Published online on the page : https://ejournal.uinbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Analysis of Drug Inventory Patterns Using the K-Means Algorithm

Dedi Setiadi[1,], Debi Gusmaliza[2]*

[1,2]Institut Teknologi Pagar Alam, Pagar Alam, Indonesia

Article Information

Article History:

Submitted: November 17, 2025
Revision: November 24, 2025
Accepted: December 3, 2025
Published: December 11, 2025

Keywords

Clustering
CRIPS-DM
Drugs
Elbow Method
K-Means

Correspondence

E-mail: dedisetiadi1212@gmail.com *

A B S T R A C T

Efficient drug inventory management is a critical challenge for the Sandar Angin Community Health Center to ensure the availability of drugs needed by customers without incurring excessive storage costs. Data mining with the K-Means algorithm was used to determine drug inventory more effectively. Drug data for the past year was used as a sample in this study. The Elbow method was used to determine the optimal number of clusters, and the results showed that three clusters were most appropriate for grouping drug sales data. The first cluster consisted of drugs with high and consistent sales, the second cluster included drugs with moderate and fluctuating sales, while the third cluster contained drugs with low and inconsistent sales. The results of this clustering provide clear guidance in drug inventory management. Drugs in the first cluster require larger stocks, the second cluster requires moderate stocks and promotional strategies tailored to the season, while the third cluster requires minimal stocks and regular evaluations to determine the continuation of its supply. The implementation of the K-Means method has proven effective in reducing storage costs, increasing customer satisfaction, and optimizing inventory management. This study concluded that data mining using the K-Means algorithm can help the Sandar Angin Community Health Center make better decisions regarding drug inventory. The results showed that out of a total of 506 drug data sets, 496 fell into cluster 0, or 98% of the data. One drug data set fell into cluster 1, and nine drug data set fell into cluster 2.

1. Introduction

Medicine is an integral component of the medical world. [1] Its function is crucial, namely to cure diseases, relieve pain, and prevent various health disorders. [2] Therefore, medication management must be carried out properly, efficiently, and effectively as part of efforts to improve the quality of healthcare services. With optimal medication management, medication availability can be ensured on time and as needed, [3] thereby minimizing risks such as stockouts or drug backlogs in healthcare facilities such as hospitals and community health centers. Furthermore, the process of distributing, purchasing, and using medication will be more efficient, which ultimately also has a positive impact on the economics of these healthcare institutions.

Therefore, careful planning in determining medication needs is essential for effective and efficient management. [4] The main goal of medication needs planning is to ensure the availability and equitable distribution of medications across all healthcare facilities, one of which can be achieved through the application of data mining technology. [5]

Data mining is the process of analyzing large data [6] sets to discover previously unknown patterns, relationships, or new information. [7] The results of this analysis can provide useful knowledge to data owners through the application of modern methods and techniques. The use of data mining plays a crucial role in various institutions and agencies [8] because it can simplify work and process large amounts of data into specific groups, thereby saving time in decision-making. [9] In the context of drug management, data mining can be used to analyze and determine drug usage patterns, which is then useful in the planning and control of drug inventories. [10]

One technique frequently used in data mining is clustering, [11] a method of grouping data into several groups or clusters based on similar characteristics between the data. [12] Some common algorithms used in this method include K-Nearest Neighbors (KNN), Fuzzy C-Means, and K-Means. Among these algorithms, K-Means is one of the most frequently used because of its ability to group data into several clusters with similar attributes, so that data with different characteristics can be arranged in a more orderly and structured manner. [13] Based on observations, the author found that the drug data processing process at the Sandar Angin Community Health Center is still not optimal, particularly in terms of planning and controlling drug stock procurement. The main problem lies in inaccuracies in analyzing drug needs, which leads to difficulties in decision-making regarding inventory planning and control. This situation results in a mismatch between the amount of drug needs and demand, potentially disrupting the smooth delivery of patient care.

