Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
Vol. 03, No. 08, August 2023
e-ISSN: 2807-8691 | p-ISSN: 2807-839X

K-Means Algorithm Method for Clustering Best-Selling Product Data at XYZ Grocery Stores

Mohamad Maulana Ridzki[1], Ijah Hadijah[2], Mukidin[3], Adelia Azzahra[4], Aisyah Nurjanah[5]
Sekolah Tinggi Ilmu Komputer (STIKOM) PolTek Cirebon, Indonesia[1,2,3]
Public Administration, Universitas Swadaya Gunung Jati Cirebon, Indonesia[4]
English Language Education, UIN Sunan Gunung Djati Bandung, Indonesia[5]
Email: maulanaridzki0805@gmail.com[1], adeliaazzahra349@gmail.com[4], aisyahn40@gmail.com[5]

Keywords
K-Means, Algorithm Method, Clustering Sales, Product Data, Grocery Stores.

ABSTRACT
This study aims to utilize the K-means clustering algorithm in data mining to categorize sales data at XYZ Grocery store. The research is essential for understanding sales patterns and enhancing inventory management strategies. The research methodology involves implementing the K-means clustering algorithm to generate centroid values for each cluster, thereby creating groups of products based on their sales performance. The findings of this study are expected to provide insights into sales trends at the store. While the abstract provides a general overview, specific results and contributions of this research are not detailed. Further studies could offer a more in-depth understanding of the practical applications of these findings in improving store management and inventory control.

INTRODUCTION

In the contemporary retail landscape, where grocery stores play a pivotal role as essential providers of daily necessities, the escalating volume of sales data necessitates sophisticated methods for effective organization and analysis (Griffin et al., 2023; Khedmati & Azin, 2020; Pelekis et al., 2023; Vásquez Sáenz et al., 2023). This study focuses on the application of the K-Means Algorithm for clustering best-selling product data, offering a promising solution to systematically categorize products in grocery stores (Ikotun et al., 2023; Kashef & Pun, 2022; Luu et al., 2023). With consumer preferences evolving and a vast array of products available, leveraging the K-Means Algorithm presents an opportunity for grocery stores to gain valuable insights into consumer behavior, optimize inventory management, and tailor marketing strategies (Kuo et al., 2022; Meng et al., 2023; van der Borgh et al., 2023). The research aims to contribute to the enhancement of data-driven decision-making processes within the grocery retail sector, providing practical guidance for store managers and stakeholders in adapting to changing market dynamics and improving overall operational efficiency.

This research lies in the challenges faced by convenience stores as providers of daily and wholesale necessities. They encounter difficulties in manually recording sales transactions using ledger books. Their inventory management system adopts a sales-based approach to optimize warehouse capacity and reduce the risk of product damage (Mudzakkir, 2018; Zhou & Sun, 2023). However, the continued use of manual recording complicates the identification of fast-selling, moderately selling, and slow-selling items (Khedmati & Azin, 2020; Meng et al., 2023). Therefore, this research describes the need to implement a system capable of categorizing item data based on sales levels, highly popular categories, moderately popular categories, and less popular categories .

The urgency of this research arises from the difficulties experienced by XYZ convenience store in categorizing items based on sales levels. Thus far, the methods tested in previous research have not been fully adequate to address this issue in the context of wholesale stock management. Consequently,

https://doi.org/10.46799/ijssr.v3i12.652

this research proposes the use of the k-means algorithm as a new approach, expected to provide a more effective solution to this problem.

The novelty of this research lies in the application of the k-means method to categorize sales data within the context of wholesale stock management in convenience stores. This is an innovative step as there has been no prior research specifically utilizing the k-means algorithm to address item categorization issues in wholesale stock management.

With the goal of developing a system using the k-means algorithm, this research aims to facilitate the identification of highly popular, moderately popular, and less popular items. The benefits of this research involve enhancing the efficiency and effectiveness of wholesale stock management for XYZ convenience store. The implications lie in the contribution of this research to decision-making regarding more optimal procurement strategies, expected to provide valuable information for store management in better managing inventory.

METHODS

K-Means Clustering Method

The K-Means Clustering Method is a widely used statistical technique for non-hierarchical clustering, employing the K-Means Clustering Algorithm as a prominent algorithm in this category (Baihaqi et al., 2019; Fauzi, 2017; Jabat & Murdani, 2019). The method consists of several key steps:

Initially, the items are divided into K initial clusters, where K denotes the predetermined cluster count. It is crucial to note that the initial centers of the clusters are obtained randomly, introducing a stochastic element to the process.

Following this, a calculation process is initiated on the list of items. Each item is assigned to a specific group based on its proximity to the nearest center, determined using the Euclidean Distance. However, it is important to acknowledge K-Means' sensitivity to the initial conditions. To mitigate this, running the algorithm multiple times with different initializations is advisable and choosing the solution with the lowest sum of squared distances is advisable (Hadi & Diana, 2020; Nasution & Eka, 2018; Purba et al., 2018).

