OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
journals.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 11, Issue. 2, August 2026. pp. 98-106
doi:10.21108/indojc.v11i2.11743

Drug Expiration Data Clustering Using the K-Means Clustering Algorithm to Determine Drug Sales Priorities (Case Study: Tanjung Sari Farma Pharmacy)

Maria Shusanti Febrianti [#1], Marzuki Syahfirin [#2], Stephany Meiyomi Br Hombing [#3]

[#] *Informatics Study Program, Faculty of Computer Science, University of Bandar Lampung*
Bandar Lampung, Indonesia
[1] maria.shusanti@ubl.ac.id
[2] marzuki@ubl.ac.id
[3] stephany.22421053@student.ubl.ac.id

Abstract

Manual monitoring of drug expiration dates still exposes community pharmacies to considerable operational risk, since staff must check each record individually before delays become financial or clinical hazards. Tanjung Sari Farma Pharmacy currently records drug data in Microsoft Excel without grouping medicines by proximity to expiration. This applied study develops a web-based monitoring feature that applies the K-Means clustering algorithm to sort drug records according to the number of days remaining before expiration, a variable referred to as Expired Day (ED). The research follows the Knowledge Discovery in Database (KDD) stages, namely data selection, preprocessing, transformation, mining, and evaluation, and the resulting feature is embedded in a system built with the Laravel framework and a MySQL database. Out of 1,782 processed drug records and three manually determined initial centroids, the clustering process converged after five iterations and produced three categories, namely Critical, Alert, and Safe. Black-box and white-box testing confirmed that every system function performed as expected, with cyclomatic-complexity scores indicating a low-to-moderate program structure. The clustering output supports sales prioritization based on the First Expired First Out (FEFO) principle, allowing the pharmacy to act on near-expiry stock faster than before.

Keywords: expired drug, FEFO, K-Means clustering, pharmacy, web-based system

Abstrak

Pengelolaan manual tanggal kedaluwarsa obat berpotensi menimbulkan risiko operasional pada apotek dengan volume stok besar, sebab petugas harus memeriksa data kedaluwarsa satu per satu sebelum keterlambatan tersebut berubah menjadi kerugian finansial maupun ancaman keselamatan pasien. Kondisi tersebut dijumpai di Apotek Tanjung Sari Farma, yang hingga kini mencatat data obat melalui Microsoft Excel tanpa fitur pengelompokan berdasarkan kedekatan masa kedaluwarsa. Penelitian terapan ini mengembangkan fitur monitoring berbasis web yang menerapkan algoritma K-Means untuk mengelompokkan data obat berdasarkan variabel jumlah hari menuju kedaluwarsa, disebut Expired Day (ED). Penelitian mengikuti tahapan Knowledge Discovery in Database (KDD), yaitu seleksi data, praproses, transformasi, penambangan data, dan evaluasi, kemudian fitur tersebut ditanamkan pada sistem yang dibangun menggunakan framework Laravel dan basis data MySQL. Dari 1.782 data obat yang diolah dengan tiga centroid awal yang ditentukan secara manual, proses klastering konvergen pada iterasi kelima dan menghasilkan tiga kategori, yaitu Kritis, Waspada, dan Aman. Pengujian Black Box dan White Box membuktikan seluruh fungsi sistem berjalan sesuai harapan, dengan nilai cyclomatic complexity yang menunjukkan struktur program tergolong rendah hingga sedang. Hasil klastering ini mendukung penentuan prioritas penjualan berdasarkan prinsip First Expired First Out (FEFO) sehingga apotek dapat menindaklanjuti stok yang mendekati kedaluwarsa lebih cepat dibandingkan cara sebelumnya.

