Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1106

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Application of K-Means Clustering for Consumable Inventory Expenditure at BKPSDM of Kudus Regency

Windraningsih [1*✉], R. Rhoedy Setiawan[1], Supriyono[1]
[1]Information System, Faculty of Engineering, Universitas Muria Kudus, Kudus, Indonesia
*Corresponding Author: 202253178@std.umk.ac.id

Article Information

Article history:

No. 1106
Rec. January 30, 2026
Rev. March 19, 2026
Acc. March 31, 2026
Pub. April 11, 2026
Page. 1413 – 1429

Keywords:

- *K-Means Clustering*
- *CRISP-DM*
- *Inventory Expenditure*
- *Web-Based System*
- *BKPSDM*

ABSTRACT

The management of consumable inventory expenditure at the Human Resources and Personnel Development Agency (BKPSDM) of Kudus Regency continues to encounter challenges due to manual and unsystematic data analysis processes. This study uses a dataset of 93 expenditure records for consumable inventory items in the 2025 period, obtained from the official inventory expenditure data of BKPSDM Kudus Regency. The main variables analyzed are the quantity of expenditures and the total expenditure value. This study aims to classify inventory usage data using the K-Means Clustering method supported by the CRISP-DM framework. The variables used include the quantity and total value of item expenditures. A web-based system was also developed to support the clustering process and provide analysis results more effectively. The K-Means++ algorithm was implemented to obtain better centroid initialization. The findings show three main clusters representing low, medium, and high expenditure levels. The system presents clustering results in tables and charts, making them easier to interpret for decision-making. This research is expected to support inventory planning and improve efficiency in inventory management at BKPSDM Kudus.

How to Cite:

Windraningsih, & et al. (2026). Application of K-Means Clustering for Consumable Inventory Expenditure at BKPSDM of Kudus Regency. Jurnal Teknologi Informasi Dan Pendidikan, 19(1), 1413-1429. https://doi.org/10.24036/jtip.v19i1.1106

1. INTRODUCTION

Effective management of consumable inventory is essential to maintain operational efficiency within public sector institutions. At the Regional Civil Service and Human

Resources Development Agency (BKPSDM) of Kudus Regency, inventory expenditure data are still processed manually, limiting the ability to analyze consumption patterns, set procurement priorities, and control stock levels effectively. As a result, inventory planning decisions are not yet supported by a comprehensive analytical framework.

Clustering techniques in data mining are widely used to identify hidden patterns in inventory and transactional data. Clustering is a technique that groups data objects based on their similarity without relying on predefined class labels [1]. K-means is an unsupervised learning algorithm, first published by Stuart Lloyd in 1984, and it is one of the most widely used clustering algorithms. K-means is a relatively simple clustering method that partitions a dataset into k clusters. The algorithm is easy to implement and execute, relatively fast, adaptable to various types of data, and widely applied in practice [2]. However, the standard K-Means algorithm is known to be sensitive to the selection of initial centroids, which can lead to inconsistent clustering results and convergence to local optima, especially on complex or large datasets [3] [4].

To overcome this limitation, Arthur and Vassilvitskii (2007) introduced K-Means++, which improves centroid initialization and enhances clustering stability. Although alternative methods such as hierarchical clustering and DBSCAN offer certain advantages, they generally require higher computational cost or complex parameter tuning. Moreover, the implementation of clustering in government institutions remains limited, particularly within a structured framework such as CRISP-DM and integrated web-based systems.

Therefore, this study applies the K-Means++ algorithm using the CRISP-DM framework to analyze consumable inventory expenditure at BKPSDM of Kudus Regency and develops a web-based system for automated analysis and visualization. This research contributes by providing a more stable clustering model and a practical decision-support tool to enhance inventory management efficiency in the public sector.

2. RESEARCH METHOD

2.1. Research Framework

Figure X illustrates the research methodology adopted in this study, which follows the CRISP-DM (Cross Industry Standard Process for Data Mining) framework. CRISP-DM provides a structured and iterative approach consisting of six major phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

The **Business Understanding** phase focuses on identifying the core problem related to the management of consumable inventory expenditure data at BKPSDM Kudus Regency and defining the analytical objectives. The primary objective of this study is to classify inventory items based on expenditure quantity and total expenditure value in order to support more effective inventory planning and budget allocation.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

The **Data Understanding** phase involves data collection, exploration, and descriptive statistical analysis of the dataset. This stage includes examining data distribution, detecting potential outliers, and assessing attribute consistency to ensure relevance to the research objectives.

The **Data Preparation** phase encompasses preprocessing activities such as data cleaning, feature selection, data transformation, and normalization where necessary. These procedures are conducted to ensure data quality and readiness prior to model implementation.

