International Journal Software Engineering and Computer Science (IJSECS)
4 (3), 2024, 965-974
Published Online December 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i3.3026.

RESEARCH ARTICLE Open Access

Analysis of Scooter Spare Parts Sales at Harapan Indah Scooter Using the K-Means Algorithm

Frencis Matheos Sarimole
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: aminspawon@gmail.com.

Tracy Olivera Lingga *
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: psucikarang@gmail.com.

Received: August 1, 2024; Accepted: October 20, 2024; Published: December 1, 2024.

Abstract: K-means clustering algorithm has been used in this study to analyze the sales performance of scooter spare parts at Harapan Indah Scooter. By using the K-means method, researchers can classify products into 3 categories according to their sales volume. The purpose of this analysis is to identify patterns in sales data and compare the characteristics of each product group. Researchers can see the output from the previous step shows three clusters: Low, Medium, and High Sales. Associating products with these categories Empowers improved tracking of sales movements and fluctuation trends in product options. The findings of this study can be useful in the field of inventory management and to develop marketing strategies to increase product sales. Companies can find out which products fall into which categories and therefore can make better decisions on how to manage stock and promotional efforts. These findings are the first step to maintain and improve sales performance and optimize Harapan Indah Scooter business.

Keywords: K-Means; Analysis; Spare Parts; Scooter; Sales; Clustering; Marketing Strategy.

1. Introduction

The use of scooters as a means of transportation has grown rapidly in many large cities, primarily due to their efficiency in tackling traffic congestion and their lower operational costs compared to cars. Scooters also offer greater flexibility for daily commuting, especially in densely populated urban areas. As the number of scooter users increases, the demand for scooter spare parts has risen accordingly. Stores like Harapan Indah Scooter are now facing challenges in managing their inventory and understanding consumer purchasing patterns to meet this growing demand effectively [1][2]. In today's competitive market, sales data analysis has become essential for businesses. Understanding sales trends and consumer behavior allows stores to make more informed and strategic decisions. Davenport *et al.* (2007) state that data analysis is key to improving business performance and competing effectively. By leveraging data correctly, stores can identify popular products and optimize available stock, helping to improve operational efficiency and customer satisfaction [1]. Moreover, sales analysis can provide valuable insights for designing promotions and offers that align more closely with consumer needs. For instance, segmenting products and sales based on historical data helps in crafting targeted marketing strategies that reach the right audience [5][6]. Using clustering methods like K-Means enables stores to group products by characteristics such as sales volume, facilitating stock and promotion planning. Anggraeni and Handayani (2023) show that K-Means clustering can be applied to ticket sales analysis, which can inform marketing strategies more effectively [7].

Efficient inventory management is a critical aspect of spare parts stores. Poorly managed stock can result in financial losses, either from overstocking, which increases storage costs, or stockouts, which can lead to missed sales opportunities. Chopra and Meindl (2007) emphasize that good inventory management is one of the key factors in the success of retail operations. Stores need an efficient inventory system, supported by accurate sales analysis, to ensure that the right products are available at the right time, avoiding unnecessary resource waste [2]. Effective inventory management also strengthens relationships with suppliers, helping to reduce operational costs in the long term. Applying the right analysis methods, such as K-Means clustering, can also assist stores in identifying sales patterns that might not be immediately obvious through traditional methods. By using this technique, stores can categorize spare parts into different groups based on sales levels, making inventory management and decision-making easier. Yulianti *et al.* (2019) argue that K-Means clustering can be used to understand customer preferences by grouping products based on relevant criteria, such as purchase frequency or sales volume [5].

