beauregard, "Optimization of Inventory Management to Prevent Drug Shortages in the Hospital Supply Chain," *Appl. Sci.*, vol. 11, no. 6, p. 2726, 2021, doi: 10.3390/app11062726.
[12] Rasim, E. Junaeti, and R. Wirantika, "Implementation of Automatic Clustering Algorithm and Fuzzy Time Series in Motorcycle Sales Forecasting," *Iop Conf. Ser. Mater. Sci. Eng.*, vol. 288, p. 12126, 2018, doi: 10.1088/1757-899x/288/1/012126.
[13] N. K. Zadeh, M. M. Sepehri, and H. Farvaresh, "Intelligent Sales Prediction for Pharmaceutical Distribution Companies: A Data Mining Based Approach," *Math. Probl. Eng.*, vol. 2014, pp. 1–15, 2014, doi: 10.1155/2014/420310.
[14] W. Dai, J.-Y. Wu, and Lu, "Applying Different Independent Component Analysis Algorithms and Support Vector Regression for IT Chain Store Sales Forecasting," *Sci. World J.*, vol. 2014, pp. 1–9, 2014, doi: 10.1155/2014/438132.
[15] C. Yu, "Design of Drug Sales Forecasting Model Using Particle Swarm Optimization Neural Networks Model," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, 2022, doi: 10.1155/2022/6836524.
[16] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J — Multidiscip. Sci. J.*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
[17] C.-N. Chen *et al.*, "Applying Simulation Optimization to Minimize Drug Inventory Costs: A Study of a Case Outpatient Pharmacy," *Healthcare*, vol. 10, no. 3, p. 556, 2022, doi: 10.3390/healthcare10030556.
[18] W. Yan-hua, Y. Liu, and W. Jing, "Hadoop-Based Parallel Algorithm for Data Mining in Remote Sensing Images," *Int. J. Performability Eng.*, vol. 15, no. 11, p. 2860, 2019, doi: 10.23940/ijpe.19.11.p4.28602870.
[19] J. Junjie, S. Wenhao, and W. Yuan, "A risk assessment approach for road collapse along tunnels based on an improved entropy weight method and K-means cluster algorithm," *Ain Shams Eng. J.*, vol. 15, no. 7, p. 102805, 2024, doi: 10.1016/j.asej.2024.102805.
[20] K. N. Prashanth Kumar, B. T. Harish Kumar, and A. Bhuvanesh, "Spectral clustering algorithm based web mining and quadratic support vector machine for learning style prediction in E-learning platform," *Meas. Sensors*, vol. 31, no. November 2023, p. 100962, 2024, doi: 10.1016/j.measen.2023.100962.
[21] X. Zheng, M. A. Abdul Hamid, and Y. Hou, "Data mining algorithm in the identification of accounting fraud by smart city information technology," *Heliyon*, vol. 10, no. 9, p. e30048, 2024, doi: 10.1016/j.heliyon.2024.e30048.
[22] S. Zhang, C. Bi, and M. Zhang, "Logistics service supply chain order allocation mixed K-Means and Qos matching," *Procedia CIRP*, vol. 188, no. 2019, pp. 121–129, 2021, doi:

10.1016/j.procs.2021.05.060.
[23] A. Usai, M. Pironti, M. Mital, and C. A. Mejri, “Knowledge Discovery Out of Text Data: A Systematic Review via Text Mining,” *J. Knowl. Manag.*, vol. 22, no. 7, pp. 1471–1488, 2018, doi: 10.1108/jkm-11-2017-0517.
[24] K. Hassani-Pak and C. J. Rawlings, “Knowledge Discovery in Biological Databases for Revealing Candidate Genes Linked to Complex Phenotypes,” *J. Integr. Bioinform.*, vol. 14, no. 1, 2017, doi: 10.1515/jib-2016-0002.
[25] A. Jahani, P. Akhavan, M. Jafari, and M. Fathian, “Conceptual Model for Knowledge Discovery Process in Databases Based on Multi-Agent System,” *Vine J. Inf. Knowl. Manag. Syst.*, vol. 46, no. 2, pp. 207–231, 2016, doi: 10.1108/vjikms-01-2015-0003.
[26] N. A. Prahastiwi, R. Andreswari, and R. Fauzi, “Students Graduation Prediction Based on Academic Data Record Using the Decision Tree Algorithm C4.5 Method,” *Jurteksi (Jurnal Teknol. Dan Sist. Informasi)*, vol. 8, no. 3, pp. 295–304, 2022, doi: 10.33330/jurteksi.v8i3.1680.
[27] J. Kovacevic, A. Kovačević, T. Miletić, J. Đuriš, and S. Ibrić, “Data Mining Techniques Applied in the Analysis of Historical Data,” *Arh. Farm. (Belgr).*, vol. 72, no. 6, pp. 701–715, 2022, doi: 10.5937/arhfarm72-41368.