Kata Kunci: apotek, First Expired First Out, K-Means clustering, masa kedaluwarsa obat, sistem berbasis web

Received on January, 2026. Accepted on June, 2026

I. INTRODUCTION

DRUG quality, safety, and effectiveness must be maintained from receipt through use by patients, and expiration date is one of the main determinants. Drugs that have passed this date tend to experience a decline in active-ingredient content, reducing therapeutic efficacy and potentially endangering users. Therefore, monitoring expiration dates is an important part of inventory management at every pharmaceutical facility, including pharmacies.

Tanjung Sari Farma Pharmacy is a medicine service provider serving the needs of its surrounding community. Initial observations and interviews showed that the pharmacy still records drug data in Microsoft Excel, including name, category, and expiration date. This recording medium, however, has no dedicated mechanism for monitoring how close each item is to expiration, so checks must be performed by reviewing records one by one.

This workflow becomes less efficient as the volume of managed drug data increases. The more items that must be checked, the greater the chance of human error, such as late identification of drugs nearing expiration or items being missed entirely. This situation demonstrates the need for a system that can make monitoring faster, more accurate, and more effective, including through web-based information technology combined with data analysis.

One technique relevant to this need is K-Means Clustering, a data-mining method that groups a set of data according to attribute similarity. In this study, it is applied to divide drug data into three groups, namely Critical, Alert, and Safe, using one main variable: the number of days remaining until expiration, referred to as Expired Day (ED). The developed web-based system is expected to present a more informative view of drug inventory, simplify monitoring, and support sales prioritization according to First Expired First Out (FEFO).

Accordingly, this study aims to: (1) design and build a web-based system to monitor drug expiration dates at Tanjung Sari Farma Pharmacy; (2) apply K-Means to group drug data by proximity to expiration; and (3) produce clustering information that can guide drug sales priorities. The remainder of this article is organized as follows: Section II presents the literature review; Section III describes the research method; Section IV presents results and discussion; and Section V presents the conclusion.

II. LITERATURE REVIEW

A. Previous Studies

Haerani et al. applied distance-based K-Means to 782 medicine/medical-device records at Selasih Regional Hospital Pharmacy to map daily usage patterns. Evaluation using the Davies-Bouldin Index identified two clusters as the most optimal scheme, distinguishing high- and low-demand medicines, and showed that Python-based cluster visualization can objectively identify medicine turnover dynamics [1].

Prasetyo et al. grouped ten types of medicine at Bhakti Asih Clinic based on stock, usage frequency, and investment value using Euclidean Distance with K=3 selected through the Elbow Method. The result divided medicines into fast-moving, moderate-moving, and slow-moving groups, which were considered helpful for procurement and inventory-storage decisions [2].

Fitriyani et al. adapted the Knowledge Discovery in Database framework to cluster Naza Pharmacy stock data over six months of operation using Google Colab and Python. The optimal cluster count was set to K=3 using the Elbow Method, with a Silhouette Coefficient score of 0.520 indicating a good and consistent cluster structure [3].

Pangestu et al. applied non-hierarchical K-Means to classify psychotropic and narcotic drug data at Sawah Besar District Community Health Center. Through three manual iterations validated with RapidMiner, the data were divided into three clusters with no membership changes in the final iteration, demonstrating the algorithm's effectiveness in mapping medicine-use intensity [4].

Syahputra and Yasir applied a quantitative K-Means approach to group expired drugs by similarity in their validity-period patterns using Euclidean Distance. They concluded that data mining can identify drugs with short remaining active periods more quickly, reducing the risk of expired-drug use [5].

Unlike these studies, which generally emphasize stock volume, sales frequency, or drug category, this study uses one numerical variable, the remaining days until expiration, as the clustering basis and integrates it directly into a web-based sales information system with FEFO-based priority recommendations.

B. Data Mining and Clustering

Data mining is the process of extracting valuable patterns, relationships, or information from large datasets to support decision-making [6]. It generally involves data cleaning, integration, selection, transformation, pattern mining, pattern evaluation, and knowledge presentation. Clustering is a data-mining technique that divides objects into groups based on attribute similarity, aiming to maximize differences between groups while minimizing differences within each group [7].