Prasetya *et al.* (2023) demonstrate that K-Means clustering, combined with the Elbow Method, can help determine the optimal number of clusters for analyzing sales patterns, providing a clearer understanding of consumer behavior and market demand [6]. By applying this method, Harapan Indah Scooter can improve its sales strategy by optimizing inventory levels and targeting promotions more effectively. Understanding which products fall into high, medium, or low sales categories will help the store make more informed stock and promotional decisions. This approach not only aids in stock planning but can also guide pricing and promotional strategies. Pambudi and Witanti (2021) highlight that clustering with K-Means allows stores to map out different levels of product sales and develop more focused marketing campaigns based on product segmentation [11]. Therefore, the use of K-Means in stores like Harapan Indah Scooter can enhance operational efficiency, improve customer satisfaction, and drive sales growth. In conclusion, the application of data analysis techniques such as K-Means clustering offers significant benefits in managing scooter spare part sales and inventory. Through proper analysis, stores can identify sales trends, understand consumer behavior, and design more effective marketing strategies. Consequently, implementing the K-Means algorithm at Harapan Indah Scooter can help optimize operational performance and boost sales in the future [9][10].

By leveraging data analysis methods such as K-Means clustering, stores like Harapan Indah Scooter can improve their ability to understand sales patterns and optimize inventory management. The findings from this approach will not only help in improving stock allocation and sales strategies, but also provide a basis for more informed decision making in a competitive market. Given the background, the purpose of the study is to demonstrate how data-driven strategies can lead to more effective operational practices, greater customer satisfaction, and improved business performance in the spare parts retail sector.

2. Research Method

This study employs a combination of data collection methods to obtain the necessary information for analysis. The first method utilized is observational research, which involves direct observation of the phenomena being studied. In this case, the researcher conducted field observations at Harapan Indah Scooter, located in Bekasi, to examine the sales patterns of scooter spare parts. The second method is interviewing, a qualitative research technique that facilitates direct communication between the researcher and the respondents to obtain in-depth information pertinent to the research topic. The researcher conducted interviews with the owner of Harapan Indah Scooter to gather relevant data for the study. Lastly, the literature review method was employed, drawing on existing academic resources such as books, journal articles, and online publications to collect secondary data that supports the research.

These methods enabled the researcher to gather the required data for subsequent analysis. The research was conducted at Harapan Indah Scooter, located at Setia Asih, Tarumajaya, Bekasi, West Java 17215. The methodological framework for this study is organized into several phases, as follows. The first phase is problem identification, a critical step in recognizing and defining the research problem. In this study, the problem identified was the lack of a clear understanding of the sales performance of scooter spare parts at Harapan Indah Scooter and the difficulty in categorizing spare parts based on consumer demand, ranging from highly preferred to less popular items. The second phase involves data collection, wherein the researcher identifies and compiles the necessary data relevant to the study. Data was directly collected from Harapan Indah Scooter through both observation and interviews with the store owner. The third phase is data analysis, which refers to the systematic process of organizing and examining the data to understand various aspects of the scooter spare parts sales. This phase begins with gathering data from multiple sources, including sales records and stock inventories of scooter spare parts. After the data is collected, it undergoes a process of cleaning and preparation for further analysis. The fourth phase, data processing, involves the preparation of the data for analysis using the K-Means clustering technique. This includes the removal of incomplete or irrelevant data and the organization of the data for clustering analysis. The cleaned data was then processed using K-Means clustering in Microsoft Excel to classify products based on sales performance, distinguishing between low and high-selling items. The final phase is data testing, an essential step to evaluate the accuracy and applicability of the processed data. In this study, data testing was conducted using RapidMiner, a widely recognized data analysis platform, which provides a suite of tools for data processing, analysis, and testing.

3. Result and Discussion

3.1 Results

In this study, the sales of scooter spare parts were analyzed using RapidMiner software for data processing, alongside Microsoft Excel with the Elbow Method to determine the optimal number of clusters (k) for the analysis. The K-means clustering algorithm was employed to classify the data into distinct groups based on sales performance. The first step in the process involved determining the value of k, which represents the number of clusters to be formed. Initial centroids were selected randomly from the available data. The next step was to calculate the Euclidean distance between each data point and the centroids in order to assign each data point to the closest centroid. After classifying the data, the centroids were recalculated as the mean of all data points assigned to each cluster. This process was repeated until the centroids stabilized, indicating that the clustering process was complete. To determine the optimal number of clusters, the Elbow Method was used. This method calculates the average within-centroid distance for various values of k. The results, shown in Table 1, indicate that when k=3, the within-centroid distance sharply decreased and then stabilized, confirming that three clusters was the optimal choice for the data. The values in Table 1 reveal the following:

Table 1. Average Within-Centroid Distance

k	Avg. Within Centroid Distance
2	799.166
3	236.426
4	118.788
5	60.994
6	36.361
7	21.883

The graph in Figure 1 illustrates this "elbow" point where the curve begins to flatten, further confirming that k=3 is the most appropriate number of clusters to form.

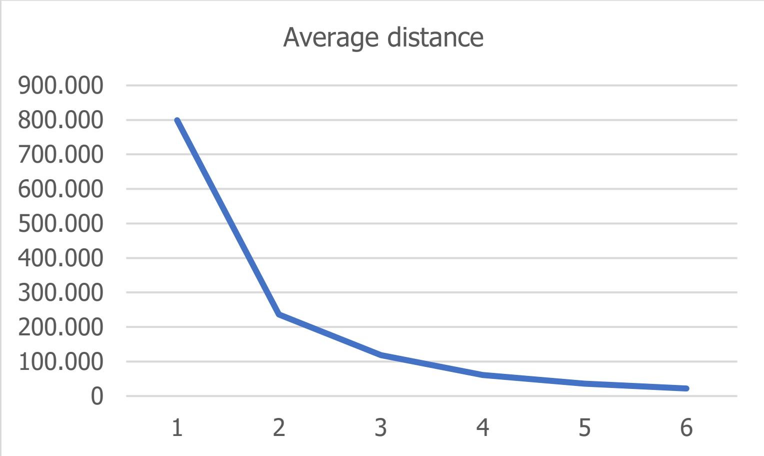

Figure 1. Elbow Method

Based on the graph above, the best number of clusters to form is a cluster with a value of k = 3, because the elbow of the graph is at point 2, namely the cluster with a value of k = 3 with an Avg dist value of 236,426. The data obtained will be calculated first, the existing data is the number of transactions and sales of scooter spare part products as shown in the following figure.

No.	Nama Produk	Jumlah Transaksi	Penjualan Produk
1.	Baterai 18650 LG M26	65	340
2.	Baterai Pack Lithium-ion 36v 7Ah	25	25
3.	Baterai Pack Lithium-ion 48v 20Ah	5	5
4.	Baterai Pack Lithium-ion 48v 10Ah	2	2
5.	Baterai Pack Lithium-ion 36v 15Ah	45	65
6.	Plat Nikel 18650 1 Meter	20	50
7.	Controller 36v 350 Watt	35	38
8.	Controller 48v 1500 Watt	3	3
9.	Handle Gas Scooter	30	30
10.	Voltmeter 0-100v	20	20
11.	Charger Lithium-ion 36v 2a	5	10
12.	Charger Lithium-ion 48v 3a	5	5
13.	BLDC 8 inch 36v 300 Watt	10	10
14.	BLDC 10 Inch 36v 250 Watt	2	2
15.	Handle Rem Sensor	10	20
16.	Kaliper Rem Scooter	5	10
17.	Switch 3 Speed	3	5
18.	Kunci Kontak	10	15
19.	Sadel Jok Scooter	2	2
20.	Reducer Step Down 12v	10	13
21.	Remote Alarm	5	6
22.	Lampu Depan Scooter	20	25
23.	Lampu Belakang Scooter	22	22
24.	Klakson 12v	3	3
25.	Helm Scooter	10	10
26.	Jok Tengah Scooter	7	7
27.	Konektor XT60	80	100
28.	Konektor XT30	50	70
29.	Kabel AWG 12 Per Meter	25	50
30.	Kabel AWG 16 Per Meter	32	63
31.	Ban Dalam Scooter 10 x 2.5"	11	11
32.	BMS 36v 15A	20	25
33.	BMS 48v 60A	10	10
34.	BMS 12v 40A	3	3
35.	BMS 12v 20A	5	15

Figure 2. Scooter Spare Parts Data

After determining the optimal k, the next step was to classify the data into three clusters: low sales, medium sales, and high sales. Initial centroids were chosen randomly from the dataset. For Cluster 0, the centroid was represented by the data "Battery Pack Lithium-ion 36V 7Ah," for Cluster 1, it was "Nickel Plate 18650 1 Meter," and for Cluster 2, it was "Charger Lithium-ion 36V 2A."