During the **Modeling** phase, the K-Means Clustering algorithm is applied with centroid initialization using the K-Means++ method to improve clustering stability and reduce sensitivity to initial centroid selection. The clustering process is performed iteratively until convergence criteria are satisfied based on minimal centroid displacement.

The **Evaluation** phase assesses the quality and validity of the generated clusters using the Silhouette Score metric. This evaluation quantitatively measures intra-cluster cohesion and inter-cluster separation to determine the strength of the clustering structure.

Finally, the **Deployment** phase represents the implementation of the clustering model into a web-based system and the presentation of analytical reports and visualizations to support managerial decision-making in inventory management.

By adopting the CRISP-DM framework, this study ensures that the data mining process is conducted systematically, rigorously, and in accordance with internationally recognized methodological standards.

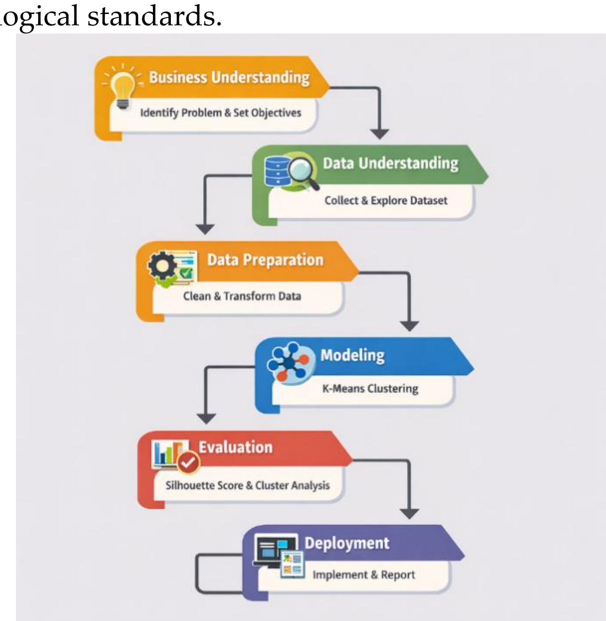

Figure 1. Research Framework Diagram Based on CRISP-DM

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1106

2.2. System Design

2.2.1. Use Case Diagram

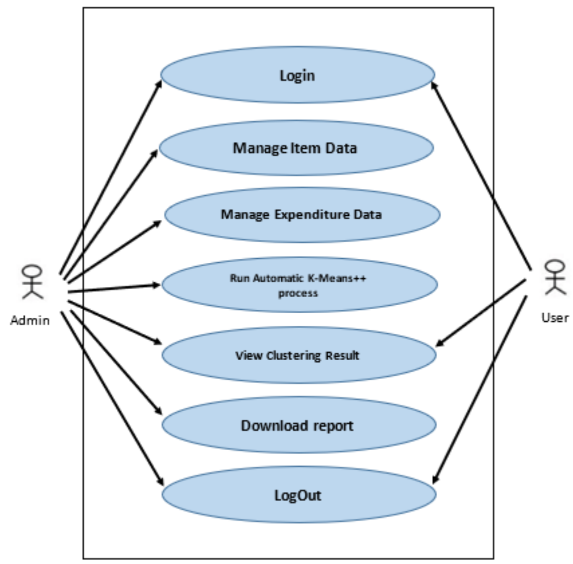

Figure 2. Use Case Diagram

Actors
1) Administrator: Has full access rights to manage and control all system features and data.
2) user: Responsible for login, logout and viewing clustering analysis results.
3) Main Use Cases
4) Authentication: Validating user access to the system.
5) Item Data Management: Managing inventory master records.
6) Expenditure Data Management: Recording and maintaining expenditure data.
7) K-Means++ Processing: Executing automated clustering analysis.
8) Result Visualization: Displaying clustering outcomes.
9) Report Generation: Producing analytical reports.
10) Logout: Securely ending user sessions.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

2.2.2. ERD (Entity Relationship Diagram)

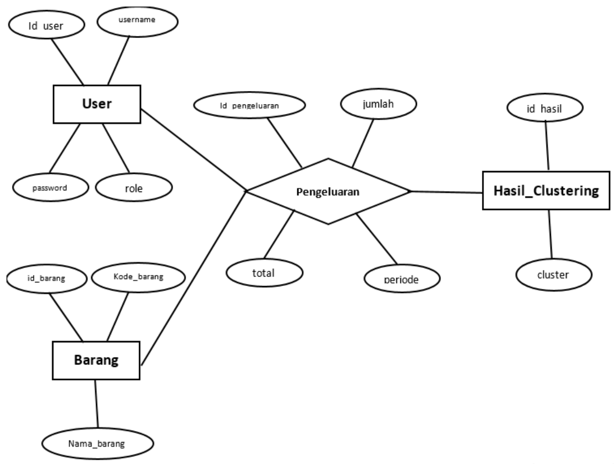

Figure 3. ERD (Entity Relationship Diagram)

The ERD defines the database structure and entity relationships to support efficient data management in the inventory expenditure clustering system. Entities: User, Item, Expenditure, Clustering_Result. Relationships: Users manage Expenditure data; Items are referenced in Expenditure records; Expenditure data generate Clustering_Result outputs.