To address this issue, drug data was clustered using data mining techniques using the K-Means algorithm. [14] This approach categorizes drug data based on their usage levels: low, medium, and high. The resulting clustering results are expected to serve as a reference for planning and controlling future drug needs, ensuring greater efficiency and targeted targeting. [15] The K-Means Clustering method is a technique for grouping large amounts of data [16] based on the similarity of attributes shared by each data item. [17] This research is motivated by problems in drug inventory management, such as stockpiling and shortages. Therefore, this study aims to group types of drugs into three clusters: C1 (low), C2 (medium), and C3 (high). The results of this clustering process are expected to be used as a basis for planning future drug needs to be more effective and efficient. [18] Thus, drug management can be optimized through appropriate stock management, adjustments in the ordering process, and cost control, thereby improving the overall quality of healthcare services.

2. Method

Research methodology is essentially the operational application and application of scientific methods in the search for truth. CRISP-DM (Cross Industry Standard Process for Data Mining) is a standard process used in data mining activities. [19] This standard is designed to ensure that each stage of data processing is structured, clearly defined, and efficient.[20] In addition to serving as a model for the data mining process, selecting the right algorithm significantly impacts the comparison and performance of the data mining methods used. CRISP-DM is widely applied as a problem-solving framework in both business and research contexts. [21] This methodology consists of six main stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The following is a model illustration of the CRISP-DM process stages. [22]

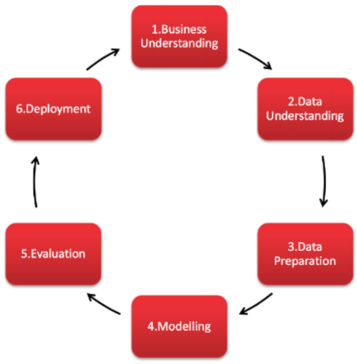

Figure 1. Research methodology

2.1. Business Understanding

This research focuses on determining the objectives of using the K-Means algorithm in the clustering process and analyzing detailed needs within the scope of the business unit or research object as a whole. Furthermore, this research also explains the objectives and limitations of the data used, taken from the Pagar Alam City Community Health Center, to serve as a basis for formulating data mining problems. This process includes stages ranging from developing an initial strategy to determining the methods needed to achieve the research objectives. [23] The initial strategy implemented in this research began with a request for drug data from the Sandar Angin Community Health Center in Pagar Alam City.

2.2. Data Understanding

The data used in this study were obtained from the Sandar Angin Community Health Center in Pagar Alam City, consisting of Drug Usage Reports and Drug Request Sheets for 2024. After the data was obtained, an exploration process was conducted to understand the structure and characteristics of the data. [24] The collected data comprised a summary of drug usage reports and requests, totaling 506 entries. The data was then processed using a predetermined algorithm to obtain the best clustering results.

2.3. Data Preparation

After completing the data comprehension phase, the next step is data processing, which aims to create the final dataset for use in the modeling phase. This phase involves compiling the final dataset based on the previously collected raw data. [25] Several steps include data cleansing, where the data is checked and cleared of errors or inconsistencies. Next, data selection is performed, identifying relevant records and attributes for analysis. Data transformation is then performed according to predetermined criteria. The assessment criteria used in this phase are as follows :

Table 1. Dataset

No	Drug Name	Unit	Initial stock	Inventory	Reception	Usage
1.	Anti-Snake Venom	Vial	0	0	1	0
2.	Acyclovir 400 mg table	Tablet	500	1.000	1.000	500
3.	Acyclovir salep	Tube	100	50	50	23
4.	Disposable syringe (1 ml)	Pcs	515	200	200	0
5.	Disposable syringe (3 ml)	Pcs	1.000	1720	2.000	720
6.	Disposable syringe (5 ml)	Pcs	510	500	500	0
7.	Alcohol (etanol) 70 % 1 liter	bottle	13	10	10	10
8.	Alcohol swab	pcs	16.000	10.000	10.000	0

