KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 3, Page: 121 - 129, September 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Data Mining Analysis Using K-Means and Naive Bayes Methods for Drug Inventory Prediction

Filsha Rifi Maharani[1]✉, Rini Sovia[2], Gunadi Widi Nurcahyo[3]
[1] Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, 25221, Indonesia
[2] Information Technology, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, 25221, Indonesia
[3] Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, 25221, Indonesia
filsharifi@gmail.com

Abstract

The management of drug inventory still encounters issues related to unstable stock fluctuations, such as drug shortages and overstocking. This condition has the potential to cause operational losses and degrade the quality of patient care. This study aims to analyze categorized drug inventory using the K-Means Clustering and Naive Bayes algorithms. The K-Means algorithm is employed to group drug data into three categories: low, sufficient, and excessive. The Naive Bayes algorithm is utilized to classify and predict the status of drug inventory requirements. The dataset was obtained from RSUD Tualang over the past two years, comprising 206 drug data points with variables including name, category, stock, price, demand, and expiration date. The test results indicate that the K-Means algorithm successfully grouped the data into three clusters, namely Low (14 items), Sufficient (125 items), and Excessive (67 items), while the Naive Bayes algorithm successfully classified the status of drug inventory requirements with an accuracy of 95.63%. The findings demonstrate that the combination of Data Mining utilizing K-Means and Naive Bayes is effective for drug inventory analysis at RSUD Tualang, thereby assisting management in identifying inventory status and supporting more accurate, efficient, and timely procurement decision-making compared to conventional methods previously employed.

Keywords: Drug inventory, Data Mining, K-Means, Naive Bayes, Prediction

KomtekInfo Journal is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

1. Introduction

Rapid technological developments, especially in the current digital era, have transformed our daily activities and interactions [1]. This impact has become increasingly apparent since the emergence of the Fourth Industrial Revolution and the adoption of artificial intelligence across various fields [2]. While technology brings numerous conveniences, it simultaneously introduces new challenges that require rapid adaptation [3]. Prediction serves as a method to anticipate unexpected future events and provide insights for decision-making [4]. It offers a statement regarding future occurrences, constructed not randomly or intuitively, but based on historical data [5]. Drug inventory prediction is a mechanism for estimating when medications in a hospital will be depleted. Predicting drug inventory helps determine whether stock levels are adequate or nearing exhaustion, ensures the appropriate quantity to treat patient conditions, and prevents overstocking. Proper inventory management can significantly reduce hospital financial expenditures and minimize the disposal of expired medications [5].

In previous research conducted by Medi et al., a prediction system was developed utilizing Linear Regression and Mean Absolute Percentage Error (MAPE) calculations to forecast outgoing drug quantities and drug procurement needs. The research yielded a system with a MAPE value of 12.4% and a system acceptance test score of 89.5% [6]. Another study by Dhea Mei Sela and Abdul Halim Hasugian applied the Simple Additive Weighting (SAW) method within a web-based decision support system. Their system was designed to assist in determining accurate drug requirements using data obtained from the Bagan Asahan Community Health Center[7].

Simulation results and real data analysis demonstrated that the system provided more accurate recommendations for drug inventory management, alongside generating routine inventory reports that hospital management can utilize for precise and efficient procurement decision-making [7]. Furthermore, research conducted by Sarah and Haryanto in 2024 detailed the design of a web-based application employing the Statistical Parabolic Projection method to estimate future drug product

Submitted: 04-06-2026 | Revised: 07-07-2026 | Accepted: 15-08-2026 | Published: 30-09-2026

inventory quantities, aiming to address the limitations of existing systems [8].

Data Mining is the process of extracting and processing data containing vital information [9]. Data mining methods encompass various approaches for comprehending and managing information [10], extracting knowledge from databases by identifying patterns through mathematical statistics, artificial intelligence, and machine learning [11]. K-Means is a prominent data mining method utilized for data grouping or classification [12]. It is a clustering algorithm that partitions data into K clusters based on the shortest distance to the cluster center, or centroid [13], with the objective of facilitating data analysis [14]. K-Means generates groupings from a dataset by iteratively calculating the distance from each data point to the group center [15]. A study applying K-Means, conducted by Dudi Irawan et al., utilized a dataset comprising 600 consumer records. The findings indicated that segmentation with k = 2 and k = 4 revealed clearer patterns compared to k = 3, with each cluster exhibiting distinct characteristics regarding risk profiles, product preferences, and banking service engagement levels [16].

Naive Bayes is a data mining method applied to address classification problems by leveraging the concepts of Bayes' Theorem [17]. It is a popular strategy for text classification due to its effective matrix assumptions, rapid implementation, and straightforward procedures, particularly for text categorization, sentiment analysis, and spam detection [18]. Naive Bayes is widely recognized for its exceptionally low error rate compared to various other classification algorithms [19]. Research incorporating Naive Bayes, conducted by Bahri et al., evaluated models using cross-validation, with results visualized via ROC curves to assess discriminatory capability. The evaluation demonstrated that the accuracy and Area Under the Curve (AUC) values of the Naive Bayes model surpassed those of Support Vector Machines (SVM), reaching an AUC of 0.901 for Naive Bayes compared to 0.818 for SVM [20].