2.2.3. Class Diagram

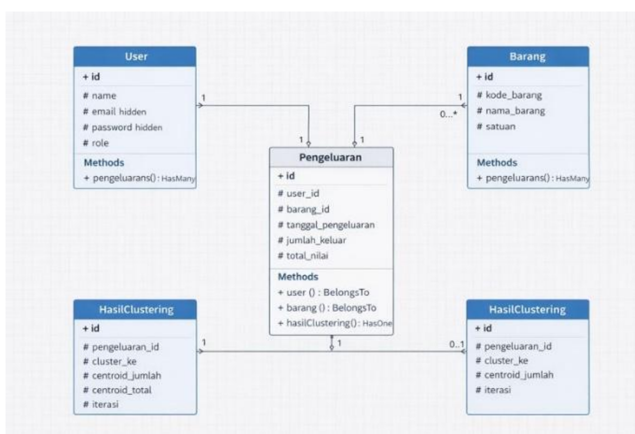

Figure 4. Class Diagram

The system consists of four main classes: User, Item, Expenditure, and ClusteringResult. User and Item have one-to-many relationships with Expenditure, which stores transaction data processed by the K-Means algorithm. Each Expenditure record is

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1106

linked to a single ClusteringResult, ensuring structured and efficient storage of clustering outputs.

2.2.4. K-Means++ Algorithm Flowchart

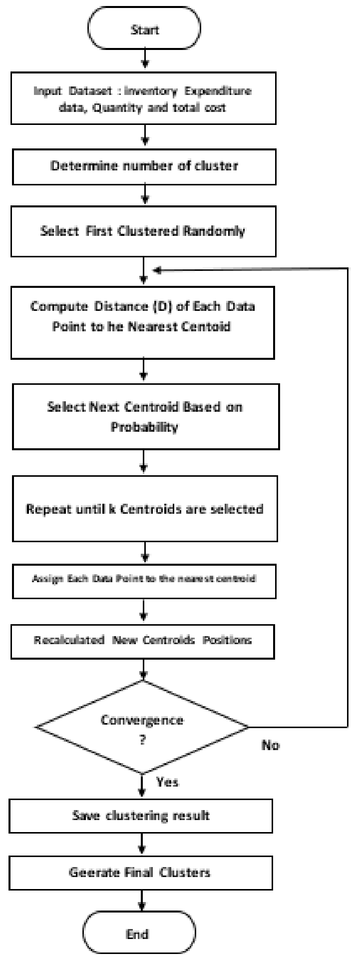

Figure 5. K-Means++ Algorithm Flowchart

The K-Means++ clustering process begins by retrieving consumable inventory expenditure data, including quantity and total cost. The data are normalized to ensure balanced variable scaling. The first centroid is randomly initialized, and subsequent centroids are selected based on squared distance probability to improve centroid distribution.

Distances between data points and centroids are computed using Euclidean Distance, after which data are assigned to the nearest cluster. Centroid positions are iteratively updated by calculating the mean of cluster members. The process continues until

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

convergence criteria are met, such as stable centroid positions or reaching the maximum iteration limit.

Once converged, the final clustering results are stored and visualized using scatter plots and expenditure distribution charts, completing the K-Means++ analysis process.

3. RESULTS AND DISCUSSION

3.1. K-Means++ and K-Means Clustering Methods Implementation Using Microsoft Excel

Table 1. Data Transformation & Modeling

No	Item Code	Item Name	Quantity of Expenditure	Total Expenditure Amount
1	ATK-006	Item-001	11	550
2	ATK-007	Item-002	4	280
3	ATK-008	Item-003	5	1350
4	ATK-009	Item-004	4	680
5	ATK-010	Item-005	13	780
6	ATK-011	Item-006	17	680
↓	↓	↓	↓	↓
64	ABK-008	Item-064	7	5600
65	ABK-013	Item-065	3	11850
66	ABK-019	Item-066	5	2500
67	ABK-025	Item-067	3	3150
68	ABK-026	Item-068	7	23870
69	ABK-027	Item-069	3	2550
↓	↓	↓	↓	↓
85	KCV-001	Item-085	153	84150
86	KCV-015	Item-086	12	7320
87	KCV-016	Item-087	1	320
88	LEL-030	Item-088	2	3040
89	LEL-031	Item-089	2	1140
90	LEL-032	Item-090	3	900
91	LEL-035	Item-091	5	300
92	LEL-036	Item-092	5	2500
93	LEL-037	Item-093	5	1650

The stages of the K-Means++ algorithm calculation are as follows:

3.1.1. Determining the Centroid Points Using the K-Means++ Algorithm

This stage involves selecting the initial centroid points before performing the clustering process.