C. K-Means Algorithm

K-Means is a non-hierarchical clustering algorithm that groups data according to their proximity to cluster centers, or centroids [8]. The grouping process is iterative: in each round, the distance from each data point to every centroid is calculated, the data point is assigned to the cluster with the smallest distance, and the centroid is updated based on the mean of its members. Iteration stops when cluster membership no longer changes, indicating convergence. The steps are: (1) set the number of clusters (K); (2) determine initial centroids; (3) calculate each data point's distance to the centroids and assign it to the nearest cluster; (4) update centroid values based on the mean of cluster members; and (5) repeat steps (3)–(4) until the centroids stabilize.

Because this study uses only one variable, Expired Day, proximity between data points is measured using one-dimensional Euclidean Distance [9], expressed as:

$$d(x,c) = \sqrt{(x-c)^2} \tag{1}$$

where d(x,c) is the distance from a data point to a centroid, x is the Expired Day value of a data point, and c is the centroid value. The centroid is updated in each iteration using:

$$C_j = \frac{\sum_i x_i}{n} \tag{2}$$

where C_j is the new centroid of cluster j, $\sum_i x_i$ is the sum of the Expired Day values of all members of that cluster, and n is the number of members.

D. First Expired First Out (FEFO)

FEFO is an inventory-management principle that prioritizes the sale or use of products with the shortest remaining time before expiration, regardless of the order in which goods arrived at the warehouse [10]. Applying this principle helps maintain pharmaceutical product quality and minimize losses from medicines that remain unsold past expiration. In this study, K-Means clustering results provide the basis for FEFO recommendations, so medicines in the cluster with the fewest remaining days automatically receive the highest sales priority.

III. Research Method

A. Research Methodology

This applied research uses a data-mining approach to analyze drug data based on expiration dates. Data processing follows the Knowledge Discovery in Database (KDD) framework, comprising data selection, preprocessing, transformation, data mining (K-Means Clustering), and evaluation. The resulting clustering feature is then integrated into the Tanjung Sari Farma Pharmacy sales information system, built using the Laravel framework and MySQL database, so that grouping results can be used directly as FEFO-based sales-priority recommendations.

B. Data Collection Techniques

Research data and information were collected using three techniques. First, direct observation of medicine receipt, storage, stock management, and sales at Tanjung Sari Farma Pharmacy confirmed that checks for medicines nearing expiration were still performed manually. Second, interviews with the pharmacy owner, administrator, and stock-management staff explored system requirements; the results confirmed that expiration date is the most important information for determining sales priority and that automatic grouping would greatly help monitoring. Third, documentation consisted of medicine-receipt data for June–August in Microsoft Excel format; selection produced a dataset of 1,782 drug records as the main research data, with attributes for drug name, stock, expiration date, and price.

C. Clustering Stages

Preprocessing checked for missing data, duplicates, and invalid formats, while transformation converted expiration dates into the numerical Expired Day (ED) value, defined as the difference between the analysis date and each medicine's expiration date. ED is the sole variable used in K-Means clustering in this study.

The number of clusters was set to K=3, representing Critical (Cluster 1), Alert (Cluster 2), and Safe (Cluster 3). Initial centroids were manually determined based on the distribution of ED values: 265, 540, and 968 days for C1, C2, and C3, respectively. The distance from each data point to all three centroids was then calculated using Equation (1), each point was assigned to the cluster with the smallest distance, and centroids were updated using Equation (2). This process was repeated until centroid values no longer changed between iterations, indicating convergence.