Table 2. Initial Centroid Determination

Centroid	TS	PP
C0	25	25
C1	20	50
C2	5	10

Subsequently, the Euclidean distance between each data point and the centroids was calculated, and the data was grouped accordingly. The centroids were recalculated based on the newly assigned data points.

Nama Produk	TS	PP	C0	C1	C2	Jarak Terdekat	Cluster
Baterai 18650 LG M26	64	340	316,04059	235,32817	337,075097	235,32817	C1
Baterai Pack Lithium-ion 36v 7Ah	25	25	1,4914718	82,91254163	25,5268743	1,491471825	C0
Baterai Pack Lithium-ion 48v 20Ah	5	5	29,021104	108,1542724	3,11713868	3,117138683	C2
Baterai Pack Lithium-ion 48v 10Ah	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Baterai Pack Lithium-ion 36v 15Ah	45	65	43,647241	40,4288224	69,1044921	40,4288224	C1
Plat Nikel 18650 1 Meter	20	50	24,010627	60,86451865	44,3170652	24,0106274	C1
Controller 36v 350 Watt	35	38	15,577325	68,18716595	41,7595427	15,57732526	C0
Controller 48v 1500 Watt	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
Handle Gas Scooter	30	30	6,4980373	76,93357713	32,5898374	6,498037258	C0
Voltmeter 0-100v	20	20	7,88825	89,0516919	18,4701143	7,888250009	C0
Charger Lithium-ion 36v 2a	5	10	25,552667	103,5280684	2,27875922	2,278759217	C2
Charger Lithium-ion 48v 3a	5	5	29,021104	108,1542724	3,11713868	3,117138683	C2
BLDC 8 inch 36v 300 Watt	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
BLDC 10 Inch 36v 250 Watt	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Handle Rem Sensor	10	20	15,92649	92,37456885	12,6942949	12,69429492	C2
Kaliper Rem Scooter	5	10	25,552667	103,5280684	2,27875922	2,278759217	C2
Switch 3 Speed	3	5	30,405759	108,912433	4,20910365	4,209103654	C2
Kunci Kontak	10	15	18,518606	97,01724934	8,10363706	8,103637058	C2
Sadel Jok Scooter	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Reducer Step Down 12v	10	13	19,815475	98,88407835	6,44037713	6,440377129	C2
Remote Alarm	5	6	28,290764	107,226347	2,19357962	2,193579625	C2
Lampu Depan Scooter	20	25	4,7894445	84,26694566	22,0594949	4,789444457	C0
Lampu Belakang Scooter	22	22	5,1209851	86,57897044	21,2916791	5,120985081	C0
Klakson 12v	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
Helm Scooter	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
Jok Tengah Scooter	7	7	26,195887	105,5606979	1,38095241	1,380952414	C2
Konektor XT60	80	100	92,114503	35,27732686	118,104886	35,27732686	C1
Konektor XT30	50	70	50,448802	35,75996938	76,0651499	35,75996938	C1
Kabel AWG 12 Per Meter	25	50	23,575325	58,97508847	46,1411125	23,57532477	C1
Kabel AWG 16 Per Meter	32	63	37,318267	44,41754431	60,8788991	37,31826746	C1
Ban Dalam Scooter 10 x 2.5"	11	11	20,548102	100,4116005	5,85559404	5,855594041	C2
BMS 36v 15A	20	25	4,7894445	84,26694566	22,0594949	4,789444457	C0
BMS 48v 60A	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
BMS 12v 40A	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
BMS 12v 20A	5	15	22,663916	98,93823953	7,1182114	7,118211402	C2