Subsequently, the algorithm recalculates the centroid centers for the newly formed clusters, accounting for any items that may have been initially missing from the clustering process. This iterative process persists until no more items are left to designate as new clusters. The primary objective of the algorithm is to optimize the grouping of items into clusters, refining the centroids with each iteration (Niu et al., 2021; Sinan et al., 2023; Wang et al., 2023).

In summary, the K-Means Clustering Algorithm employs an iterative approach to partition items into clusters, refining centroid centers by calculating Euclidean distances until convergence (Windarto, 2017). The random initialization of cluster centers introduces a stochastic element, contributing to this method's versatility and widespread application in clustering analysis (Chen et al., 2021; Miao et al., 2023). Additionally, it is important to address the sensitivity to initial conditions and consider running the algorithm multiple times for robust results.

Table file design

Table 1. Test Plan

Test Class	Test Grains	Types of Testing
Login	Password Verification	*Black box*
Add Product Data	Add Product Data	*Black box*
Transaction Data	Add Transaction Data	*Black box*
Sales Data Report	Print Sales Data	*Black box*
Sales Cluster	Sales Cluster Results	*Black box*

Table 2. Product Data

Test Cases and Results (Normal Data)			
Input Data	What to expect	Observation	Conclusion
Product Data	Data is saved to the database and can be managed again	Product data is inputted completely and is in accordance with the provisions	Accepted
Press the save button	The save button is available and the save data function can be used	Information appears that the data was successfully saved	Accepted

Test Design

Following the completion of system development, the subsequent phase involves testing the developed system using the black box method, specifically focusing on the Error Guessy method in this research. Black box testing centers on evaluating the functionality embedded within the system. The testing process encompasses critical aspects, including assessing the system interface's functionality. This involves evaluating the interface's ability to execute its functions seamlessly. Additionally, the research delves into scrutinizing the system's capability to effectively run its interface, ensuring that it operates as intended. Another essential dimension of black box testing involves evaluating the system's proficiency in handling inputs beyond its predefined boundaries. This includes assessing the system's adaptability to inputs that may fall outside the expected parameters. Lastly, the testing process extends to the system's ability to handle security issues, encompassing its resilience to potential threats and its effectiveness in safeguarding sensitive data. Through the systematic application of the Error Guessy method in testing these dimensions, the research aims to validate the robustness and reliability of the developed system. This comprehensive evaluation provides valuable insights into the system's performance, identifying potential areas for improvement or refinement.

RESULTS

Cluster Center Distance Calculation Results

To measure the distance between the data and the center of the cluster, Euclidian distance is used, then the distance matrix will be obtained as follows:

$d |= x – y = | \sum i \; n1 = (xi – yi)^2$ x = cluster center

y = data

In this case, the initial center of the cluster has been chosen: C1 is very marketable (80, 100), C2 sells well (40, 100) and C3 is undersold (10, 100). Then the distance calculation is carried out from the rest of the data sample with the cluster center, for example with M(a, b) where a is the stock of goods and b is the number of transactions sold.

D1 = (150, 22)
D2 = (100, 42)
D3 = (90, 21)
D4 = (50, 20)
D5 = (120, 87)

Calculate the Euclidean Distance of all data to the center of the first cluster.

$D1(p,c) = (22\text{-}80)^2+(150\text{-}100)^2 = 76.58$
$D1(p,c) = (22\text{-}40)^2+(150\text{-}100)^2 = 53.14$

$D1(p,c) = (22\text{-}10)^2\text{+}(150\text{-}100)^2 = 51.42$

$D2(p,c) = (100\text{-}80)^2\text{+}(42\text{-}100)^2 = 38.00$
$D2(p,c) = (100\text{-}40)^2\text{+}(42\text{-}100)^2 = 2.00$
$D2(p,c) = (100\text{-}10)^2\text{+}(42\text{-}100)^2 = 32.00$

$D3(p,c) = (90\text{-}80)^2\text{+}(21\text{-}100)^2 = 45.12$
$D3(p,c) = (90\text{-}40)^2\text{+}(21\text{-}100)^2 = 10.77$
$D3(p,c) = (90\text{-}10)^2\text{+}(21\text{-}100)^2 = 27.86$

$D4(p,c) = (22\text{-}80)^2\text{+}(150\text{-}100)^2 = 78.10$
$D4(p,c) = (22\text{-}40)^2\text{+}(150\text{-}100)^2 = 53.85$
$D4(p,c) = (22\text{-}10)^2\text{+}(150\text{-}100)^2 = 50.99$