D. System Design

The system was designed using UML diagrams, including a use-case diagram to model the pharmacy owner's interaction with the system; flowchart and activity diagrams to show the login, dashboard, drug-data management, and K-Means clustering processes; and a sequence diagram to represent interactions among the interface, controller, database, and clustering module. The main functional requirements include user authentication; drug-data management (add, edit, delete, search, and import from Excel); calculating the days remaining until expiration; running the K-Means algorithm; and presenting clustering results as the basis for sales-priority recommendations. Non-functional requirements include hardware at least equivalent to an Intel Core i3 with 4 GB RAM, and supporting software comprising Visual Studio Code, XAMPP, MySQL, Laravel, and a web browser.

IV. RESULTS AND DISCUSSION

A. Dataset and Preprocessing

The final dataset comprised 1,782 drug records selected from Tanjung Sari Farma Pharmacy's goods-receipt archive. Each record was transformed into an Expired Day value, the number of days remaining until the medicine's expiration date as of the analysis date. Table I presents a sample of the transformed results.

TABLE I
SAMPLE OF DATA TRANSFORMATION RESULTS

No.	Drug Name	Expiration Date	Expired Day
1	Abun DL	01/12/2029	1262
2	Acifa 400	01/03/2029	987
3	Acifar Cream	01/05/2027	317
4	Acital Chewable	01/10/2027	470
...	...	...	...
1782	Zoralin Tab 200 mg	01/01/2029	928

B. Initial Centroid Determination

Initial centroid values were manually determined based on the distribution of Expired Day values in the dataset, as presented in Table II.

TABLE II
Initial Centroids

Cluster	Initial Centroid Value (days)
C1	265
C2	540
C3	968

C. Distance Calculation and Iteration

As an illustration, Abun DL with an ED value of 1262 days has distances of 997 from C1, 722 from C2, and 294 from C3, and is therefore assigned to Cluster 3 because that is the smallest distance. Conversely, Acifar Cream with ED 317 days has distances of 52 from C1, 223 from C2, and 651 from C3, and is assigned to Cluster 1. The same calculations were performed for all 1,782 records to obtain the initial grouping, which was then used to calculate new centroids using Equation (2).

In the first iteration, Cluster 1 contained 466 records with a total ED of 130,834 (mean 280.76); Cluster 2 contained 543 records with a total ED of 320,338 (mean 589.94); and Cluster 3 contained 773 records with a total ED of 799,477 (mean 1,034.25). The distance-assignment-update process was repeated until centroid values stabilized, as summarized in Table III.

TABLE III
Centroid Value Changes per Iteration

Iteration	C1	C2	C3
1	265.00	540.00	968.00
2	280.75	589.94	1,034.25
3	286.04	602.23	1,039.10
4	292.71	609.14	1,039.10
5	292.71	609.14	1,039.10

The centroid values in the fifth iteration did not differ from those in the fourth iteration, so clustering was considered converged and stopped at that iteration. Convergence in only five rounds indicates that the three initial centroids were sufficiently representative of the data distribution, given the relatively separated Expired Day ranges among medicines expiring soon, later, and much later.

D. Clustering Result Categories

Based on the final centroid positions, Cluster 1 (centroid $\approx$ 293 days) was designated Critical because it contains medicines with the shortest remaining expiration period and requires the highest sales priority. Cluster 2 (centroid $\approx$ 609 days) was designated Alert and should be monitored regularly, while Cluster 3 (centroid $\approx$ 1039 days) was designated Safe because its remaining validity period is relatively long. This division is consistent with the pharmacy policy of giving special attention to medicines with less than six months remaining before expiration, and also forms the basis for FEFO-based sales recommendations in the system.

E. System Implementation and Testing

The clustering results above were implemented in a web-based system built using Laravel and MySQL. The following figures show the main system interfaces and their respective functions.

Users, namely the pharmacy owner or staff, access the system through a login page by entering a registered username and password. The system authenticates users before allowing them to proceed, as shown in Fig. 1.