Figure 3. Distance Calculation and Data Grouping for the 1st Iteration

After the first iteration, the new centroid values were obtained, as shown in Table 3:

Table 3. New Centroid Values

Centroid	TS	PP
C0	24.57	26.43
C1	45.14	105.43
C2	6.00	7.95

The new centroids were then used for the second iteration, and since the centroids did not change further, the process was stopped. The results of these calculations can be seen in Table 3 (first iteration) and Table 4.6 (second iteration), indicating that the clustering process had converged.

Nama Produk	TS	PP	C0	C1	C2	Jarak Terdekat	Cluster
Baterai 18650 LG M26	64	340	316,04059	235,32817	337,075097	235,32817	C1
Baterai Pack Lithium-ion 36v 7Ah	25	25	1,4914718	82,91254163	25,5268743	1,491471825	C0
Baterai Pack Lithium-ion 48v 20Ah	5	5	29,021104	108,1542724	3,11713868	3,117138683	C2
Baterai Pack Lithium-ion 48v 10Ah	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Baterai Pack Lithium-ion 36v 15Ah	45	65	43,647241	40,4288224	69,1044921	40,4288224	C1
Plat Nikel 18650 1 Meter	20	50	24,010627	60,86451865	44,3170652	24,0106274	C1
Controller 36v 350 Watt	35	38	15,577325	68,18716595	41,7595427	15,57732526	C0
Controller 48v 1500 Watt	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
Handle Gas Scooter	30	30	6,4980373	76,93357713	32,5898374	6,498037258	C0
Voltmeter 0-100v	20	20	7,88825	89,0516919	18,4701143	7,888250009	C0
Charger Lithium-ion 36v 2a	5	10	25,552667	103,5280684	2,27875922	2,278759217	C2
Charger Lithium-ion 48v 3a	5	5	29,021104	108,1542724	3,11713868	3,117138683	C2
BLDC 8 inch 36v 300 Watt	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
BLDC 10 Inch 36v 250 Watt	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Handle Rem Sensor	10	20	15,92649	92,37456885	12,6942949	12,69429492	C2
Kaliper Rem Scooter	5	10	25,552667	103,5280684	2,27875922	2,278759217	C2
Switch 3 Speed	3	5	30,405759	108,912433	4,20910365	4,209103654	C2
Kunci Kontak	10	15	18,518606	97,01724934	8,10363706	8,103637058	C2
Sadel Jok Scooter	2	2	33,259953	112,0659413	7,17152979	7,171529793	C2
Reducer Step Down 12v	10	13	19,815475	98,88407835	6,44037713	6,440377129	C2
Remote Alarm	5	6	28,290764	107,226347	2,19357962	2,193579625	C2
Lampu Depan Scooter	20	25	4,7894445	84,26694566	22,0594949	4,789444457	C0
Lampu Belakang Scooter	22	22	5,1209851	86,57897044	21,2916791	5,120985081	C0
Klakson 12v	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
Helm Scooter	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
Jok Tengah Scooter	7	7	26,195887	105,5606979	1,38095241	1,380952414	C2
Konektor XT60	80	100	92,114503	35,27732686	118,104886	35,27732686	C1
Konektor XT30	50	70	50,448802	35,75996938	76,0651499	35,75996938	C1
Kabel AWG 12 Per Meter	25	50	23,575325	58,97508847	46,1411125	23,57532477	C1
Kabel AWG 16 Per Meter	32	63	37,318267	44,41754431	60,8788991	37,31826746	C1
Ban Dalam Scooter 10 x 2.5"	11	11	20,548102	100,4116005	5,85559404	5,855594041	C2
BMS 36v 15A	20	25	4,7894445	84,26694566	22,0594949	4,789444457	C0
BMS 48v 60A	10	10	21,95961	101,6938168	4,49363367	4,493633671	C2
BMS 12v 40A	3	3	31,846891	110,7593443	5,79017077	5,790170772	C2
BMS 12v 20A	5	15	22,663916	98,93823953	7,1182114	7,118211402	C2

Figure 4. Distance Calculations and Data Grouping - Iteration 2

To validate the results, the data was tested using RapidMiner Studio Version 10.1, a software platform that allows for comparison between manually processed data and results generated by the software. The process was initiated by selecting "Start" and creating a new "Blank Process" in the software, as shown in Figure 5. After that, the data was imported in CSV format using the "Read CSV" operator, as illustrated in Figure 5.