$D5(p,c) = (120\text{-}80)^2\text{+}(87\text{-}100)^2 = 22.36$
$D5(p,c) = (120\text{-}40)^2\text{+}(87\text{-}100)^2 = 53.85$
$D5(p,c) = (120\text{-}10)^2\text{+}(87\text{-}100)^2 = 82.46$

From the calculation of Eucludien Distance to the first center point is obtained as follows:

Table 3. 0th iteration

Data	C1 Sells Highly	C2 Laku	C3 Undersells	Closest Distance
D1	76,58	53,14	51,42	3
D2	38,00	2,00	32,00	2
D3	45,12	10,77	27,86	2
D4	78,10	53,85	50,99	3
D5	22,36	53,85	82,46	1

It can be concluded from the table above is C1 = 1 data, C2 = 2 data and C3 = 2 data. To calculate the 1st iteration, a new cluster will be created in the following way:

The determination of the new cluster will be calculated by the formula: Number of Results / Value of Results.

C1 = (90+66+89)/3 = (81.67)
C1 = (120+70+50)3 = (80.00)
C2 = (42+36+56+29+33+39+
61+45)/8 = (43.50)
C2 = (100+90+41+80+50+50+
50+79)/8 = (67.50)
C3 = (22+21+20+15+7+14+11+6+15+15+15+18+
14+13+19+17+6+13+20+9+26+7)/21 = (15.05)
C3 = (150+50+70+80+66+55+67+40+45+49+
55+77+90+130+140+125+134+
122+100+50+80)/21 = (84.52)

In the same way calculate the distance of each point to the 1st center point:

$D1(p,c) = (150-81)^2+(22-80)^2 = 91.98$
$D1(p,c) = (150-43)^2+(22-67)^2 = 85.26$
$D1(p,c) = (150-15)^2+(22-84)^2 = 65.84$

$D2(p,c) = (100-81)^2+(42-80)^2 = 44.42$
$D2(p,c) = (100-43)^2+(42-67)^2 = 32.53$
$D2(p,c) = (100-15)^2+(42-84)^2 = 31.08$

$D3(p,c) = (90-81)^2+(21-80)^2 = 61.49$
$D3(p,c) = (90-43)^2+(21-67)^2 = 31.82$
$D3(p,c) = (90-15)^2+(21-84)^2 = 8.09$

$D4(p,c) = (22-81)^2+(150-80)^2 = 68.58$
$D4(p,c) = (22-43)^2+(150-67)^2 = 29.30$
$D4(p,c) = (22-15)^2+(150-84)^2 = 34.88$

$D5(p,c) = (120-81)^2+(87-80)^2 = 40.86$
$D5(p,c) = (120-43)^2+(87-67)^2 = 70.13$
$D5(p,c) = (120-15)^2+(87-84)^2 = 82.92$

Table 4. 1st iteration.

Data	C1 Sells Highly	C2 Laku	C3 Undersells	Closest Distance
D1	91,98	85,26	65,84	3
D2	44,42	32,53	31,08	3
D3	61,49	31,82	8,09	3
D4	68,58	29,30	34,88	2
D5	40,86	70,13	82,92	1

It can be concluded from the table above is C1 = 1 data, C2 = 1 data and C3 = 3 data. To calculate the 2nd iteration, a new cluster will be created in the following way:
The determination of the new cluster will be calculated by the same formula: Number of Results/Value of Results.

C1 = (90+89)/2 = (89.5)
C1 = (120+50)/2 = (85)
C2 = (20+66+15+36+15+36+15+18+56+
26+33+39+61+45)/12 = (35.83)
C2 = (50+44+40+41+45+49+80+50+
50+50+50+79)/12 = (52.33)
C3 = (22+42+21+15+7+14+11+6+14+13+
19+17+29+6+13+6+20+9+7)/18 = (15.38)
C3 = (150+100+90+70+80+66+55+67+55+77+
90+130+140+125+134+122+80)/18 = (96.16)

In the same way calculate the distance of each point to the center of the 2nd cluster:

$D1(p,c) = (150-89)^2+(22-85)^2 = 93.71$
$D1(p,c) = (150-35)^2+(22-52)^2 = 98.64$

$D1(p,c) = (150\text{-}15)^2+(22\text{-}96)^2 = 54.19$

$D2(p,c) = (100\text{-}89)^2+(42\text{-}85)^2 = 49.81$
$D2(p,c) = (100\text{-}35)^2+(42\text{-}52)^2 = 48.04$
$D2(p,c) = (100\text{-}15)^2+(42\text{-}96)^2 = 26.45$

$D3(p,c) = (90\text{-}89)^2+(21\text{-}85)^2 = 68.68$
$D3(p,c) = (90\text{-}35)^2+(21\text{-}52)^2 = 40.48$
$D3(p,c) = (90\text{-}15)^2+(21\text{-}96)^2 = 8.05$