The process of predicting drug stock begins with the application of Data Mining to ensure that historical data on medication usage has been successfully collected and managed. This prepared data is subsequently processed using the K-Means algorithm to categorize inventory status into low, sufficient, or excessive levels. The segmentation results from K-Means then serve as labels or crucial inputs for the Naive Bayes model, which performs advanced classification to forecast future inventory patterns, thereby enabling management to establish proactive and accurate drug procurement priorities.

2. Methods

This research framework focuses on three core elements: historical drug inventory records from RSUD Tualang spanning the last two years (the data aspect), the integration of the K-Means Clustering algorithm for categorization and Naïve Bayes for classification (the methodological aspect), and the translation of these analytical findings into practical insights for hospital administration (the system aspect). Utilizing a quantitative Data Mining approach, the workflow is executed systematically through distinct phases—beginning with problem formulation, data collection, and preprocessing, followed by algorithmic execution via K-Means and Naïve Bayes, and concluding with result generation. This methodology was chosen for its proficiency in handling extensive pharmaceutical inventories to yield measurable, objective patterns and forecasts. The scope centers strictly on predicting medication stock levels at RSUD Tualang using parameters such as drug nomenclature, packaging units, classifications, starting and ending stock volumes, pricing, consumption demand, and expiry dates. Ultimately, these factors are categorized into three distinct inventory conditions—low, sufficient, and excessive—to enhance overall inventory control efficiency, with the complete stepwise procedure outlined in Figure 1.

Figure 1 above illustrates the research framework is a conceptual structure that defines the core elements of a study, encompassing several essential components that help organize and guide scientific investigation. With this framework, researchers can ensure that all critical aspects are systematically covered and integrated into the research process.

The research workflow starts with problem identification and data collection, which is followed by data processing involving cleaning, transformation, and validation. The refined data is then analyzed using K-Means clustering beginning with optimal parameter determination, centroid initialization, data assignment, and centroid updates resulting in three clusters. The resulting groupings are used as class labels in the Classification process (Naive Bayes), which includes input attributes and cluster labels, calculation of prior and likelihood probabilities, computation of posterior values, class prediction, and ultimately produces the final 3 clusters. The next stage is Program Implementation, concluding with the presentation of Results and Discussion.