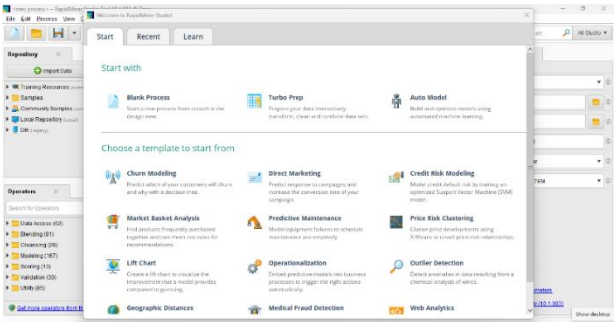

Figure 5. Main Page of RapidMiner Studio

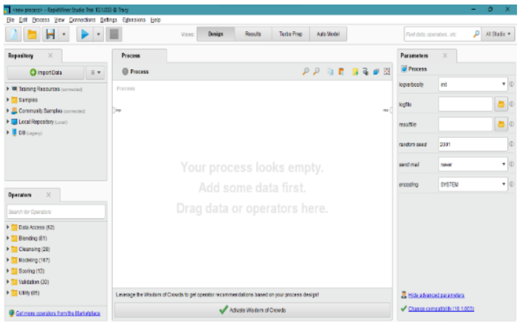

Figure 6. Blank Process in RapidMiner Studio v.10.1

After uploading the data, the "K-Means" operator was applied to perform the clustering, with the number of clusters set to 3. The clustering process was connected to the "Result" node to display the output, as shown in Figure 7. The results of the clustering were displayed in the "Example Set" window, which offers various views and visualizations, such as Data View, Statistics View, and Visualizations. The graphical representation of the clusters can be seen in Figure 9, where the clustering results for three groups were shown clearly.

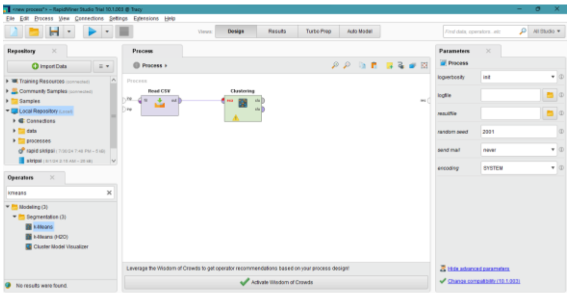

Figure 7. Data Clustering Process in RapidMiner

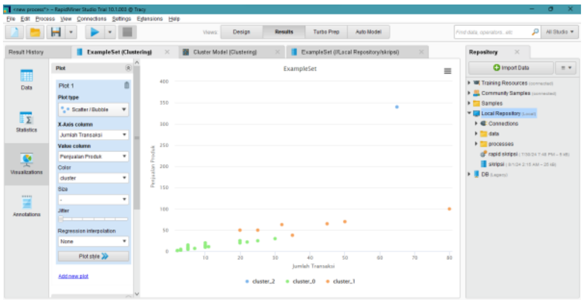

Figure 8. Cluster Visualization in RapidMiner

The final clustering results revealed three groups, each representing products with different sales performance. Cluster 0 contained 27 items, representing low sales, Cluster 1 had 7 items for medium sales, and Cluster 2 contained just 1 item for high sales. Products in Cluster 0 included items such as the "Battery Pack Lithium-ion 36V 7Ah," "Controller 48V 1500W," and "Charger 36V 2A." Cluster 1 included items like the "Battery Pack Lithium-ion 36V 15Ah" and "Nickel Plate 18650 1 Meter," while Cluster 2 consisted of a single high-selling item, the "Battery 18650 LG M26." These results are summarized in Figure 9, providing insight into the sales distribution of scooter spare parts.