$D4(p,c) = (22\text{-}89)^2+(150\text{-}85)^2 = 77.82$
$D4(p,c) = (22\text{-}35)^2+(150\text{-}52)^2 = 16.00$
$D4(p,c) = (22\text{-}15)^2+(150\text{-}96)^2 = 46.35$

$D5(p,c) = (120\text{-}89)^2+(87\text{-}85)^2 = 35.00$
$D5(p,c) = (120\text{-}35)^2+(87\text{-}52)^2 = 86.68$
$D5(p,c) = (120\text{-}15)^2+(87\text{-}96)^2 = 77.90$

Table 5. 2nd iteration

Data	C1 Sells Highly	C2 Laku	C3 Undersells	Closest Distance
D1	93,71	98,64	54,19	3
D2	49,81	48,04	26,45	3
D3	68,68	40,48	8,05	3
D4	77,82	16,00	46,35	2
D5	35,00	86,68	77,90	1

It can be concluded from the table above is C1 = 1 data, C2 = 1 data and C3 = 3 data. To calculate the 2nd iteration, a new cluster will be created in the following way:
The determination of the new cluster will be calculated by the formula: Number of Results / Value of Results.

C1 = (90+56+89)/3 = (78.33)
C1 = (120+50+80)/3 = (83.33)
C2 = (20+14+11+66+15+36+15+18+
14+26+33+39+61+45)/8 = (29.5)
C2 = (50+66+55+44+40+41+45+49+55+
50+50+50+50+79)/8 = (51.71)
C3 = (22+42+21+15+7+6+13+
19+17+6+13+20+9+7)/14 = (16.4)
C3 = (150+100+90+70+80+67+77+90+130+
140+125+134+122+100+80)/14 = (103.67)

In the same way count each center point to the 3rd point:

$D1(p,c) = (150\text{-}78)^2+(22\text{-}83)^2 = 87.28$
$D1(p,c) = (150\text{-}29)^2+(22\text{-}51)^2 = 98.57$
$D1(p,c) = (150\text{-}16)^2+(22\text{-}103)^2 = 46.47$

$D2(p,c) = (100\text{-}78)^2+(42\text{-}83)^2 = 39.97$
$D2(p,c) = (100\text{-}29)^2+(42\text{-}51)^2 = 49.88$
$D2(p,c) = (100\text{-}16)^2+(42\text{-}103)^2 = 25.86$

$D3(p,c) = (90\text{-}78)^2+(21\text{-}83)^2 = 57.72$
$D3(p,c) = (90\text{-}29)^2+(21\text{-}51)^2 = 39.22$
$D3(p,c) = (90\text{-}16)^2+(21\text{-}103)^2 = 14.22$

$D4(p,c) = (22\text{-}78)^2+(150\text{-}83)^2 = 61.19$
$D4(p,c) = (22\text{-}29)^2+(150\text{-}51)^2 = 9.65$
$D4(p,c) = (22\text{-}16)^2+(150\text{-}103)^2 = 53.79$

$D5(p,c) = (120\text{-}78)^2+(87\text{-}83)^2 = 38.48$
$D5(p,c) = (120\text{-}29)^2+(87\text{-}51)^2 = 91.23$
$D5(p,c) = (120\text{-}16)^2+(87\text{-}103)^2 = 75.39$

Table 6. Iteration-3

Data	C1 Sells Highly	C2 Laku	C3 Undersells	Closest Distance
D1	93,71	98,64	54,19	3
D2	49,81	48,04	26,45	3
D3	68,68	40,48	8,05	3
D4	77,82	16,00	46,35	2
D5	35,00	86,68	77,90	1

It can be concluded from the table above is C1 = 1 data, C2 = 1 data and C3 = 3 data. To calculate the 2nd iteration, a new cluster will be created in the following way:
The determination of the new cluster will be calculated by the formula: Number of Results / Value of Results.

C1 = (90+56+89)/3 = (78.33)
C1 = (120+80+50)/3 = (83.33)
C2 = (20+15+14+66+11+66+6+15+36+
15+18+14+26+33+39+61+45)/16 = (29.5)
C2 = (50+70+66+55+44+67+41+45+49+55+
50+50+50+50+79)/16 = (53.81)
C3 = (22+42+21+7+13+19+17+29+6+
13+20+9+7)/14 = (17.30)
C3 = (150+100+90+80+77+90+130+1
40+125+134+122+100+80)/14 = (109.07)

In the same way calculate each point to the center of the point kr-4:

$D1(p,c) = (150\text{-}78)^2+(22\text{-}83)^2 = 87.28$
$D1(p,c) = (150\text{-}29)^2+(22\text{-}53)^2 = 96.32$
$D1(p,c) = (150\text{-}17)^2+(22\text{-}109)^2 = 41.19$