3.1.1.1. Selecting the first centroid randomly from the dataset.

Table 2. centroid randomly from the dataset

data ke	X	Y
1	11	***550***

3.1.1.2. Calculating the distance of all data points in the dataset from the selected centroid using the following formula.

$$d1 = (11 - 11)^2 + (550 - 550)^2 = 0$$
$$d2 = (11 - 4)^2 + (550 - 280)^2 = 72{,}949$$
etc..

Table 3. calculating the distance of all

data ke	X	Y	hasil
1	11	***550***	-
2	4	***280***	72.949
3	5	***1.350***	640.036
:	:	:	:
88	2	***3.040***	6.200.181
89	2	***1.140***	348.181
90	3	***900***	122.564
91	5	***300***	62.536
92	5	***2.500***	3.802.536
93	5	***1.650***	1.210.036

3.1.1.3. The data point with the highest probability value is selected as the new centroid. Based on the calculation results, data point number 6 has the maximum probability value of 16.936; therefore, it is chosen as the second centroid (C2).

Table 4. centroid 2

data ke	X	Y
1	11	***550***
17	*680*	*16.936*

3.1.1.4. Steps 2 and 3 are repeated iteratively until the required number of clusters (k) is fulfilled and all centroids have been determined.

Table 4. centroid 2

data ke	X	Y
1	11	***550***
6	*17*	*680*
21	*290*	*5.800*

3.1.2. Computation Process Using the K-Means Clustering Algorithm

At this stage, the collected data are prepared according to the system used for data mining analysis. The transformation process is carried out by inputting consumable inventory expenditure data into the data mining application.

3.1.2.1. Determination of the Number of Clusters

The number of clusters is determined based on inventory expenditure characteristics, namely C1, C2, and C3. Therefore, the number of clusters (k) is set to three.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

3.1.2.2. Initialization of Initial Cluster Centroids

The selection of initial centroids in the K-Means algorithm is a critical step, as the clustering results are highly dependent on the initial centroid positions. In this calculation, the initial centroid positions are defined as follows :

Table 5. Initial Centroid Positions

record	centroid	X	Y
1	1	11	550
6	2	17	680
21	3	290	5.800

3.1.2.3. Euclidean Distance Comparison

The Euclidean Distance formula is used to calculate the distance between each data point and the centroid :

$$\textbf{distance} = \sqrt{(x1 - c1)^2 + (x2 - c2)^2} \tag{3}$$

x1: expenditure quantity of the data
x2: total expenditure value of the data
c1, c2: centroid coordinates

3.1.2.4. Classification of Data Based on the Nearest Centroid

After calculating the distance to all centroids, the smallest distance is identified using the MIN function. Each data point is then assigned to a cluster using the following criteria:

= IF(C0=MIN(Range C0 to C2), "0", IF(C1=MIN(Range C0 to C2), "1", "2")) (4)

C0: Euclidean distance to Centroid 0
C1: Euclidean distance to Centroid 1
C2: Euclidean distance to Centroid 2

3.1.2.5. Updating the Centroid Values

Once all data points are assigned to clusters based on their nearest centroids, the next step is to update the centroid values. The new centroid for each cluster is calculated as the mean of all data points belonging to the respective cluster.

Table 6. Initial Clustering Calculation Results

Record	X	Y	C1	C2	C3	Minimum	cluster
1	11	550	0,000	130,138	5257,408	0,000	1
2	4	280	270,091	400,211	5527,404	270,091	1
3	5	1.350	800,022	670,107	4459,117	670,107	2
4	4	680	130,188	13,000	5127,982	13,000	2
5	13	780	230,009	100,080	5027,637	100,080	2
6	17	680	130,138	0,000	5127,273	0,000	2
↓	↓	↓	↓	↓	↓	↓	↓
64	7	5.600	5050,002	4920,010	346,539	346,539	3
65	3	11.850	11300,003	11170,009	6056,804	6056,804	3
66	5	2.500	1950,009	1820,040	3312,284	1820,040	2
67	3	3.150	2600,012	2470,040	2665,496	2470,040	2
68	7	23.870	23320,000	23190,002	18072,216	18072,216	3
69	3	2.550	2000,016	1870,052	3262,648	1870,052	2
↓	↓	↓	↓	↓	↓	↓	↓
85	153	84.150	83600,121	83470,111	78350,120	78350,120	3
86	12	7.320	6770,000	6640,002	1545,213	1545,213	3
87	1	320	230,217	360,355	5487,615	230,217	1
88	2	3.040	2490,016	2360,048	2774,985	2360,048	2
89	2	1.140	590,069	460,245	4668,891	460,245	2
90	3	900	350,091	220,445	4908,398	220,445	2
91	5	300	250,072	380,189	5507,379	250,072	1
92	5	2.500	1950,009	1820,040	3312,284	1820,040	2
93	5	1.650	1100,016	970,074	4159,775	970,074	2

After the initial cluster assignment, centroid positions are updated by computing the mean of all data points within each cluster to better represent group characteristics.