Table 2. Unit criteria

Unit	Information
Vial	1
Tablet	2
Tube	3
Pcs	4
bottle	5

Table 3. Initial stock criteria

Initial stock	Information
0-50	1
60-100	2
110-500	3
510-1.000	4
10.000-20.000	5

Table 4. Inventory criteria

Supply	Information
0-50	1
100-500	2
550-1.000	3
1.000-5.000	4
6.000-10.000	5

Table 5. Acceptance criteria

Reception	Information
0-50	1
100-500	2
550-1.000	3
1.000-5.000	4
6.000-10.000	5

Table 6. Usage criteria

Usage	Information
0-50	1
100-500	2
550-1.000	3

2.4. Modelling

At this stage, the research process begins to involve the application of machine learning to determine the techniques, tools, and algorithms used in data mining. [26] The modeling applied in this study uses the clustering method with the K-Means algorithm. The aim of this stage is to group drug data based on their usage levels using the 2024 dataset, so that drug inventory patterns at the Community Health Center can be identified. The steps for applying the K-Means algorithm in this modeling process are as follows:

2.4.1. Prepare the Dataset

The dataset comes from data that has been archived and stored in the form of a Microsoft Excel file, then the data will be used as a dataset, the attributes used in K-Means Clustering are Drug Name, Unit, Initial Stock, Inventory, Receipt of Use.

2.4.2. Determine the Number of Clusters

Based on the prepared dataset, the clustering process is carried out by forming three groups, namely high, medium, and low categories.

2.4.3. Determine the centroid randomly

Because in this study three groups (clusters) were determined, three centroids were required as cluster center points which were determined randomly in the initial stage of the clustering process.

2.4.4. Calculate the distance of the data to the centroid

To calculate the distance to the centroid, the Euclidean Distance formula is used, which is one of the most common distance measurement methods used in mathematics, statistics, and computer science, especially in machine learning algorithms, [27] with the formula :

$$d(x,y) = \sum_{k=1}^{n}(x_i - y_i)^2 \quad (1)$$

Information :

- d(x,y = distance between data x and data y
- xi = i-th testing data
- yi = i-th training data

In this calculation stage, the initial cluster center value is determined for each cluster based on each variable used. In the first iteration, the initial cluster center value (initial centroid) is determined randomly. Then, in subsequent iterations, the cluster center value is updated by calculating the average of the data included in each cluster. The iteration process will continue until the new centroid value generated no longer changes or has reached a predetermined maximum number of iterations. If the new centroid value is the same as the centroid value from the previous iteration, the clustering process is considered to have converged and the calculation is stopped.

2.4.5. Update Centroid Point

The calculation of new centroid points is carried out based on the grouping results in the first iteration to be used in the next iteration process.

2.4.6. Repeat steps 3 to 5 until the value of the centroid point no longer changes

The calculation is performed continuously until all predetermined data has been processed. After the distance between each data point and the centroid is calculated, the next step is to group the data based on the closest distance to each centroid. Next, an iteration is performed to compare the grouping results. If the centroid position changes, the calculation process is repeated until the centroid value stabilizes or no longer changes.

2.5. Evaluation

At this stage, the main focus is to ensure that the resulting model complies with the K-Means Clustering method standards and that no steps are missed from the initial process to the final modeling stage. Evaluation is carried out using the Elbow Method to test the accuracy and effectiveness of the clustering results. The Elbow Method is a technique used to determine the most optimal number of clusters by analyzing the percentage comparison between the number of clusters and the Sum of Squared Error (SSE) value. [28] The relationship between the value of k (the number of clusters) and SSE is inversely proportional—the larger the value of k, the smaller the SSE value will be until it reaches the optimal point that forms a right angle, which indicates the best number of clusters to use. [29]

$$SSE = \sum_{k=1}^{k}\sum_{xi}[Xi - Ck]^2 \quad (2)$$

Information :

K = cluster c

Xi = distance of object i data

Ck = pusat cluster ke-i

2.6. Deployment

In the final stage of the CRISP-DM process,[30] the knowledge or information obtained from the analysis is disseminated. This information is then implemented in a report, and each resulting cluster is analyzed. [31] This step aims to ensure that the research results are easily understood and utilized by the Sandar Angin Community Health Center in Pagar Alam City as a basis for planning future drug needs.