$D2(p,c) = (100\text{-}78)^2+(42\text{-}83)^2 = 39.97$

D2(p,c) = $(100-29)^2+(42-53)^2$ = 48.52
D2(p,c) = $(100-17)^2+(42-109)^2$ = 26.31

D3(p,c) = $(90-78)^2+(21-83)^2$ = 57.72
D3(p,c) = $(90-29)^2+(21-53)^2$ = 36.70
D3(p,c) = $(90-17)^2+(21-109)^2$ = 19.43

D4(p,c) = $(22-78)^2+(150-83)^2$ = 67.19
D4(p,c) = $(22-29)^2+(150-53)^2$ = 8.08
D4(p,c) = $(22-17)^2+(150-109)^2$ = 59.14

D5(p,c) = $(120-78)^2+(87-83)^2$ = 38.48
D5(p,c) = $(120-29)^2+(87-53)^2$ = 91.29
D5(p,c) = $(120-17)^2+(87-109)^2$ = 73.51

Table 7. Iteration-4

Data	C1 Sells Highly	C2 Laku	C3 Undersells	Closest Distance
D1	87,28	96,32	41,19	3
D2	39,97	48,52	26,31	3
D3	57,72	36,70	19,43	3
D4	67,19	8,08	59,14	2
D5	38,48	91,29	73,51	1

It can be concluded from the table above is C1 = 1 data, C2 = 1 data and C3 = 3 data. From the results of iteration-0 and the last iteration-4 obtained the same result, the iteration is declared over.

System Design Results

Store owners must log in first before logging in to the application system.

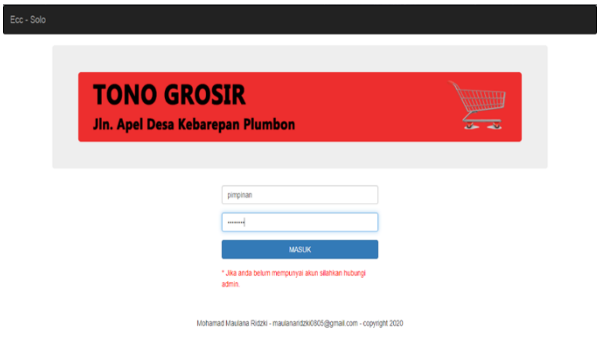


Figure 1. Login Page

After the user successfully enters, it will be redirected to the main page or dashboard of the website. The display can be seen in the picture as follows.

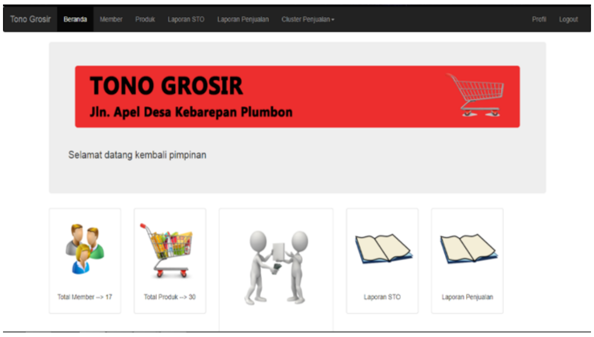
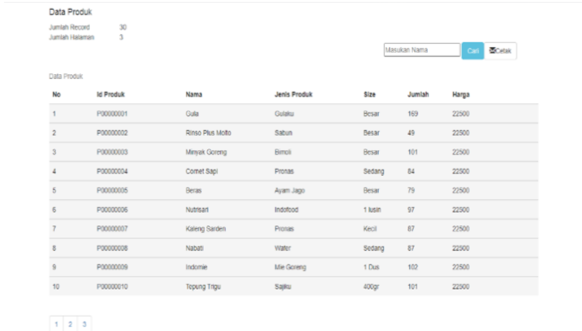


Figure 2. Dashboard Page

The page is about the price of the product, the product data, the number of products and the date of the product. The product data view can be seen as follows.