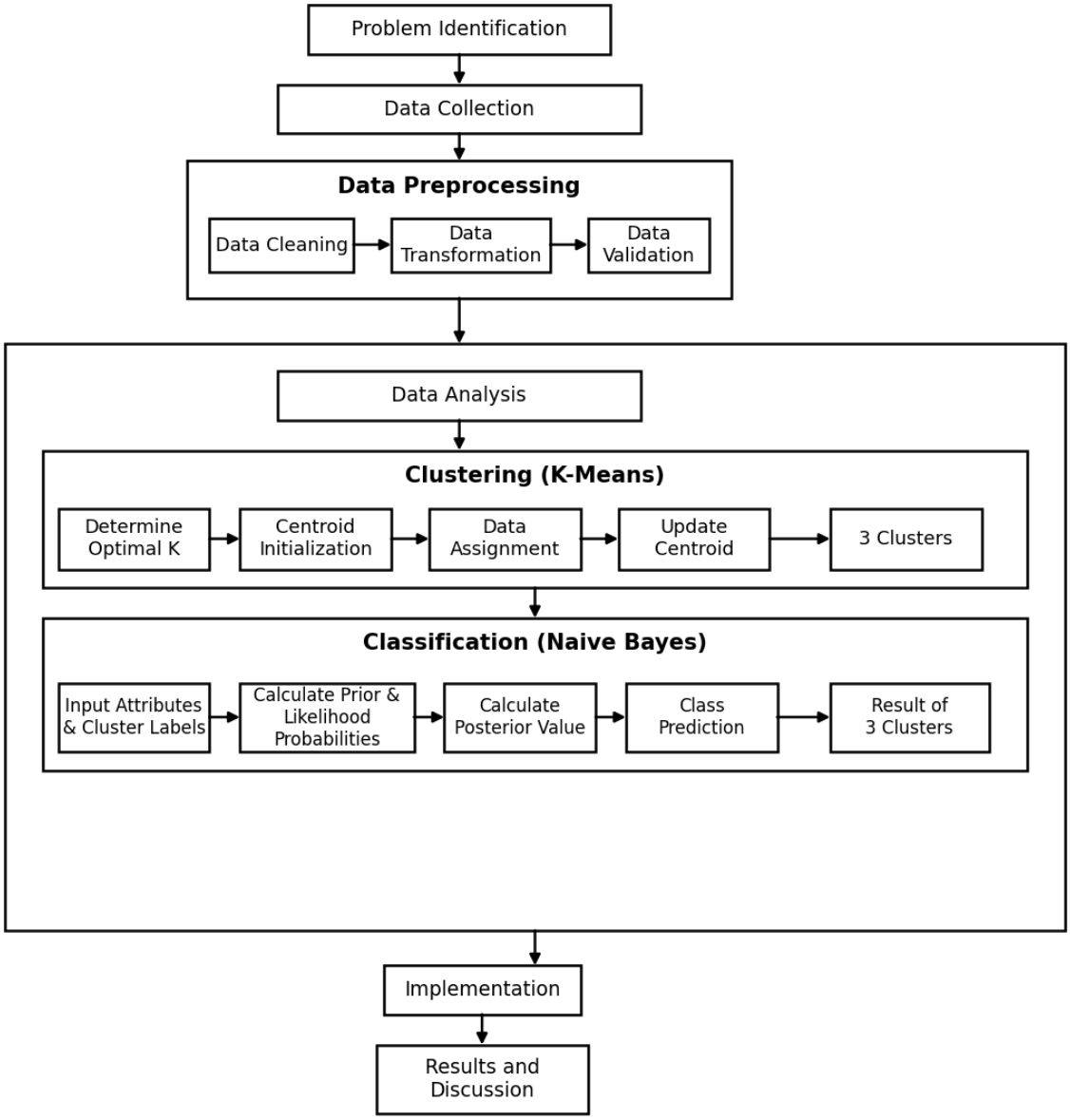


Figure 1. Research Framework

2.1 K-Means

K-Means Clustering is an unsupervised learning algorithm that groups data into several clusters based on the similarity of their characteristics or features [21]. Another approach involves the development of a set of functions that evaluate various characteristics of the clustering based on a number of clustering parameters [22]. The K-Means method falls under the category of distance-based clustering algorithms that partition data into multiple clusters, and this algorithm can only be applied to numeric attributes [23]. K-Means is used to group drug data based on usage patterns or categories, whereas Naïve Bayes predicts the appropriate inventory level for the future based on existing transaction or stock data [24]. The K-Means process yields a newly formed dataset with an additional cluster column, and this information is utilized to strengthen the input features. The Naive Bayes algorithm processes this data to

determine the final classification result [25]. The steps of the K-Means algorithm according to [26] are:

1. Determine k as the desired number of clusters.
2. Determine random values for the centroid (initial cluster) centers as many as k.
3. Calculate the distance of each input data point to each centroid using the Euclidean Distance formula until the shortest distance from each data point to the centroid is found. The formula for the Euclidean Distance formula is as follows:

$$De = \sqrt{(xi - sj)^2}(yi - tj)^2 \quad (1)$$

Xi is the criteria data and sj is the centroid of the ke-j cluster.
4. Classify the data based on its proximity to the centroid (shortest distance).
5. Repeat steps 3 to 5 if the members of each cluster remain unchanged. Through these stages, the K-Means algorithm is capable of objectively grouping drug data based on its numerical patterns without requiring prior class labels. The results of this grouping subsequently serve as the basis (labels) for the Naïve Bayes algorithm to perform drug inventory status classification in the next stage.

2.2 Naïve Bayes

Naive Bayes is a data mining method applied to address classification problems by leveraging the concepts of Bayes' Theorem [17]. It is a classification method used to solve multi-class classification problems, termed "Naive" because the calculation of probabilities for each created class is simplified, thereby facilitating its implementation [27]. This method operates by calculating the posterior probability for each category based on the attribute values within a dataset. Once the calculation process is complete, each data sample is classified into the category that possesses the highest posterior probability [28]. The Naive Bayes method is a popular strategy for text classification due to its effective matrix assumptions, rapid implementation, and straightforward procedures, particularly for text categorization, sentiment analysis, and spam detection [18]. Naive Bayes is widely recognized for its exceptionally low error rate compared to various other classification algorithms [19]. The application process of Naïve Bayes becomes more structured and reproducible on other datasets. The stages of Naïve Bayes consist of [29]:

1. Divide the data into training and testing sets.
2. Calculate Prior Probability: The initial probability of each class is calculated using the following formula:

$$Pc_i = \frac{s_i}{s} \quad (2)$$

P(Ci) is the amount of training data from category Ci, and N is the total amount of training data.
3. To calculate the likelihood probability, the model determines the probability for every attribute or feature across each individual class. Given that these attributes consist of numeric data, the mean and standard deviation derived from formulas (3) and (4) are applied in the calculations:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (3)$$

μ is the mean, x_i is the i-th data value, N is the total number of data points (sample size), and $\sum$ is the summation of all data values

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (4)$$

where σ is the sample standard deviation, x_i is the i-th data value, μ is the previously calculated mean, N is the total number of data points (sample size), and $\sum$ is the summation of all squared differences between the data and the mean.
4. Calculate Posterior Probability: Use Bayes' Theorem to obtain the probability of each class with the following formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (5)$$

where $f(x)$ is the normal distribution probability density function, x is the random variable value, μ is the population mean, σ is the population standard deviation, e is Euler's constant (approximately 2.71828), and π is the pi constant (approximately 3.14159).
5. Determine the Classification Class: Choose the class with the highest posterior probability. This represents the final result of the drug status classification.

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)} \quad (6)$$

where $P(A|B)$ is the probability of event A given that event B has occurred, $P(B|A)$ is the probability of event B given that event A has occurred, $P(A)$ is the prior probability of event A, and $P(B)$ is the prior probability of event B.