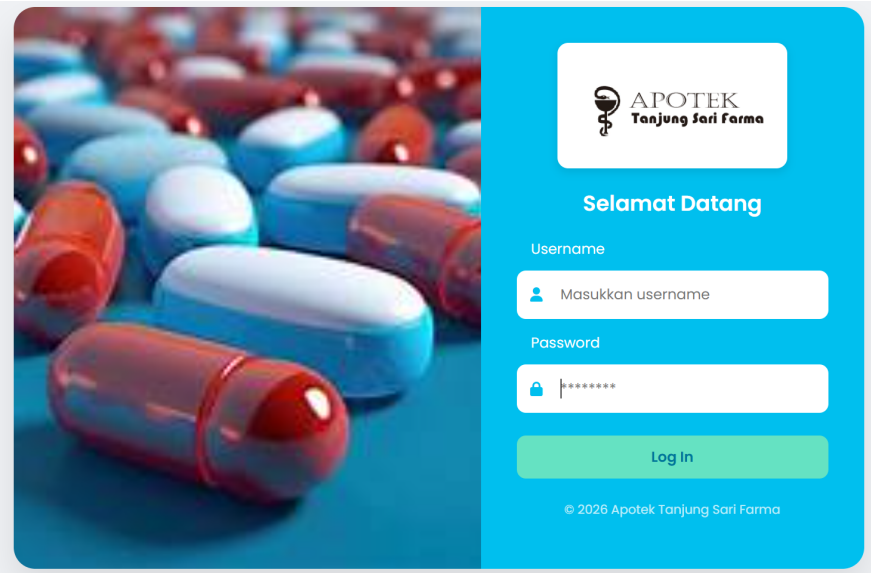

Fig. 1. The login page

After successful authentication, the system displays a dashboard serving as an information center. It summarizes the number of drug records in each cluster—Cluster 1 (Critical), Cluster 2 (Alert), and Cluster 3 (Safe)—and the total number of drug records stored, as shown in Fig. 2.

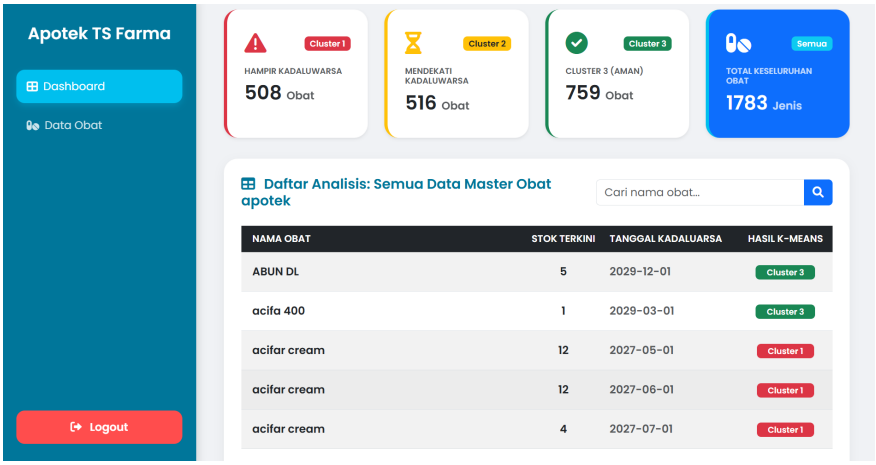

Fig. 2. The dashboard

The Drug Data page displays all medicines stored in the database with their cluster status, and provides search by medicine name and filtering by cluster to accelerate monitoring, as shown in Fig. 3.

To add new inventory, users complete the Add New Drug Data form, which includes distributor, drug name, category, initial stock, cost price, expiration date, and batch number. Data saved through this page, shown in Fig. 4, are automatically included in the next clustering process.

The system also provides a centroid-shift iteration history page that records centroid values at each iteration until convergence, making the K-Means clustering process traceable, as shown in Fig. 5.

Users can also update stored drug data through the Edit Drug Data page, including changes to name, category, stock, or expiration date, as shown in Fig. 6. Each saved change also affects clustering results in the next iteration.