Data Produk

Jumlah Record 30
Jumlah Halaman 3

Masukan Nama | Cari | Cetak

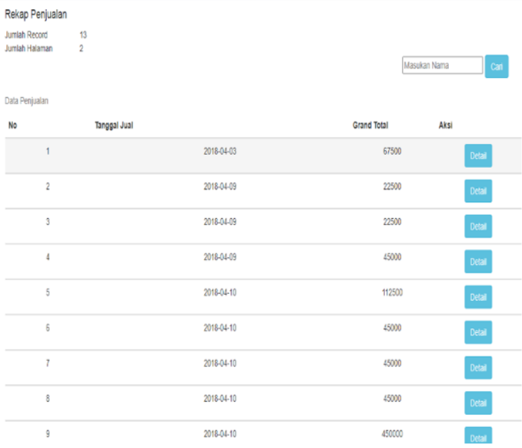
Data Produk

No	Id Produk	Nama	Jenis Produk	Size	Jumlah	Harga
1	P00000001	Gula	Gulaku	Besar	169	22500
2	P00000002	Rinso Plus Molto	Sabun	Besar	49	22500
3	P00000003	Minyak Goreng	Bimoli	Besar	101	22500
4	P00000004	Cornet Sapi	Pronas	Sedang	84	22500
5	P00000005	Beras	Ayam Jago	Besar	79	22500
6	P00000006	Nutrisari	Indofood	1 lusin	97	22500
7	P00000007	Kaleng Sarden	Pronas	Kecil	87	22500
8	P00000008	Nabati	Wafer	Sedang	87	22500
9	P00000009	Indomie	Mie Goreng	1 Dus	102	22500
10	P00000010	Tepung Trigu	Sajiku	400gr	101	22500

1 2 3

Figure 3. Product Data Page

To be able to see the details of the report, you must enter the sales report page. The sales report table contains data on the name of the product of the item, the data on the number of items that have been sold, it looks as follows.



Rekap Penjualan

Jumlah Record 13
Jumlah Halaman 2

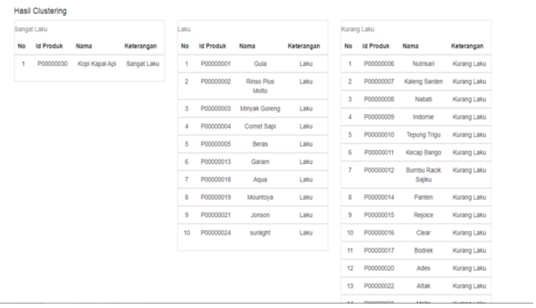
Masukan Nama | Cari

Data Penjualan

No	Tanggal Jual	Grand Total	Aksi
1	2018-04-03	67500	Detail
2	2018-04-09	22500	Detail
3	2018-04-09	22500	Detail
4	2018-04-09	45000	Detail
5	2018-04-10	112500	Detail
6	2018-04-10	45000	Detail
7	2018-04-10	45000	Detail
8	2018-04-10	45000	Detail
9	2018-04-10	450000	Detail

Figure 4. Sales Report page.

For sales cluster results, the store owner must enter on the sales cluster menu. The sales cluster contains the best-selling, best-selling and less-selling products in the following appearance.



Hasil Clustering

Sangat Laku

No	Id Produk	Nama	Keterangan
1	P00000030	Kopi Kapal Api	Sangat Laku

Laku

No	Id Produk	Nama	Keterangan
1	P00000001	Gula	Laku
2	P00000002	Rinso Plus Molto	Laku
3	P00000003	Minyak Goreng	Laku
4	P00000004	Comet Sapi	Laku
5	P00000005	Beras	Laku
6	P00000013	Garam	Laku
7	P00000018	Aqua	Laku
8	P00000019	Mountoya	Laku
9	P00000021	Jonson	Laku
10	P00000024	sunlight	Laku

Kurang Laku

No	Id Produk	Nama	Keterangan
1	P00000006	Nutrisari	Kurang Laku
2	P00000007	Kaleng Sarden	Kurang Laku
3	P00000008	Nabati	Kurang Laku
4	P00000009	Indomie	Kurang Laku
5	P00000010	Tepung Trigu	Kurang Laku
6	P00000011	Kecap Bango	Kurang Laku
7	P00000012	Bumbu Racik Sajiku	Kurang Laku
8	P00000014	Panten	Kurang Laku
9	P00000015	Rejoice	Kurang Laku
10	P00000016	Clear	Kurang Laku
11	P00000017	Bodrek	Kurang Laku
12	P00000020	Ades	Kurang Laku
13	P00000022	Attak	Kurang Laku

Figure 5. Sales clustering figures.

System Testing

Table 8. Login Testing.

Test Cases and Results (Normal Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**
Username: admin Password: admin	Admin is listed in the username field, password is listed in the password field	Admin is listed in the username field, password is listed in the password field	Accepted
Press the Login button	User data is searched in the user table, if available, go to the home page	The login button may work properly	Accepted
Test Cases and Results (Incorrect Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**
Username: admin Password: admin	Admin is listed in the email field, testing is listed in the password field	Admin is listed in the mail field, ****** is listed in the password field	Accepted
Press the Login button	User data not found in user table, failed login and displays error	Login failed and displays error information	Accepted

From the results of functionality tests conducted on the system, several conclusions can be drawn regarding two different use cases. First, the system behaves as expected in use cases with normal data, where usernames and passwords are as expected. The login process was successful, with admin listed in the username field and password listed in the password field. The login button works well, allowing users to enter the home page without problems. Second, the system responds well to input errors in use cases with incorrect data. Despite a typo error in the field, the system can recognize and respond appropriately. In this case, even though the admin is listed in the email field and testing in the password field, the system still provides clear error information to the user. In addition, when user data is not found in the database, the system successfully resolves the situation by displaying the appropriate error message. Overall, the results of this test show that the system has been successfully implemented, can manage inputs effectively, and provides appropriate responses in normal situations and errors.