The Naïve Bayes algorithm is capable of probabilistically classifying drug data based on the statistical patterns of the training data. The combination of K-Means clustering results and Naïve Bayes classification yields a more accurate drug inventory status prediction system, thereby assisting RSUD Tualang in transitioning from a reactive inventory management approach to a proactive and data-driven one.

3. Results and Discussions

This section provides a comprehensive description of the data, the results of both methods, the analysis of the constructed classification model, and an in-depth discussion of the findings obtained from the research stages that have been conducted.

3.1 Dataset

The data utilized in this study consist of historical drug inventory data over the past two years, totaling 206 data points. Data collection in this research was conducted through direct observation at RSUD Tualang, Tualang District, Siak Regency, Riau, as well as interviews with the hospital's pharmaceutical management regarding drug inventory management patterns. The initial sample dataset is presented in Table 1.

Table 1 dataset

No	Drug Name	Unit	Category	Initial Stock	Final Stock	Drug price(Rp)	Demand	Expiration Date
1	Betahistine Mesylate 6 mg	tablet	vertigo	5000	3000	565	6000	01/11/2026
2	Dexamethasone 0,5 mg	tablet	Anti-inflamasi dan anti-alergi	2000	500	105	6000	01/12/2026
3	Natrium diklofenak 50 mg	tablet salut enterik	Pereda nyeri dan anti-inflamasi	5000	5000	195	200	01/05/2030
4	ipratropium bromida 0,5 mg+salbutamol 2,5 mg	ampul	Bronkodilator untuk asma atau PPOK	2000	1650	4.995	6000	01/08/2027
5	Levofloksasin inf 500 mg/100Ml	botol 100 mL	Antibiotik golongan kuinolon	40	0	37.550	2500	01/06/2029

The table above presents a sample of the first five data points representing the characteristics of the drug dataset used in this study. The dataset contains various important attributes related to pharmaceutical inventory management, including sequence number, drug name and dosage, packaging unit, functional category or drug classification, initial and final stock quantities, unit drug price, demand level, and the expiration date of each product. This raw data will subsequently undergo preprocessing and advanced analysis stages to optimize drug inventory management.

3.2 Variable Analysis

At this stage, the collected raw data undergoes a transformation process, where expiration dates are converted into numerical formats to enable further processing by computer algorithms. Table 2 presents the results of transforming non-numerical or complex attributes into structured numerical variables ($X_1 to X_4$).

Table 2 Data after transformation

No	X1	X2	X3	X4
1	5000	3000	6000	5
2	2000	500	6000	6
3	5000	5000	200	47
4	2000	1650	6000	14
5	40	0	2500	36

Based on Table 2 above, these variables represent data attributes that have been normalized or transformed, where X1 represents the initial stock, X2 represents the final stock, X3 represents the demand level, and X4 represents a derived variable or another conversion result. This transformed data is subsequently ready to be used as input in the clustering analysis process.

3.3 K-Means Classification Results

The findings from this clustering offer a transparent overview of the medication inventory status at RSUD Tualang, enabling healthcare providers and decision-makers to design targeted drug stock forecasting programs, with the complete data points comprising initial stock (X1), final stock (X2), demand (X3), remaining months to expiration (X4), and drug status (Y) detailed in Table 3.

Table 3 K-Means Results

No	X1	X2	X3	X4	Cluster(Y)
1	5000	3000	6000	5	Excessive
2	2000	500	6000	6	Excessive
3	5000	5000	200	47	Low
4	2000	1650	6000	14	Excessive
5	40	0	2500	36	Sufficient

Table 3 shows the results of drug grouping using the K-Means method based on the attributes of Initial Stock, Final Stock, Demand, and Remaining Months to Expiration. It can be observed that drug inventory with specific circulation characteristics tends to fall into the same inventory status cluster. Medications with extremely high inventory quantities but minimal demand levels predominantly fall into the Excessive Cluster. The visualization of K-Means is presented in Figure 2.