3.1.2.6. Iteration

The K-Means algorithm iteratively performs distance computation, cluster assignment, and centroid updating until cluster memberships stabilize and no further centroid changes occur. In this study, steps 2–5 are repeated until convergence is achieved to ensure optimal and representative clustering results.

Table 7. Iteration Process Results

Record	X	Y	C1	C2	C3	Minimum	cluster	information
1	11	550	221,740	1079,663	12087,126	221,740	1	secure
2	4	280	48,351	1349,704	12357,144	48,351	1	secure
3	5	1.350	1021,667	279,860	11287,146	279,860	2	secure
4	4	680	351,669	949,725	11957,146	351,669	1	updated
5	13	780	451,732	849,657	11857,121	451,732	1	updated
6	17	680	351,862	949,653	11957,112	351,862	1	updated
↓	↓	↓	↓	↓	↓	↓	↓	↓
64	7	5.600	5271,667	3970,358	7037,172	3970,358	2	updated
65	3	11.850	11521,667	10220,356	788,031	788,031	3	secure
66	5	2.500	2171,667	870,415	10137,153	870,415	2	secure
67	3	3.150	2821,668	1520,402	9487,165	1520,402	2	secure
68	7	23.870	23541,667	22240,350	11232,966	11232,966	3	secure
69	3	2.550	2221,668	920,437	10087,161	920,437	2	secure
↓	↓	↓	↓	↓	↓	↓	↓	↓
85	153	84.150	83821,797	82520,462	71513,000	71513,000	3	secure
86	12	7.320	6991,670	5690,349	5317,169	5317,169	3	secure
87	1	320	9,374	1309,736	12317,154	9,374	1	secure
88	2	3.040	2711,669	1410,415	9597,169	1410,415	2	secure
89	2	1.140	811,673	489,846	11497,155	489,846	2	secure
90	3	900	571,671	729,764	11737,150	571,671	1	updated
91	5	300	28,335	1329,696	12337,141	28,335	1	secure
92	5	2.500	2171,667	870,415	10137,153	870,415	2	secure
93	5	1.650	1321,667	23,028	10987,148	23,028	2	secure

Centroid positions converge at the 12th iteration, indicating stable cluster centers that define the Final Cluster Centers. These final centroids represent the optimal and converged expenditure groupings across three clusters.

Table 8. Final Cluster Center at Iteration 12

Record	X	Y	C1	C2	C3	Minimum	cluster	information
1	11	550	1294,747	13257,906	83600,121	1294,747	1	secure
2	4	280	1564,788	13527,914	83870,132	1564,788	1	secure
3	5	1.350	494,885	12457,914	82800,132	494,885	1	secure
4	4	680	1164,808	13127,914	83470,133	1164,808	1	secure
5	13	780	1064,741	13027,904	83370,118	1064,741	1	secure
6	17	680	1164,732	13127,902	83470,111	1164,732	1	secure
↓	↓	↓	↓	↓	↓	↓	↓	↓
64	7	5.600	3755,282	8207,917	78550,136	3755,282	1	secure
65	3	11.850	10005,279	1958,007	72300,156	1958,007	2	secure
66	5	2.500	655,384	11307,915	81650,134	655,384	1	secure
67	3	3.150	1305,347	10657,920	81000,139	1305,347	1	secure
68	7	23.870	22025,271	10062,114	60280,177	10062,114	2	secure
69	3	2.550	705,414	11257,919	81600,138	705,414	1	secure
↓	↓	↓	↓	↓	↓	↓	↓	↓
83	31	465	1379,799	13342,902	83685,089	1379,799	1	secure
84	46	690	1155,088	13117,919	83460,069	1155,088	1	secure
85	153	84.150	82305,380	70342,219	0,000	0,000	3	secure
86	12	7.320	5475,271	6487,910	76830,129	5475,271	1	secure
87	1	320	1524,819	13487,919	83830,138	1524,819	1	secure
88	2	3.040	1195,367	10767,921	81110,141	1195,367	1	secure
89	2	1.140	704,898	12667,918	83010,137	704,898	1	secure
90	3	900	944,840	12907,916	83250,135	944,840	1	secure
91	5	300	1544,781	13507,913	83850,131	1544,781	1	secure

Based on Table 5, C0, C1, and C2 denote the centroid values for each cluster at every iteration. The Minimum column represents the shortest distance between each data point and its nearest centroid, while the Cluster column indicates the assigned group. After the iterative process converges, the centroid values remain stable, and these final values are defined as the Final Cluster Centers.