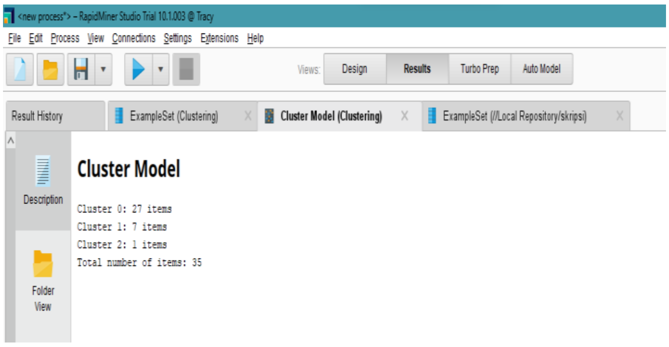

Figure 9. Distribution of Products in Each Cluster

The clustering analysis successfully identified three distinct groups based on sales volume. These groups allow for a clearer understanding of product demand and provide valuable information for inventory management, pricing strategies, and targeted marketing efforts.

3.2 Discussion

The general framework configuration represents the final design of the system to be developed. Below is the The use of the Elbow Method to determine the optimal number of clusters was a critical step in the analysis. The method revealed that the best choice for k was 3, as shown in Figure 1, where the decrease in the average within-centroid distance slowed significantly after k=3. This indicates that three clusters were sufficient to represent the various sales levels, and further increasing the number of clusters would not yield substantial improvements in the grouping. The results confirm that the Elbow Method is an efficient technique for identifying the appropriate number of clusters in a dataset like the one used in this study.

After determining the optimal number of clusters, the K-means algorithm successfully grouped the scooter spare parts into three clusters based on their sales performance. Cluster 0 consisted of products with low sales, such as the "Battery Pack Lithium-ion 36V 7Ah," "Controller 48V 1500W," and other low-demand items. Cluster 1 contained products with medium sales, including items like the "Battery Pack Lithium-ion 36V 15Ah" and "Nickel Plate 18650 1 Meter." Lastly, Cluster 2, with only one item, the "Battery 18650 LG M26," represented high-sales products. The grouping of products into these three categories allowed for a clearer understanding of sales performance and provided actionable insights for managing inventory more efficiently.

The distribution of products into these clusters aligns with the expectations of a retail business, where some products naturally perform better than others. Cluster 0, which includes 27 items, represents products with low demand. These items are likely to accumulate in inventory and could be at risk of obsolescence if not managed carefully. Managing these low-sales items effectively is essential to prevent overstocking and minimize storage costs. Retailers may consider discounting, bundling, or promoting these products to boost their sales.

Cluster 1, which includes 7 items, represents products with medium sales. These products have a moderate turnover and could benefit from targeted promotional strategies to increase their sales volume. Identifying the characteristics of products in Cluster 1—such as price, features, or customer demand—could help refine marketing strategies to boost their performance further. Additionally, better demand forecasting can assist in optimizing stock levels for these items, ensuring that they are available without excessive overstock. Cluster 2, which contains just one product, the "Battery 18650 LG M26," represents high-sales items. These products are the most popular and generate significant revenue. Understanding the factors that contribute to the high sales of these items—such as superior quality, better customer preference, or competitive pricing—can help businesses replicate the success of these items with other products. For example, enhancing the visibility of high-sales products in marketing campaigns or increasing stock availability can drive even more sales. Furthermore, promoting complementary products to these high-demand items may create cross-selling opportunities.

The results of this analysis offer several practical applications for inventory management. By classifying products into low, medium, and high sales clusters, businesses can manage stock more effectively, ensuring that they focus on maintaining adequate stock levels for high-sales products, avoid overstocking low-demand items, and adjust marketing strategies for medium-sales products. Additionally, these insights can help businesses plan better promotional activities, such as discounts or special offers, for products in the low-sales group, or focus on bundling for medium-sales items to enhance their appeal.