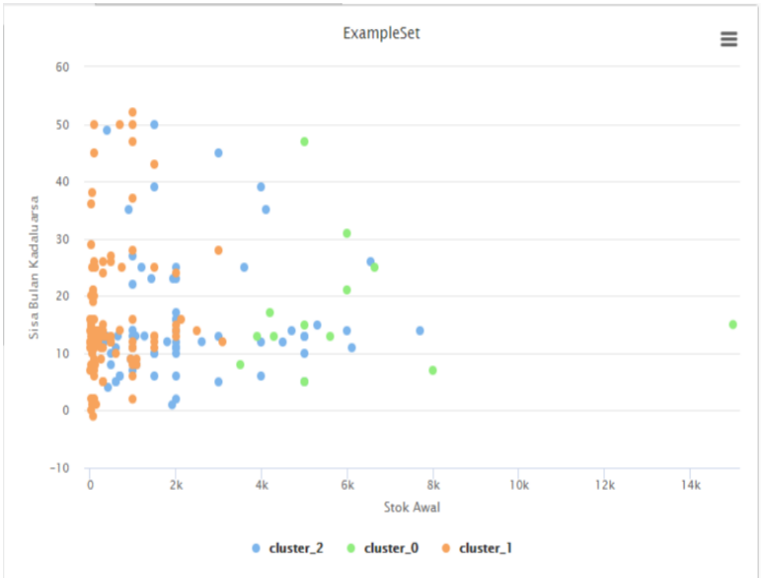


Figure 2 K-Means Scatter Plot Graph

The scatter plot graph visualizes the distribution of the three data clusters based on the Initial Stock and Remaining Months to Expiration variables, demonstrating that cluster_0 (green) is dominated by data with high initial stock, cluster_1 (orange) features low initial stock but a long remaining expiration period, while cluster_2 (blue) tends to lie within medium to low levels of stock and remaining months to expiration.

3.4 Naïve Bayes Classification Results

The highest classification result, or final probability, represents the largest probability value derived from the calculation of the posterior probability for each analyzed class. This value serves as the test data based on the Naïve Bayes calculations. A sample of the test data resulting from the Naïve Bayes classification is presented in Table 4.

Table 4 Naïve Bayes Results

No	X1	X2	X3	X4	Cluster(Y)
1	8	8	200	12	Cukup Obat
2	100	90	2500	11	Cukup Obat
3	100	60	2500	11	Cukup Obat
4	72	52	2500	12	Cukup Obat
5	72	52	2500	14	Cukup Obat

Based on Table 4, which presents the results of integrating the K-Means Clustering and Naïve Bayes methods, the model successfully classified the drug inventory status into three main categories: Drug Shortage, Sufficient Stock, and Overstock. The visualization of the Naïve Bayes results is presented in Figure 3.

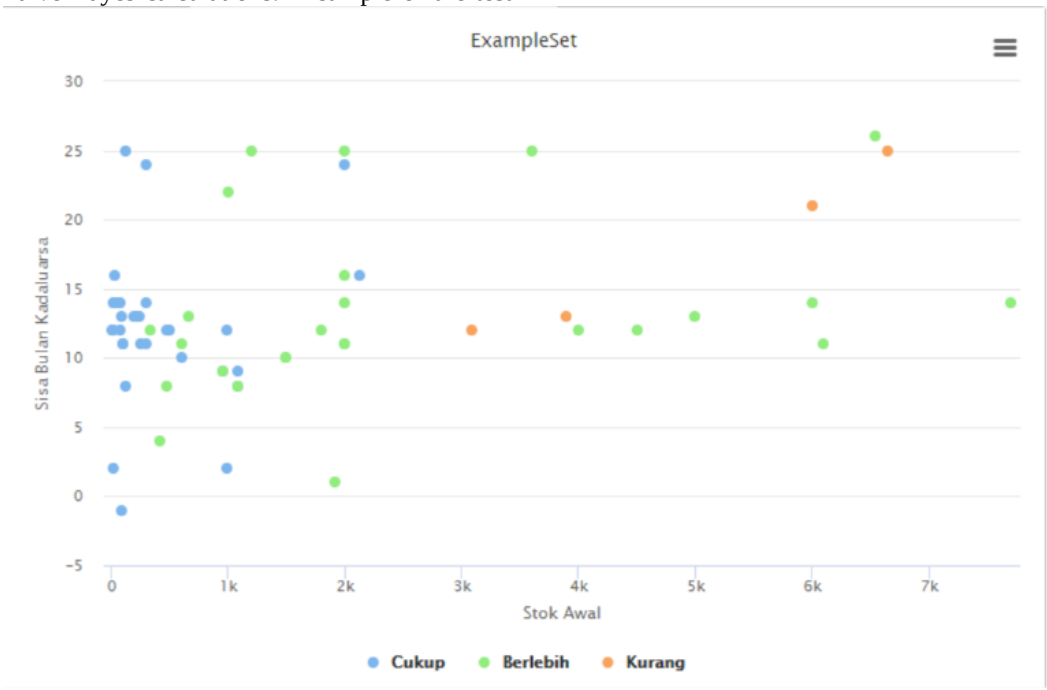


Figure 3 Naive Bayes Scatter Plot Graph

The scatter plot graph visualizes the distribution of data based on the Initial Stock and Remaining Months to Expiration variables with category labels for Sufficient (blue), Excessive (green), and Low (orange), demonstrating that the Sufficient category is concentrated at low initial stocks, the Excessive category is widely spread up to high initial stocks, whereas the Low category generally resides at medium to high initial stock levels.

3.5 Evaluation Results

The clustering process using the clustering procedure successfully partitioned the 206 pharmaceutical inventory records into three distinct categories—namely the low cluster containing 14 items, the sufficient cluster encompassing 125 items, and the excessive cluster consisting of 67 items using initial stock, final stock, consumption demand, and remaining expiration months as the baseline attributes. The centroid iteration process achieved convergence at the third iteration, indicated by no changes in the number of members within each cluster, thereby rendering the grouping results stable and representative of the data characteristics. These clustering results were subsequently utilized as class labels in the classification stage using the Naïve Bayes algorithm.