3.2. Web-Based Application Implementation

The K-Means Clustering application for consumable inventory expenditure is developed using the PHP programming language with the Laravel framework. The application provides several main features, including:

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1106

3.2.1. Login Page: This page is a crucial component of the application's security system. Its function is to authenticate users before they can access the system's features and to ensure that only authorized users are allowed to log in.

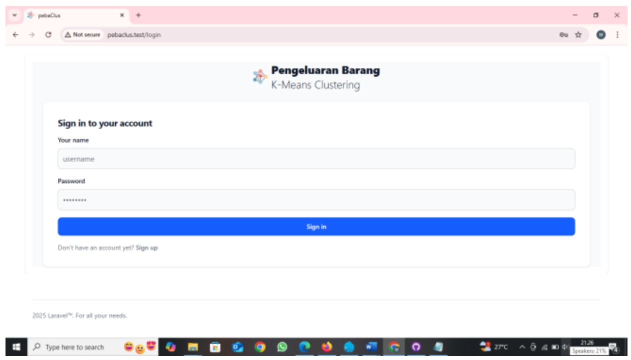

Figure 6. Login Page

3.2.2. Dashboard Menu: Displays the main interface of the Consumable Inventory Expenditure K-Means Clustering application.

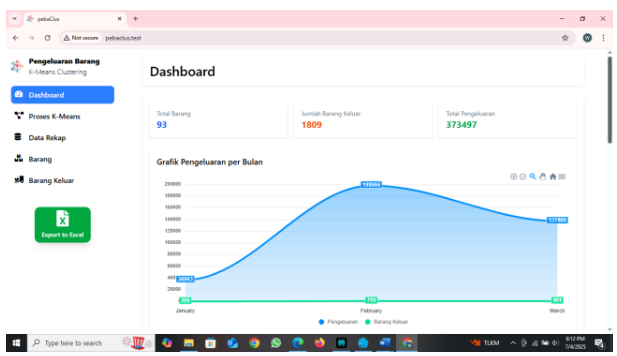

Figure 7. Initial Interface of the K-Means Clustering Application

3.2.3. Item Menu: Displays the master data of inventory items

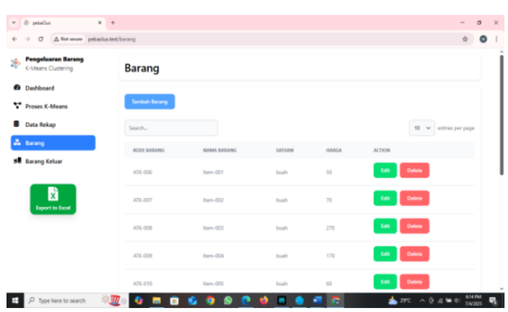

Figure 8. Item Menu Interface

3.2.4. Monthly Outgoing Items Menu: Displays the history of inventory expenditure based on monthly periods.

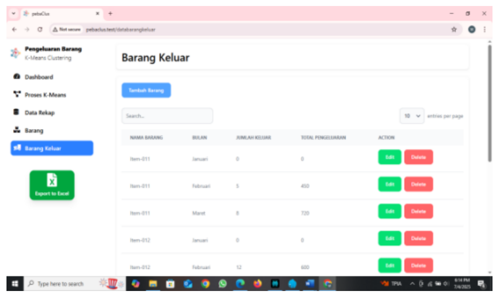

Figure 9. Monthly Outgoing Item Menu Interface

3.2.5. Menu Outgoing Goods Recap Data: Presents the total quantity and total value of goods expenditures within a specified time period.

Figure 10. Outgoing Goods Recap Data Menu Interface

3.2.6. Clustering Results Menu: Displays the results of clustering analysis, scatter plot visualizations, and downloadable reports in Excel format.

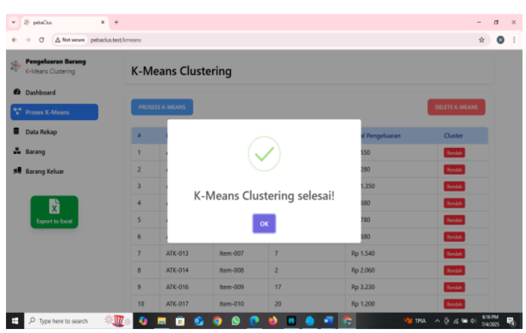

Figure 11. Clustering Process Interface

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1106

3.2.7. When the Clustering Process is clicked, a dialog box appears indicating that the K-Means Clustering process has been completed.

3.3. Clustering Analysis Results

The clustering analysis results form three groups in the Cluster column. The group names are displayed using descriptive words rather than numerical labels, as typically shown in Microsoft Excel, making the results easier for users to understand. The results can also be downloaded as an Excel file.