3. Results and Discussion

3.1. Result

This research resulted in the application of data mining using the K-Means method to determine drug inventory levels at the Sandar Angin Community Health Center. The data analyzed were drug sales data for the past year. The application of the K-Means algorithm resulted in three main clusters that represent data grouping patterns based on the level of similarity of their characteristics. Each cluster describes a group of data with certain characteristics, such as high, medium, and low levels of drug demand. The formation of these three clusters aims to facilitate the analysis process in identifying trends in drug needs at the Sandar Angin Community Health Center, so that inventory management strategies can be designed more efficiently and based on data (data-driven decision-making). And the results of the clustering process show a clearly identified sales distribution pattern.

Cluster 0 : This group includes drugs with high sales and consistent demand across the months. Most drugs in this cluster are used to treat chronic conditions, such as hypertension and diabetes, which require regular patient consumption. Drugs in this category exhibit high and stable demand, so their availability must be maintained in sufficient quantities to prevent shortages.

Cluster 1 : This cluster consists of drugs with moderate sales but exhibiting fluctuating demand patterns. Most drugs in this group are cold and cough medications, which typically experience increased sales during certain periods or seasons, such as during seasonal changes or unpredictable weather conditions. This fluctuating sales pattern necessitates moderate stock management and the implementation of appropriate promotional strategies during specific periods.

Cluster 2 : This cluster includes drugs with low sales and unstable distribution patterns. Most drugs in this group are rarely needed by patients or are new drugs with low public awareness. They have low and inconsistent demand, necessitating limited stock availability and regular evaluation to assess their continued availability.

Tabel 7. Caster results

No	Drug Name	0	1	2	Cluster
1.	name Water for injection amp 20 ml	0,001557	0	0	0
2.	name Water for injection amp 25 ml	0,001557	0	0	0
3.	name of 1 ml syringe	0,003114	0	0	0
4.	name of 3 ml syringe	0,00327	0	0	0
5.	name of 5 ml syringe	0,00327	0	0	0
6.	name of 0,5 ml syringe	0,00327	0	0	0
7.	name of 10 ml syringe	0,003114	0	0	0
8.	name Albendazole suspension 200 mg/5 ml	0,001713	0	0	0
…	…	…	…	…	…
506	name Albendazole suspension 400 mg	0,001557	0	0	0

3.2. Discussion

3.2.1. Preprocessing Data

Before clustering, the data undergoes several preprocessing stages, including :

1. Cleaning Data : Eliminating data that is incomplete or has invalid values.
2. Data Normalization : Using the min-max normalization method to normalize the number of sales to be in the range of 0 to 1.
3. Data Transformation : Converting categorical data into numeric data using one-hot encoding technique.

This process continues until the data distribution appears normal, indicating there are no outliers. The data can be processed. The next step is to display the data distribution.

Program
```
plt.figure(figsize=(15, 5))
plt.subplot(1, 3, 1)
plt.hist(df['persediaan'], bins=15, color='skyblue', edgecolor='black')
plt.title('Persediaan')
plt.xlabel('Value')
plt.ylabel('Count')
attrition_counts = df['sisa stok'].value_counts()
plt.subplot(1, 3, 2)
plt.bar(attrition_counts.index, attrition_counts.values, color=['lightgreen', 'lightcoral'])
plt.title('Sisa Stok')
plt.xlabel('Value')
plt.ylabel('Count')
bulan_counts = df['bulan'].value_counts()
plt.subplot(1, 3, 3)
plt.bar(bulan_counts.index, bulan_counts.values, color=['lightgreen', 'lightcoral'])
plt.title('Bulan')
plt.xlabel('Value')
plt.ylabel('Count')
plt.tight_layout()
plt.show()
```

The visualization results of the data distribution generated from the program code above are presented in Figure 2.