During the classification phase, the Naïve Bayes model determined final class predictions for each record by computing prior and likelihood probabilities derived from the preceding K-Means clusters, subsequently followed by posterior probability calculations. To validate these outcomes, the model's predictions were cross-checked against the real-world operational conditions of the pharmacy inventory, revealing that 197 out of the 206 evaluated records aligned accurately as valid, while 9 items proved inconsistent or invalid, as depicted in Figure 4.

Derived from the performance evaluation data, the suggested model exhibits steady excellence across every testing metric encompassing accuracy, precision, recall, and the F1-score reaching a consistent performance level of 95.63%. Accuracy is calculated by comparing the number of correctly classified data points to the total test data, yielding a value of 95.63% (197/206). Because the classification in this study is single-label multi-class with three mutually exclusive categories (Low, Sufficient, Excessive), precision and recall are calculated using a micro-average approach, where total correct predictions at the aggregate level become true positives while simultaneously eliminating the need to separate false positives and false negatives across classes. With this approach, a precision value of 95.63%, a recall of 95.63%, and an F1-score of 95.63% are obtained. The equality of these four metric values is a mathematical consequence commonly found in multi-class classification evaluation using the micro-average scheme, as also discovered in similar inventory classification studies.

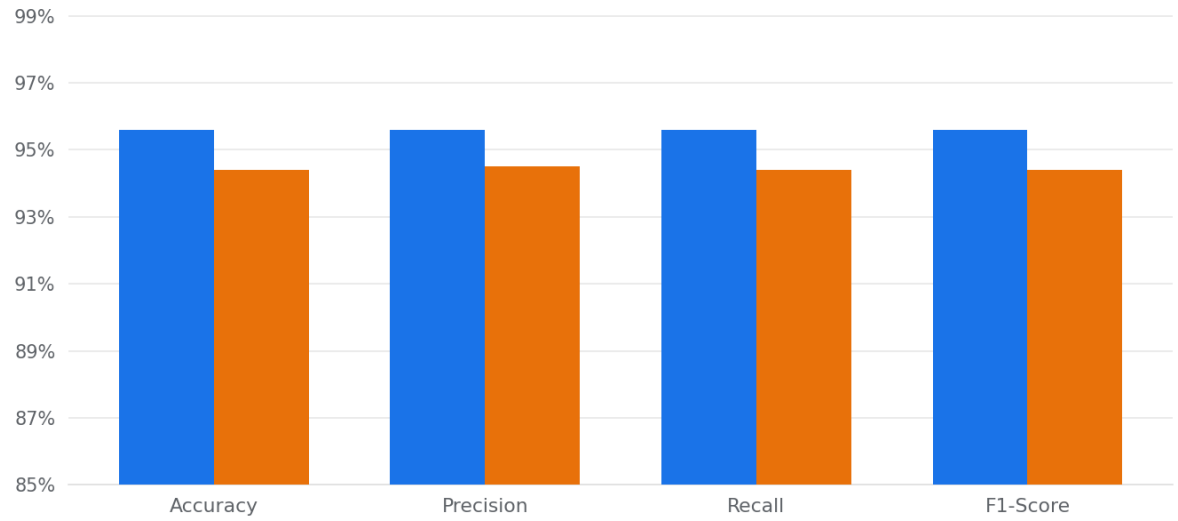


Figure 4. Evaluation Graph

The evaluation results indicate that the combination of K-Means Clustering and Naive Bayes is capable of providing an overview of drug inventory status while generating predictions with favorable performance. The accuracy value of 95.63% proves that the model can accurately classify inventory status (low, sufficient, excessive), whereas the K-Means grouping results help explain the relationships among initial stock, final stock, demand, and remaining months to expiration toward the inventory status categories. The findings of this study can provide more structured information to support RSUD Tualang in determining drug procurement priorities and formulating more accurate, efficient, and timely inventory management strategies.

4. Conclusions

This research aims to analyze drug inventory at RSUD Tualang through a combination of the K-Means Clustering and Naive Bayes algorithms, and the results demonstrate that this approach is capable of addressing the stock instability issues previously faced by the hospital. The K-Means algorithm successfully grouped 206 drug inventory data points based on the attributes of initial stock, final stock, demand, and remaining months to expiration into three clusters: the Low cluster comprising 14 items, the Sufficient cluster comprising 125 items, and the Excessive cluster comprising 67 items, with a centroid iteration process that converged and stabilized at the third iteration. These grouping results were subsequently utilized as class labels for the Naive Bayes algorithm, which was able to classify drug inventory requirement status with equivalent accuracy, precision, recall, and F1-score of 95.63%, where out of 206 test data points, 197 data points were found to be consistent with actual field conditions. These findings fulfill the research objective by showing that the integration of clustering and classification methods can yield an objective, measurable, and consistent drug inventory prediction system. The main contribution of this study lies in the application of a two-stage approach utilizing K-Means clustering results as input labels for the Naive Bayes classification model thereby producing a more structured inventory status prediction system compared to applying either algorithm separately or using conventional methods previously relied upon. Practically, these findings can be utilized by the pharmaceutical management of RSUD Tualang as a decision-support tool in determining drug procurement priorities, preventing stockouts and overstocking, and minimizing the risk of expired medications. Theoretically, this study reinforces empirical evidence that a Data Mining combination based on clustering and classification is effectively applied to the healthcare inventory management domain. For future development, subsequent research is recommended to expand data coverage across multiple hospitals and integrate additional variables to enhance the generalization and accuracy of the prediction model.