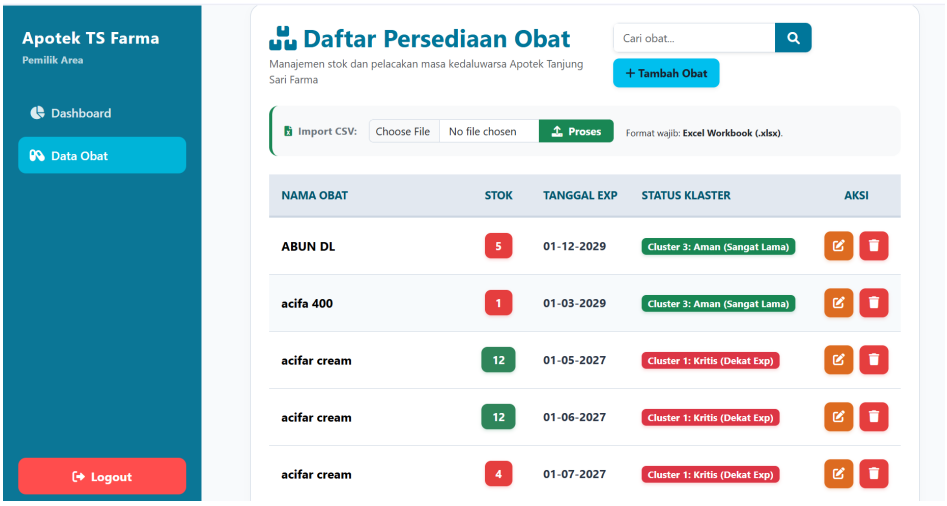

Fig. 3. The drug data page

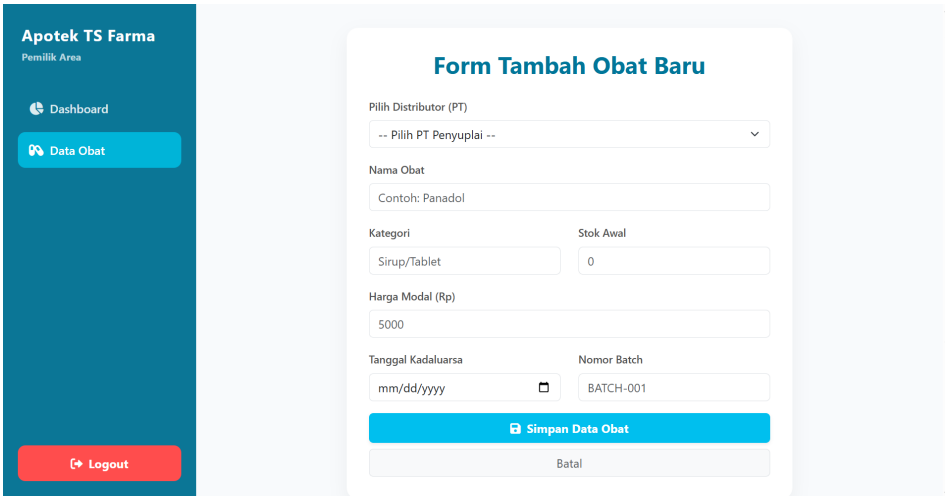

Fig. 4. The add-drug-data page

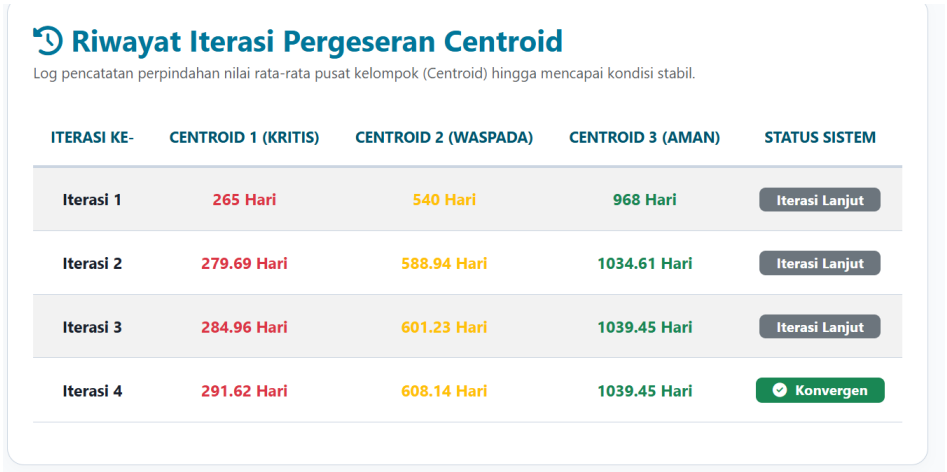

Fig. 5. Centroid-shift iteration history

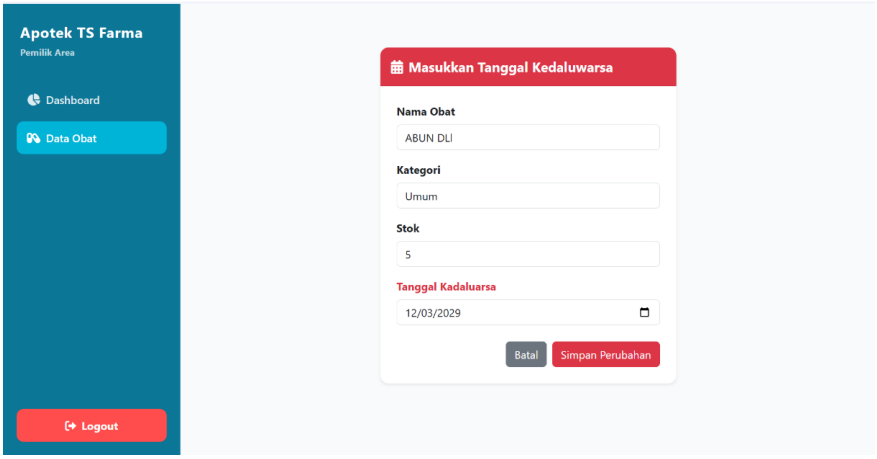

Fig. 6. The edit-drug-data page

Whenever data are added, imported, or updated, the system automatically recalculates the Expired Day value and runs K-Means clustering so that drug-priority status is consistently updated to reflect current inventory.

Functional black-box testing of the login, dashboard, drug-data, K-Means clustering, and logout modules showed valid results for every test scenario, with no functional errors found in the system interface. White-box testing of program logic, calculated using cyclomatic complexity, yielded values of 3 for the login module, 1 for the dashboard, 3 for the K-Means clustering module, and 1 for the logout module. These values indicate low-to-moderate program-flow complexity, making it easier to understand, test, and maintain, and confirm that all independent execution paths ran without logical errors.

F. Discussion

The results show that K-Means can effectively group drug data using only one numerical variable, Expired Day, without requiring complex additional attributes. This approach is consistent with several related studies applying K-Means to pharmaceutical data [1, 2, 3, 4, 5], but differs in its choice of variable: previous studies often emphasized stock volume or sales frequency, whereas this study focuses on expiration time, which is directly relevant to FEFO. Rapid convergence and valid results from all black-box and white-box tests indicate that this clustering feature is suitable for use as part of pharmacy operations.

Nevertheless, using a single variable means that clustering does not yet consider other factors, such as sales turnover or stock volume, which can also affect procurement decisions in some cases. Future work incorporating additional variables, comparison algorithms, or automatic notifications could improve this sales-priority recommendation system.