Table 9. Product Data Testing

Test Cases and Results (Normal Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**

Product Data	Data is saved to the database and can be managed again	Product data is inputted completely and is in accordance with the provisions	Accepted
Press the save button	The save button is available and the save data function can be used	Information appears that the data was successfully saved	Accepted
Test Cases and Results (Incorrect Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**
Product data	Data is saved to the database and can be managed again	Admin is listed in the mail field, ****** is listed in the password field	Accepted
Press the save button	The save button is available and the save data function can be used	Login failed and displays error information	Accepted

From the results of the product data management system functionality test, positive conclusions can be drawn regarding two test scenarios, namely cases with normal data and cases with incorrect data. In tests with normal data, product data input runs smoothly, and the data entered is in accordance with applicable regulations. The process of using the save button is also successful, with the system providing information that the data has been successfully saved into the database. This positive response indicates that the system can manage the data well, provide the expected results, and allow data management again.

On the other hand, in tests with incorrect data, the system also showed satisfactory performance. Even if there is an error in the use of the field, the system can still recognize and respond to the situation properly. The use of inappropriate fields, such as admin in the email field and certain characters in the password field, does not stop the process of storing product data in the database. The save button still works, and even if an error occurs in the login, the system provides clear error information after pressing the save button.

In general, the test results show that the product data management system has been successfully designed and implemented well. Its ability to manage normal data input and respond to error situations provides confidence that the system is reliable for effective and efficient management of product data. This conclusion supports the success of the system in meeting user expectations and needs in terms of product data management and storage.

Table 10. Transaction Data Testing

Test Cases and Results (Normal Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**
Product Data	Data is saved to the database and can be managed again	Product transaction data is inputted completely and in accordance with the provisions	Accepted

Pressing the transaction button	The process button is available and the transaction data process function can be used	Information appears that the transaction data is in process	Accepted
Test Cases and Results (Incorrect Data)			
Input Data	**What to expect**	**Observation**	**Conclusion**
Product data	Data can be processed and according to the provisions	Data is not available and does not comply with the conditions	Accepted
Press the process button	The process button is available and the save data function can be used	Information appears that the data is incomplete and not in accordance with the provisions	Accepted

From the results of system functionality tests on use cases with normal data, it can be concluded that the system can manage product and transaction data well. The entered product data is successfully saved into the database and can be re-managed as needed. Next, the transaction button works fine, and after pressing it, information appears stating that the transaction data is being processed. This test concludes that the system can manage product transactions effectively, provide appropriate responses, and ensure that data is inputted completely and according to the provisions.

The system also responds well to error situations in use cases with incorrect data. The process button is still available and working even if product data is unavailable and does not comply with the conditions. After pressing the process button, information appears stating that the data is incomplete and not in accordance with the provisions. The conclusion that can be drawn is that the system can recognize data discrepancies and provide clear information to users. Even if there is an error, the system can still continue the process and guide users to complete the data in accordance with applicable regulations.

The test results show that the system performs satisfactorily in managing product and transaction data. Its ability to handle normal and incorrect data reflects a solid design and functionality that matches user expectations. This conclusion indicates that the system can reliably support the product transaction process efficiently, provide a good user experience, and ensure compliance with applicable Top.

CONCLUSION

In this study, after going through the system design and implementation stages, the k-means algorithm successfully reached the 4th iteration, forming three clusters with details of the amount of data in each cluster. The test run shows data consistency at the initial and final iterations, certifying that the data clustering process has been completed. After implementation, this application proved successful in helping XYZ Wholesale stores identify very salable, salable, and undersold products. The information generated by the application provides a valuable strategic foundation, facilitates product stock management, improves management efficiency, and provides guidance for more optimal strategic decision-making.

REFERENCES

Baihaqi, W. M., Indartono, K., & Banat, S. (2019). Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee. *Paradigma - Jurnal Komputer Dan Informatika*, *21*(2), 243–248. https://doi.org/10.31294/p.v21i2.6149

Chen, L., Shan, W., & Liu, P. (2021). Identification of concrete aggregates using K-means clustering and level set method. *Structures*, *34*, 2069–2076. https://doi.org/10.1016/j.istruc.2021.08.048

Fauzi, A. (2017). *Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore*.