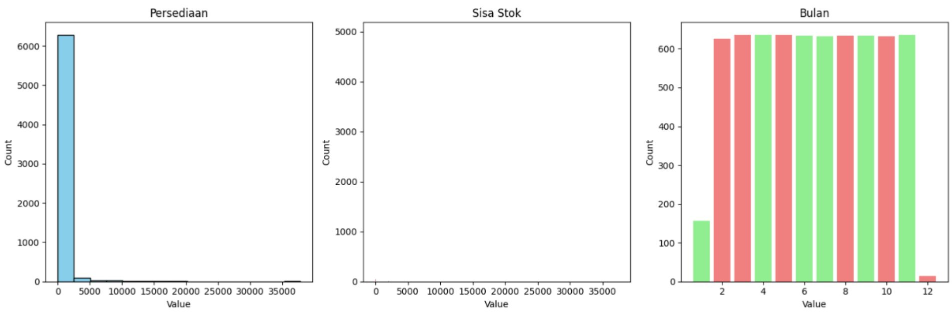

Figure 2. Data distribution

From the distribution, it can be seen that there are several things that can be analyzed, most of the stock amounts are in the range of 0 - 2500, there is no remaining stock and there are more drugs in months 2 - 11.

Program
```
sisa_yes = df[df['sisa stok'] == 1]['bulan']
sisa_no = df[df['sisa stok'] == 0]['bulan']
plt.figure(figsize=(10, 6))
plt.hist(sisa_yes, bins=15, color='red', alpha=0.5, label='Sisa Yes')
plt.hist(sisa_no, bins=15, color='green', alpha=0.5, label='Sisa No')
plt.title('Hubungan antara Bulan dan Sisa Stok')
plt.xlabel('Bulan')
plt.ylabel('Sisa Stok')
plt.legend()
plt.grid(True)
plt.show()
```

The code above is used to see the relationship between months and remaining stock, and the visualization results can be seen in Figure 3. As follows

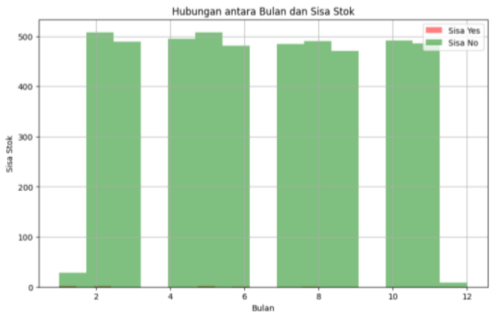

Figure 3. the relationship between months and remaining stock

From the graph above, it can be seen that, in each month the remaining stock always runs out, and there is very little left in months 1, 2, 5 and 8.

3.2.2. Evaluasi

In the evaluation stage, elbow testing is used with the code as seen in the code below :

Program

```
plt.figure(figsize=(12,4))

plt.subplot(121)
plt.plot(list_cluster,list_innertia,'r-o')
plt.xlabel('Jumlah Cluster')
plt.ylabel('Innertia')
plt.grid()

plt.subplot(122)
plt.plot(list_cluster,list_silhouette,'r-o')
plt.xlabel('Jumlah Cluster')
plt.ylabel('Silhouette')
plt.grid()
```

The results of the elbow method testing obtained through the program code above are then visualized in Figure 4 below. This visualization aims to show the relationship between the number of clusters and the within-cluster sum of squares (WCSS) value, thus making it easier for researchers to determine the optimal number of clusters. With this graphical representation, the process of determining the most efficient k can be carried out more systematically and based on empirical evidence.

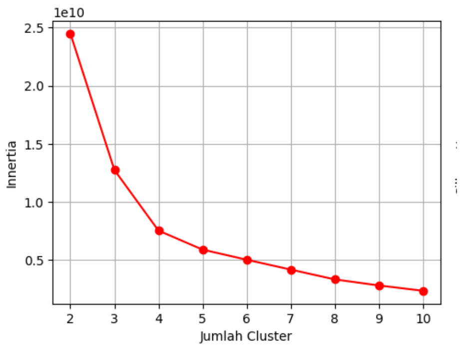

Figure 4. Elbow testing

From the graph above, it can be seen that the inertia value is good. It can be seen in the red circle that the inertia value has decreased drastically on the 3rd line, which shows that the optimal number of clusters is 3.

3.2.3. Deployment

Program

```
df_percentage = baseline.to_frame()
   for i in range(k):
        df_percentage[i] = df_cluster[i]/baseline

df_percentage
```

The code is used in the cluster analysis process to produce proportional values that facilitate interpretation and can then be used in visualization, model evaluation, the results of which are shown in Figure 5.