References

[1] A. Thahir, N. D. Hani, and M. J. Masri, "DAMPAK PERKEMBANGAN TEKNOLOGI TERHADAP POLA INTERAKSI SOSIAL PESERTA DIDIK DI SEKOLAH DAN KELUARGA," *Al-Irsyad J. Educ. Sci.*, vol. 4, no. 1, pp. 13–26, Feb. 2025, doi: 10.58917/aijes.v4i1.140.

[2] N. Imtikhani, A. Nugroho, and T. Farina, "Pengaruh Artificial Intelligence dalam Menggantikan Peran Manusia di Dunia Kerja Ditinjau dari Peraturan Pemerintah No 33 Tahun 2013 Tentang Perluasan Kesempatan Kerja," *J. Sos. Teknol.*, vol. 5, no. 5, pp. 1475–1496, Jun. 2025, doi: 10.59188/jurnalsostech.v5i5.32118.

[3] I. Antika, Y., Sabila, M., Natalia, D., Sari, D. A., & Nuddin, "Digitalisasi Pelayanan Publik di Era Revolusi Industri 4.0," *J. Ris. Multidisiplin Edukasi*, vol. 2, no. 5, pp. 358–369, 2025, doi: https://doi.org/10.71282/jurmie.v2i5.337.

[4] Z. Cao, L. Song, and S. Zhang, "Deep Learning-Driven Bus Short-Term OD Demand Prediction via a Physics-Guided Adaptive Graph Spatio-Temporal Attention Network," pp. 1–26, 2025.

[5] A. Ramazani, W. Fuadi, and R. Meiyanti, "Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Fuzzy Time Series Pada Perumda Aceh Utara," *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 8, no. 2, pp. 200–209, Jul. 2025, doi: 10.47324/ilkominfo.v8i2.360.

[6] M. Medi, S. P. A. Alkadri, and B. C. Octariadi, "Sistem Prediksi Kebutuhan Obat di Puskesmas Berbasis Website Menggunakan Metode Regresi Linear," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 3, p. 1215, Dec. 2023, doi: 10.35889/jutisi.v12i3.1543.

[7] D. Sela and A. H. Hasugian, "Metode Simple Additive Weighting Untuk Menentukan Jumlah Persedian Stok Obat Tbc Pada Puskesmas Bagan Asahan," *J. Process.*, vol. 19, no. 2, Oct. 2024, doi: 10.33998/processor.2024.19.2.1961.

[8] H. R. Sahra and E. V. Haryanto, "Rancang Bangun Peramalan Stok Obat Dengan Metode Statistical Parabolic Projection Pada PT. Tempo Scan Pacifik, Tbk Berbasis Web," *JID (Jurnal Info Digit.*, vol. 2, no. 3A, pp. 1812–1825, 2024.

[9] P. Hsu and S. Chen, "Communications of the Association for Information Systems Data Mining Augmented Survey Research," vol. 54, pp. 491–518, 2024.

[10] M. Y. Zidane, B. Nurina Sari, I. Maulana, A. Primaya, and G. Garno, "PENERAPAN DATA MINING DALAM KLASIFIKASI DATA TRANSAKSI PRODUK KOPERASI DI SMK PGRI 2 KARAWANG," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 263–269, Dec. 2024, doi: 10.36040/jati.v9i1.12196.

[11] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN PRODUK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 615–620, Feb. 2024, doi: 10.36040/jati.v8i1.8360.

[12] P. A. Gbadega, Y. Sun, and lufunke A. OBalogun, "Optimized energy management in Grid-Connected microgrids leveraging K-means clustering algorithm and Artificial Neural network models," vol. 336, no. 1, 2025.

[13] R. Rahmad, S. Defit, and R. Sovia, "Analisis Data Mining dengan Metode K-Means Clustering Dalam Pengelompokan Penggunaan Alat Kontrasepsi," *Bull. Comput. Sci. Res.*, vol. 5, no. 5, pp. 1174–1181, Aug. 2025, doi: 10.47065/bulletincsr.v5i5.750.

[14] Jaelani, Octaviana, and Elkin Rilvani, "Studi Literatur: Perbandingan Algoritma K-Means dan Fuzzy C-Means dalam Analisis Clustering," *J. Comput. Sci. Technol.*, vol. 5, no. 2, pp. 52–58, Aug. 2025, doi: 10.54840/jcstech.v5i2.382.

[15] Siti Maghfiroh and Zaehol Fatah, "Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTS miftahul ulum bengkak," *J. Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 133–140, Oct. 2025, doi: 10.35473/jamastika.v4i2.4527.

[16] D. Irawan, G. Wijaya, and T. T. Warisaji, "Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 6, no. 1, pp. 47–53, Mar. 2025, doi: 10.37148/bios.v6i1.162.