[28] J. C. D. Vera, G. M. N. Ortiz, C. Molina, and M. A. Vila, “Knowledge Redundancy Approach to Reduce Size in Association Rules,” *Informatica*, vol. 44, no. 2, 2020, doi: 10.31449/inf.v44i2.2839.
[29] P. Ristoski and H. Paulheim, “Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey,” *SSRN Electron. J.*, 2016, doi: 10.2139/ssrn.3199217.
[30] J. L. Dias, M. K. Sott, C. C. Ferrão, J. C. Furtado, and J. A. R. Moraes, “Data mining and knowledge discovery in databases for urban solid waste management: A scientific literature review,” *Waste Manag. Res.*, vol. 39, no. 11, pp. 1331–1340, 2021.
[31] H. Gao, S. Gajjar, M. Kulahci, Q. Zhu, and A. Palazoglu, “Process Knowledge Discovery Using Sparse Principal Component Analysis,” *Ind. Eng. Chem. Res.*, vol. 55, no. 46, pp. 12046–12059, 2016, doi: 10.1021/acs.iecr.6b03045.
[32] F. Asrin, S. Saide, S. Ratna, and A. Wenda, “Knowledge Data Discovery (Frequent Pattern Growth): The Association Rules for Evergreen Activities on Computer Monitoring,” pp. 807–816, 2020, doi: 10.1007/978-3-030-51156-2_93.
[33] M. Relich and K. Bzdyra, “Estimating New Product Success With the Use of Intelligent Systems,” *Found. Manag.*, vol. 6, no. 2, pp. 7–20, 2014, doi: 10.1515/fman-2015-0007.
[34] S. Cheung, M. Nakamoto, and Y. Hamuro, “NYSOL: A User-Centric Framework for Knowledge Discovery in Big Data,” *Int. J. Knowl. Eng.*, vol. 1, no. 3, pp. 214–218, 2015, doi: 10.18178/ijke.2015.1.3.037.
[35] Q. Qian, Q. Zhao, Y. Chen, and Z. Jiang, “Knowledge Discovery in Evolution of the Structure and Form of Scientific Model: J-System Theory,” 2019, doi: 10.2991/mmsta-19.2019.1.
[36] Y. S. Siregar, B. O. Sembiring, H. Hasdiana, A. R. Dewi, and H. J. P. Harahap, “Algortihm C4.5 in Mapping the Admission Patterns of New Students in Engingeering Computer,” *Sinkron*, vol. 6, no. 1, pp. 80–90, 2021, doi: 10.33395/sinkron.v6i1.11154.
[37] E. N. Ekwonwune, C. I. Ubochi, and A. E. Duroha, “Data Mining as a Technique for Healthcare Approach,” *Int. J. Commun. Netw. Syst. Sci.*, vol. 15, no. 09, pp. 149–165, 2022, doi: 10.4236/ijcns.2022.159011.
[38] K. Kasikumar, M. N. M, and R. M. Suresh, “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks,” *Int. J. Data Min. Tech. Appl.*, vol. 7, no. 1, pp. 172–176, 2018, doi: 10.20894/ijdmta.102.007.001.027.
[39] G. Kaundal, “To Enhance the Security in Data Mining Using Integration of Cryptograhic and Data Mining Algorithms,” *Iosr J. Eng.*, vol. 4, no. 6, pp. 34–38, 2014, doi: 10.9790/3021-04623438.
[40] Y. T. Assegid and R. Gangarde, “Effective Pattern Discovery for Text Mining and Compare PDM and PCM,” *Int. J. Eng. Trends Technol.*, vol. 35, no. 5, pp. 189–194, 2016, doi: 10.14445/22315381/ijett-v35p242.
[41] A. Papadimitriou and V. Tsoukala, “Evaluating and enhancing the performance of the K-Means clustering algorithm for annual coastal bed evolution applications,” *Oceanologia*, vol. 66, no. 2, pp. 267–285, 2024, doi: https://doi.org/10.1016/j.oceano.2023.12.005.
[42] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, “A dynamic K-means clustering for data mining,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 2, pp. 521–526, 2019, doi: 10.11591/ijeecs.v13.i2.pp521-526.
[43] A. R. Hedar, A. M. M. Ibrahim, A. E. Abdel-Hakim, and A. A. Sewisy, “K-means cloning: Adaptive spherical K-means clustering,” *Algorithms*, vol. 11, no. 10, pp. 1–21, 2018, doi:

10.3390/a11100151.
[44] A. Naik and L. Samant, “Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime,” *Procedia Comput. Sci.*, vol. 85, pp. 662–668, 2016, doi: https://doi.org/10.1016/j.procs.2016.05.251.