V. CONCLUSION

This study successfully designed and implemented a web-based system that helps Tanjung Sari Farma Pharmacy monitor drug expiration dates in a more structured way than the previous manual recording process. K-Means Clustering grouped 1,782 drug records into three categories—Critical, Alert, and Safe—based on the Expired Day variable, with the process converging in the fifth iteration. The grouping results were shown to serve as the basis for sales prioritization according to the First Expired First Out principle, while helping reduce the risk of losses from medicines passing their expiration dates before being sold. Black-box and white-box tests also confirmed that all functions performed validly with low-to-moderate program complexity.

Future research is recommended to add supporting variables such as stock volume and sales turnover, test comparison clustering algorithms such as Fuzzy C-Means or DBSCAN, and equip the system with automatic notifications and real-time transaction-data integration so that sales-priority recommendations become more

adaptive to changing inventory conditions.

Acknowledgment

The author thanks Tanjung Sari Farma Pharmacy for its permission and willingness to support the data-collection process, and the Informatics Study Program, Faculty of Computer Science, University of Bandar Lampung for the guidance provided during this research.

References

[1] E. Haerani, E. Budianita, A. Nazir, and W. Mahesa, "Penerapan K-Means Clustering Pada Data Obat/Alkes di Apotik RSUD Selasih", Semin. Nas. Teknol. Informasi, Komun. dan Ind., vol. 1, no. 1, pp. 220–229, 2023.

[2] L. A. Prasetyo, I. Himawan, and R. A. Sihombing, "Pengelompokan Persediaan Obat Dengan Metode K-", J. Rekayasa Sist. Inf. dan Teknol., vol. 2, no. 2, pp. 759–768, 2024.

[3] 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, pp. 2841–2848, 2024, doi: 10.23960/jitet.v12i3.4921.

[4] F. P. Ferdy Pangestu, N. Y. Nur Yasin, R. C. Ronald Chistover Hasugian, and Y. Yunita, "Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat", J. Sisfokom (Sistem Inf. dan Komputer), vol. 12, no. 1, pp. 53–62, 2023, doi: 10.32736/sisfokom.v12i1.1461.

[5] S. Syahputra and A. Yasir, "Penerapan Data Mining Dalam Menghitung Pengelompokan Obat-Obatan Kadaluarsa Menggunakan K-Means", Device J. Inf. Syst. Comput. Sci. Inf. Technol., vol. 4, no. 2, pp. 171–181, 2023, doi: 10.46576/device.v4i2.4060.

[6] D. K. Damanik, A. Antoni, and M. Z. Siambaton, "Algoritma K-Means untuk Klasterisasi Obat di Puskesmas Simpang Limun Kota Medan", Hello World J. Ilmu Komput., vol. 3, no. 2, pp. 100–110, 2024, doi: 10.56211/helloworld.v3i2.549.

[7] E. Wahyuni, F. Yunita, B. Rianto, U. I. Indragiri, I. Hilir, and K. Clustering, "IMPLEMENTASI METODE K-MEANS CLUSTERING DALAM MENENTUKAN PRODUK FAST, MEDIUM DAN SLOW MOVING PADA APOTEK ZERINA", vol. 8, pp. 1–10, 2026.

[8] Y. A. Pranoto and N. Vendyansyah, "Implementasi Metode K-Means Clustering pada Aplikasi Pengelompokan Daerah Produksi Garam Implementation of the K-Means Clustering Method in Salt Production Area Grouping Applications", J. Bumigora Inf. Technol., vol. 5, no. 1, pp. 45–52, 2023, doi: 10.30812/bite/v5i1.2690.

[9] E. Tasia, E. Saputra, F. Muttakin, and A. Marsal, "Pengembangan Sistem Informasi Berbasis FEFO untuk Pengendalian Obat Kadaluwarsa di Apotek Rahman", J. Pendidik. dan Teknol. Indones., vol. 5, no. 1, pp. 23–38, 2025, doi: 10.52436/1.jpti.554.