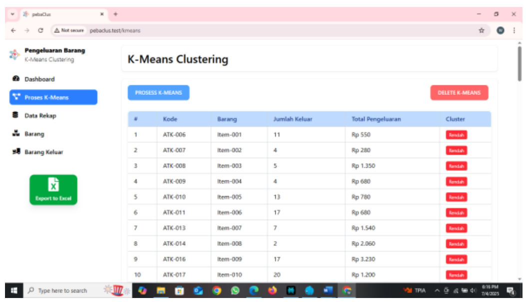

Figure 13. Clustering Analysis Results Interface

3.4. Scatter Plot

In addition to the analysis results, a scatter plot is displayed to illustrate the clustering formed. Three clusters are generated, and the scatter plot visualizes these clusters using three different colors to clearly distinguish each group.

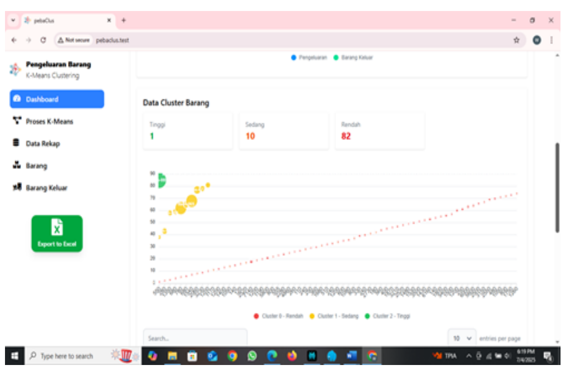

Figure 14. Scatter Plot Interface

3.5. Interpretation of Results

The clustering results of this study consist of three groups:

1) Cluster 0 (High) records the highest number of units, totaling 84,150 units, indicating intensive operational activities or major projects that require substantial inventory

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

consumption. This cluster plays a critical role in inventory expenditure management planning during the observed period. From a managerial perspective, items within this cluster should be prioritized in procurement planning, budget allocation, and stock monitoring to prevent shortages and operational disruptions.

2) Cluster 1 (Medium) demonstrates a moderate and stable level of inventory expenditure management planning, primarily associated with routine operational activities without additional or exceptional demand. Managerially, this cluster requires balanced stock control and periodic evaluation to maintain supply stability while avoiding overstocking.
3) Cluster 2 (Low) reflects the lowest level of inventory expenditure management planning, indicating minimal operational demand and limited inventory utilization during the period of analysis. Items in this cluster should be carefully reviewed to optimize storage costs and prevent unnecessary budget allocation, potentially through reduced procurement frequency or inventory rationalization strategies.

Overall, categorizing inventory expenditures into High, Medium, and Low clusters establishes an analytical basis for enhancing procurement strategies and optimizing budget utilization in public sector organizations. By distinguishing high-priority items (Cluster 0), ensuring controlled management of routine items (Cluster 1), and streamlining low-demand items (Cluster 2), policymakers can distribute resources more effectively while reducing the likelihood of stockouts or excess inventory. This cluster-oriented approach fosters greater transparency, accountability, and data-driven decision-making in inventory management, particularly within government institutions where fiscal planning and operational sustainability are essential.

4. CONCLUSION

This study concludes that the implementation of K-Means Clustering with K-Means++ successfully classified 93 inventory expenditure records at BKPSDM of Kudus Regency based on quantity and total expenditure variables. The evaluation findings suggest a well-structured clustering model characterized by clear separation between the High, Medium, and Low expenditure categories.

The findings provide measurable evidence that the K-Means++ initialization improves clustering stability and reduces the risk of suboptimal centroid selection compared to standard K-Means. The resulting clusters offer actionable insights for prioritizing procurement planning, optimizing budget allocation, and controlling inventory levels more systematically.

Scientifically, this research contributes by integrating the K-Means++ algorithm within the CRISP-DM framework and implementing it into a web-based decision-support system tailored to a government institution context. This integration demonstrates the practical applicability of data mining techniques in enhancing transparency, efficiency, and evidence-based inventory management in the public sector.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, *Konsep dasar dan teknik lanjutan dalam data mining*, 3rd ed. Elsevier, 2012.

[2] Wu, X. & Kumar, V., *The Top Ten Algorithms in Data Mining*. London: CRC Press. 2009.

[3] Zhao, S., Yang, R., Meng, F., and Cai, L., *The standard K-Means algorithm is known to be sensitive to the selection of initial centroids, which can lead to inconsistent clustering results and convergence to local optima, especially on complex or large datasets*. Elsevier, 2025.

[4] Khalish, F., Piranti, N.M., and Martadireja, O., "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means", Jurnal Ilmiah Ilmu Pendidikan, vol. 8, no. 5, pp. 5392-5397, 2025.