[17] P. Zhahiran Herlambang, N. Rahaningsih, and I. Ali, "OPTIMASI MANAJEMEN PERSEDIAAN DENGAN KLASIFIKASI NAIVE BAYES DI TOKO BAJA MANDIRI," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 5651–5657, May 2025, doi: 10.36040/jati.v9i4.13900.

[18] A. S. Perdana, Istian Kriya Almanfaluti, and Alshaf Pebrianggara, "Klasifikasi Ulasan Pelanggan pada Aplikasi MitraShopee Menggunakan Metode Support Vector Machine dan Naïve Bayes," *Teknologi*, vol. 15, no. 1, pp. 19–31, Jun. 2025, doi: 10.26594/teknologi.v15i1.5577.

[19] M. Afriansyah, Joni Saputra, V. Y. P. Ardhana, and Yuan Sa'adati, "ALGORITMA NAIVE BAYES YANG EFISIEN UNTUK KLASIFIKASI BUAH PISANG RAJA BERDASARKAN FITUR WARNA," *J. Inf. Syst. Manag. Digit. Bus.*, vol. 1, no. 2, pp. 236–248, Jan. 2024, doi:

10.59407/jismdb.v1i2.438.
[20] D. S. Bahri and U. Pamulang, "Analisis Perbandingan Algoritma Naïve Bayes dengan SVM (Support Vektor Machine) Dalam Mendiagnosis Penyakit Asma Comparative Analysis of Naïve Bayes and Support Vector Machine (SVM) Algorithms in Diagnosing Asthma," vol. 5, no. 2, pp. 189–198, 2025, doi: 10.59395/m9qa0357.
[21] F. Zafira, B. Irawan, and A. Bahtiar, "PENERAPAN DATA MINING UNTUK ESTIMASI STOK BARANG DENGAN METODE K-MEANS CLUSTERING," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 156–161, Feb. 2024, doi: 10.36040/jati.v8i1.8319.
[22] S. A. B. Telaumbanua, F. Setiadi, and S. Nurjanah, "Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset," *bit-Tech*, vol. 7, no. 3, pp. 984–994, Apr. 2025, doi: 10.32877/bt.v7i3.2287.
[23] 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, Apr. 2025, doi: 10.23960/jitet.v13i2.6174.
[24] P. Lu and Z. Zhang, "The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering," vol. 5, no. 5, pp. 1–17, 2025.
[25] H. Yzzogh and H. Benaboud, "Flooding distributed denial of service detection in software-defined networking using k-means and naïve Bayes," vol. 15, no. 1, pp. 817–826, 2025, doi: 10.11591/ijece.v15i1.pp817-826.
[26] M. A. K-means, J. Firmandanu, A. Hiswara, and F. N. Khasanah, "Pengelompokkan Sembako Terlaris," vol. 1, no. 2, pp. 71–74, 2025.
[27] H. Hartono, A. Hajjah, and Y. N. Marlim, "PENERAPAN METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI JUDUL BERITA," *J. Simantec*, vol. 12, no. 1, pp. 37–46, Dec. 2023, doi: 10.21107/simantec.v12i1.19398.
[28] N. P. Sari, "Analisis Performa Algoritma CNN dalam Klasifikasi Citra Medis Berbasis Deep Learning," *J. Komput.*, vol. 2, no. 2, pp. 87–92, Jun. 2024, doi: 10.70963/jk.v2i2.113.
[29] F. Handayani, R. Firdaus, A. Wahyudi, H. Fu'adah Amran, and R. Medikawati Taufiq, "Kombinasi Algoritma Gaussian Naïve Bayes Dan Adaboost Untuk Meningkatkan Akurasi Dalam Klasifikasi Penyakit Diabetes," *J. FASILKOM*, vol. 15, no. 2, pp. 238–244, Aug. 2025, doi: 10.37859/jf.v15i2.9279.

Biographies of Authors



Filsha Rifi Maharani I was born in Tualang on August 8, 2003, and completed her Bachelor of Computer Science degree majoring in Informatics Engineering at Universitas Putra Indonesia YPTK Padang, graduating in May 2025, with beginner-level data science competencies encompassing relational database management using SQL, tabular data processing with Excel and Python, basic web development, and visualization dashboard creation utilizing Tableau. I can be contacted via email: filsharifi@gmail.com



Rini Sovia Born in Solok, April 5, 1976. Lecturer at Putra Indonesia University YPTK Padang. He completed his education in informatics management in 1999, his master's degree in informatics engineering in 2006, and his doctoral education in information technology in 2023. His areas of expertise include artificial intelligence (AI), expert systems (ES), data mining (DM), decision support systems (DSS), and databases. I can be contacted via email: rini_sovia@upiyptk.ac.id



Gunadi Widi Nurcahyo was born in Temanggung, 14 March 1969. He was graduated bachelor's degree in informatics management at Universitas Putra Indonesia YPTK Padang in 1992. He completed his Master and PhD in Computer Science at Universiti Teknologi Malaysia in 2003. I an be contacted via email: gunadiwidi@yahoo.co.id