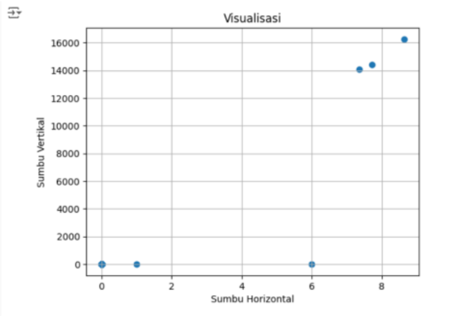

Figure 5. Data visualization

4. Conclusion

Based on the evaluation and modeling results, three clusters were formed to classify the data based on monthly periods. Clusters 1 and 2 demonstrate a relatively low risk of drug stockouts, based on the remaining stock and its distribution each month. Meanwhile, cluster 0 requires special attention due to its significantly higher risk of stockouts compared to the other clusters.

References

[1] D. Fitriyani, M. Jajuli, and G. Garno, "Implementasi Algoritma K-Means Untuk Klasterisasi Dalam Pengelolaan Persediaan Obat (Studi Kasus: Apotek Naza)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024.

[2] S. Farzin, F. N. Chianeh, M. V. Anaraki, and F. Mahmoudian, "Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID)," *J. Clean. Prod.*, vol. 266, p. 122075, 2020.

[3] R. D. Parinduri, S. Defit, and G. W. Nurcahyo, "Implementasi Algoritma Apriori dalam Data Mining untuk Optimalisasi Stok Obat di Apotik," *J. KomtekInfo*, pp. 89–97, 2024.

[4] R. M. Febrialdo and Y. Wendra, "Penerapan Data Mining Memprediksi Penjualan Obat Menggunakan Metode K-Nearest Neighbor (Studi Kasus: Apotek Difana)," *J. Sains Inform. Terap.*, vol. 4, no. 3, pp. 526–533, 2025.

[5] R. A. N. Rolan, "Penerapan Data Mining Pengelompokan Penjualan Obatdengan Algoritma K-Means Clustering Pada Toko Obat Sanjaya," *J. Data Sci. Penusa*, vol. 1, no. 1, pp. 242–247, 2024.

[6] M. I. Habibi, A. Nazir, E. Haerani, and E. Budianita, "Application of Data Mining for Ceramic Sales Data Association Using Apriori Algorithm," *Knowbase Int. J. Knowl. Database*, vol. 4, no. 2, pp. 105–114, 2024.

[7] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, 2024.

[8] D. Puspita, D. Setiadi, and R. Wulandari, "Implementation of the Linear Discriminant Analysis (LDA) Method in Potato Fruit Classification," *Jusikom J. Sist. Komput. Musirawas*, vol. 9, no. 2, pp. 89–97, 2024.

[9] P. W. Rahayu, I. N. Bernadus, and A. I. Datya, "Penerapan Data Mining Dalam Mengetahui Pola Transaksi Pembelian Obat Menggunakan Algoritma Apriori Di Apotek Kharisma Farma Tiga," *J-Icon J. Komput. Dan Inform.*, vol. 12, no. 1, pp. 44–55, 2024.

[10] B. T. Admaja, "Data Mining Using K-Means Algorithm for Clustering Snack Sales at CV Sinar Pangan Utama," *J. Artif. Intell. Eng. Appl.*, vol. 5, no. 1, pp. 946–953, 2025.

[11] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024.

[12] I. Rinaldi and H. Harmayani, "Penggunaan algoritma clustering K-Means untuk mengelompokkan pola cuaca," *J. Sci. Soc. Res.*, vol. 8, no. 1, pp. 343–348, 2025.

[13] F. Khalish, N. M. Piranti, and O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," *JIIP-Jurnal Ilm. Ilmu Pendidik.*, vol. 8, no. 5, pp. 5392–5397, 2025.