Griffin, E. C., Keskin, B. B., & Allaway, A. W. (2023). Clustering retail stores for inventory transshipment. *European Journal of Operational Research*, *311*(2), 690–707. https://doi.org/10.1016/J.EJOR.2023.06.008

Hadi, F., & Diana, Y. (2020). PENGKLUSTERAN PENJUALAN BAHAN BANGUNAN MENGGUNAKAN ALGORITMA K-MEANS. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, *4*(1), 22. https://doi.org/10.35145/joisie.v4i1.629

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/J.INS.2022.11.139

Jabat, J. T., & Murdani, M. (2019). Penerapan Data Mining Pada Penjualan Produk Retail Menggunakan Metode Clustering. *Pelita Informatika: Informasi Dan Informatika*, *8*(1), 26–32.

Kashef, R., & Pun, H. (2022). Predicting l-CrossSold products using connected components: A clustering-based recommendation system. *Electronic Commerce Research and Applications*, *53*. https://doi.org/10.1016/J.ELERAP.2022.101148

Khedmati, M., & Azin, P. (2020). An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications*, *159*. https://doi.org/10.1016/J.ESWA.2020.113546

Kuo, R. J., Rakhmat Setiawan, M., & Nguyen, T. P. Q. (2022). Sequential clustering and classification using deep learning technique and multi-objective sine-cosine algorithm. *Computers and Industrial Engineering*, *173*. https://doi.org/10.1016/J.CIE.2022.108695

Luu, E., Xu, F., & Zheng, L. (2023). Short-selling activities in the time of COVID-19. *British Accounting Review*, *55*(4). https://doi.org/10.1016/J.BAR.2023.101216

Meng, Q., Huang, H., Li, X., & Wang, S. (2023). Short-selling and corporate default risk: Evidence from China. *International Review of Economics and Finance*, *87*, 398–417. https://doi.org/10.1016/J.IREF.2023.04.026

Miao, Y., Li, S., Wang, L., Li, H., Qiu, R., & Zhang, M. (2023). A single plant segmentation method of maize point cloud based on Euclidean clustering and K-means clustering. *Computers and Electronics in Agriculture*, *210*. https://doi.org/10.1016/j.compag.2023.107951

Mudzakkir, B. D. (2018). Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *2*(2), 34–40.

Nasution, Y. R., & Eka, M. (2018). Penerapan Algoritma K-Means Clustering Pada Aplikasi Menentukan Berat Badan Ideal. *ALGORITMA: Jurnal Ilmu Komputer Dan Informatika*, *2*(1).

Niu, G., Ji, Y., Zhang, Z., Wang, W., Chen, J., & Yu, P. (2021). Clustering analysis of typical scenarios of island power supply system by using cohesive hierarchical clustering based K-Means clustering method. *Energy Reports*, *7*, 250–256. https://doi.org/10.1016/j.egyr.2021.08.049

Pelekis, S., Pipergias, A., Karakolis, E., Mouzakitis, S., Santori, F., Ghoreishi, M., & Askounis, D. (2023). Targeted demand response for flexible energy communities using clustering techniques. *Sustainable Energy, Grids and Networks*, *36*. https://doi.org/10.1016/J.SEGAN.2023.101134

Purba, W., Tamba, S., & Saragih, J. (2018). The effect of mining data k-means clustering toward students profile model drop out potential. *Journal of Physics: Conference Series*, *1007*, 012049. https://doi.org/10.1088/1742-6596/1007/1/012049

Sinan, M., Leng, J., Shah, K., & Abdeljawad, T. (2023). Advances in numerical simulation with a clustering method based on K–means algorithm and Adams Bashforth scheme for fractional order laser chaotic system. *Alexandria Engineering Journal*, *75*, 165–179. https://doi.org/10.1016/j.aej.2023.05.080

van der Borgh, M., Nijssen, E. J., & Schepers, J. J. L. (2023). Unleash the power of the installed base: Identifying cross-selling opportunities from solution offerings. *Industrial Marketing Management*, *108*, 122–133. https://doi.org/10.1016/J.INDMARMAN.2022.11.010

Vásquez Sáenz, J., Quiroga, F. M., & Bariviera, A. F. (2023). Data vs. information: Using clustering techniques to enhance stock returns forecasting. *International Review of Financial Analysis*, *88*. https://doi.org/10.1016/J.IRFA.2023.102657

Wang, X., Shao, Z., Shen, Y., & He, Y. (2023). Research on fast marking method for indicator diagram of pumping well based on K-means clustering. *Heliyon*, *9*(10). https://doi.org/10.1016/j.heliyon.2023.e20468

Windarto, A. P. (2017). Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *International Journal of Artificial Intelligence Research*, *1*(2), 26. https://doi.org/10.29099/ijair.v1i2.17

Zhou, Q., & Sun, B. (2023). Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem. *Data and Information Management*, 100064. https://doi.org/10.1016/j.dim.2023.100064