[5] Arthur, D. & Vassilvitskii, S., "k-means++: The Advantages of Careful Seeding." SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035, https://dl.acm.org/doi/10.5555/1283383.1283494.

[6] G. Richards, *Warehouse management: A complete guide to improving efficiency and minimizing costs in the modern warehouse*, Kogan Page Publishers, 2014.

[7] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. SPSS Inc., 2000.

[8] D. Kurniawansyah and J. Devitra, "Sistem informasi untuk manajemen stok barang habis pakai di instansi lingkungan hidup," *Jurnal Manajemen Sistem Informasi*, vol. 8, no. 4, pp. 568–581, 2023.

[9] T. B. Pamungkas, S. Maesaroh, and P. Ardiansyah, "Implementasi data mining pada stok penggunaan barang di GMF Aeroasia menggunakan algoritma K-means clustering," *Jurnal Ilmiah Sains dan Teknologi*, vol. 7, no. 2, pp. 112–123, 2023.

[10] Supriyono, K . F. Ferine, D. Puspitasari, Rulinawaty, and E. Timotius, "Implementation of data mining with Apriori techniques to determine the pattern of purchasing of agricultural equipment (Case study: XYZ Store)," *Journal of Physics: Conference Series*, vol. 1933, no. 1, 2021.

[11] R. R. Setiawan, "Desain informasi untuk penelitian database berbasis web," *Jurnal Sistem Informasi dan Teknologi*, vol. 1, no. 2, pp. 3–10, 2018.

[12] R. Setiawan, D. L. Fithri, A. P. Utomo, and F. Nugraha, "Penerapan optimalisasi marketplace untuk pemasaran produk pada UMKM keripik pare Alena Desa Damaran, Kecamatan Kota Kabupaten Kudus," *Muria Jurnal Layanan Masyarakat*, 2022.

[13] T. S. Ardan, S. A. Mujiatno, and B. A. Akbar, "Implementasi data analytics menggunakan metodologi CRISP-DM (Studi kasus data bike sales in Europe pada platform Kaggle)," *JSiI | Jurnal Sistem Informasi*, vol. 12, no. 2, pp. 10–15, 2025.

[14] A. Gunawan, S. Ningsih and D.A. Lantana, *Pengantar BASIS DATA*, PT. Literasi Nusantara Abadi Grup, 2023.

[15] Alfandi. M. Sofyan & Fatah. Zaehol, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Umama Hijab Kaliwates jember", *Jurnal Riset Sistem Informasi*, vol. 1, no. 4, hlm. 94–102, 2024.

[16] H. K. N. Iman, N. Latifah, S. Supriyono, and F. Nugraha, "Pemanfaatan metode decision tree dengan algoritma C4.5 untuk prediksi potensi kunjungan wisatawan," *Jurnal Sistem Komputer dan Informatika (JSON),* vol. 5, no. 3, p. 611, 2024.

[17] F. Nurdiyansyah and I. Akbar, "Implementasi algoritma K-means untuk menentukan persediaan barang pada poultry shop," *Jurnal Teknologi dan Manajemen Informatika,* vol. 7, no. 2, pp. 86–94, 2021.

[18] A. M. H. Pardede et al., "Implementation of data mining to classify the consumer's complaints of electricity usage based on consumer's locations using clustering method," *Journal of Physics: Conference Series,* vol. 1363, p. 012079, 2019.

[19] A. P. Rizkyandri, Jasmir, and Y. Arvita, "Implementasi Metode K-Means Clustering Untuk Menentukan Persediaan Barang Pada Toko SS Baby Shop," *Jurnal Informatika dan Rekayasa Komputer (JAKAKOM)*, vol. 3, no. 2, pp. 712–718, 2023.

[20] D. Ruswanti, D. Susilo and Riani, "Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45," *GO INFOTECH: JURNAL ILMIAH STMIK AUB*, vol. 30, no. 1, pp.111-121, 2024.

[21] N. C. Ningsih, D. Novaliendry, Y. Hendriyani, and D. Asmara, "Grouping of Student Achievement Based on Student Names in Class VII of SMPN 28 Sarolangun Using the K-Means Clustering Method," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 18, no. 2, pp. 864-882, 2025.

[22] A. B. P. Pradata, Supriyono, and D. L. Fithri, "Web-Based Expert System for Dental Disease Diagnosis Using the Certainty Factor Method: A Case Study at Al-Fatah Primary Clinic," *Jurnal Teknologi Informasi dan Pendidikan,* vol. 18, no. 2, pp. 984-1005, 2025.

[23] A. F. Hakim, Y. Irawan, and R. R. Setiawan, "Implementation of the CRISP-DM Methodology and Naive Bayes Algorithm on a Raw Material Requirement Prediction System to Reduce Food Waste (Case Study: Adamsafee Bakery, Resto, & Cafe)," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 18, no. 2, pp. 968-983., 2025.