[14] M. R. Nur, A. Ranggareksa, F. Adiba, and A. Husna, "Clustering The Use Of Puskesmas Soreang Medicine Using The Fuzzy C-Means Method," *J. Heal. Innov. Technol.*, vol. 1, no. 1, pp. 1–8, 2025.

[15] R. S. Qosimah, G. W. Sasmito, and D. Apriliani, "Drug Data Management Application at Telu Tegal Pharmacy Using Website-Based K-Means Algorithm," *J. Appl. Informatics Sci.*, vol. 1, no. 1, pp. 18–23, 2025.

[16] M. A. vafky Ideal and I. Fitriyanto, "Implementation of a K-Means-Based Intelligent Patient Complaint Clustering System to Identify Handling Priorities," *Knowbase Int. J. Knowl. Database*, vol. 5, no. 1, pp. 69–80, 2025.

[17] M. A. Septianto, A. Faqih, and A. R. Rinaldi, "Klasterisasi data produksi pertanian di Kabupaten Cirebon dengan algoritma K-Means," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025.

[18] L. N. Azizah, L. Utari, L. T. Ningrum, and D. Rahmiyati, "Penerapan Algoritma K-Means untuk Pemetaan Kebutuhan Obat Sesuai Prioritas dalam Pengadaan," *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 6, no. 1, pp. 162–169, 2025.

[19] D. Setiadi, S. Sasmita, and M. Yolanda, "Penerapan Algoritma Regresi Linier Berganda Untuk Memprediksi Hasil panen Padi Di Kota Pagar Alam," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 5, no. 2, pp. 337–438, 2024.

[20] D. Setiadi, "Analisis Prediksi Harga Beras Berbasis Kualitas Menggunakan Algoritma K-Nearest Neighbord," *J. Ilm. Inform. Glob.*, vol. 15, no. 3, pp. 106–115, 2024.

[21] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021.

[22] D. Setiadi, S. Sasmita, and Y. I. Mukti, "Optimization Of Agricultural Production In South Sumatera Using Multiple Linear Regression Algorithm," *Knowbase Int. J. Knowl. Database*, vol. 4, no. 2, pp. 168–179, 2024.

[23] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for data science: Methods, processes and frameworks," *SBC Rev. Comput. Sci.*, vol. 4, no. 1, pp. 28–43, 2024.

[24] G. Wulandari, S. Sasmita, and F. Rahmadayanti, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Obat," *J. Tek. Inf. dan Komput.*, vol. 7, no. 1, pp. 144–150, 2024.

[25] F. D. Agustiar, B. N. Sari, and I. Maulana, "Penerapan Data Mining Untuk Pengelompokan Produk Penjualan Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 58–67, 2025.

[26] D. B. Saputra, V. Atina, and F. E. Nastiti, "Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit Menggunakan Algoritma Random Forest," *IDEALIS Indones. J. Inf. Syst.*, vol.

7, no. 2, pp. 240–247, 2024.

[27] M. Fauzie, "Klasifikasi Pengadaan Obat Jaminan Kesehatan Nasional Melalui E-Purchasing Menggunakan Algoritma K-Nearest Neighbor," *J. Komputer, Inf. dan Teknol.*, vol. 4, no. 2, p. 10, 2024.

[28] M. A. F. Razan, N. J. Alifah, Q. A'yuni, and M. Wati, "Application of K-Means Clustering Algorithm for Air Quality Pattern Analysis in Jakarta," *J. Teknol. DAN ILMU Komput. PRIMA*, vol. 8, no. 1, pp. 64–80, 2025.

[29] C. A. Fransiska and D. Dafid, "Application of K-Means Clustering Algorithm for Disease Grouping at Blessing Dental Care Clinic," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 3, pp. 1119–1132, 2025.

[30] D. Setiadi, A. Arif, and A. Oktaria, "Comparison of K-Nearest Neighbor and Naive Bayes Algorithms for Tuberculosis Diagnosis Classification," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 1, pp. 176–187, 2025.

[31] A. Rahmawati, A. Hananto, and F. Nurapriani, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Untuk Optimalisasi Penjualan Sembako," *J. Ilm. Ilk. Komput. Inform.*, vol. 8, no. 2, pp. 233–243, 2025.