Furthermore, the use of RapidMiner Studio for testing the clustering results validated the efficiency of the manual data processing approach. The software's ability to quickly generate results and visualize clusters allowed for a comparison between manual and automated clustering methods, confirming the accuracy and reliability of the findings. The software's capabilities in visualizing cluster distribution, such as in Figure 9, further support the results by clearly presenting the differences in product performance based on sales. The clustering results from this study offer actionable insights into how scooter spare parts are performing in the market, which can directly inform decisions on inventory management, sales strategies, and marketing. By applying the K-means algorithm and the Elbow Method, businesses can identify underperforming products, optimize stock levels, and design targeted promotions to improve sales across different product categories. This method can be further applied to other sectors in retail or inventory management to improve operational efficiency and profitability.

4. Related Work

This study applies the K-Means clustering algorithm to analyze the sales of scooter spare parts, which shares similarities with several other studies that have used clustering techniques for sales analysis across different industries. Pambudi and Witanti (2021) utilized K-Means clustering to analyze sales data for Ayu Collection, an online store, grouping products based on sales performance to optimize inventory management and sales strategies. This approach mirrors the current research, where spare parts are similarly categorized into low, medium, and high-sales groups, helping improve inventory control and operational efficiency [11]. Similarly, Sharyanto and Lestari (2022) employed K-Means clustering in e-commerce for customer segmentation using the RFM (Recency, Frequency, Monetary) model, focusing on consumer purchase patterns. While their work segments customers, this study focuses on segmenting products based on sales volume [12].

Surapati and Jannah (2024) also applied K-Means clustering to analyze customer interests in purchasing K-pop merchandise. This study used clustering to identify customer segments and adjust marketing strategies accordingly, a concept somewhat similar to segmenting products based on sales to enhance product promotions and demand forecasting [13]. Triyandana *et al.* (2022) employed K-Means for grouping food and beverage menu items by sales volume, optimizing product offerings, and managing inventory. This parallels the approach in this study, where spare parts are classified to optimize stock levels and improve sales performance [14]. Beyond retail product analysis, K-Means clustering has been applied in other sectors, such as education and telecommunications. Sulistiyawati and Supriyanto (2021) used K-Means to group students based on academic performance, identifying those in need of additional resources [15]. In telecommunications, Handoko *et al.* (2020) applied K-Means clustering to assess the sales performance of Telkomsel data packages, identifying high and low-performing packages to optimize sales strategies [17]. This is similar to the current study, where spare parts are grouped to identify top performers and slow-moving items.

K-Means clustering has also been used in the automotive sector, specifically for analyzing motorbike spare parts sales. Hutabarat and Sindar (2019) applied K-Means to categorize motorcycle spare parts based on sales, enabling better stock management by identifying low-demand products that could be discounted or promoted. This study follows a similar pattern, analyzing scooter spare parts to improve inventory control and sales efficiency [18]. Darmi and Setiawan (2017) also employed K-Means for product classification, enabling businesses to identify slow-moving products and optimize stock [19]. Some studies have used K-Means clustering in customer relationship management (CRM). Montero (2022) applied K-Means to group customers based on their purchasing behavior, aiming to optimize CRM strategies and improve customer retention [20]. Ridzki (2023) used K-Means to classify best-selling products in grocery stores, helping businesses focus on high-performance products and adjust their sales strategies accordingly [21]. The integration of K-Means with RFM models in some studies, such as the one by the Primskystore team (2023), also mirrors the way K-Means has been applied to product-level sales analysis, enhancing decision-making and sales management [22].

The primary difference between this study and previous research lies in the specific focus on scooter spare parts as the product category. While many studies like Hutabarat and Sindar (2019) focused on motorbike spare parts, or Handoko *et al.* (2020) on data packages [17], this research examines a niche market within the automotive sector, targeting a specific store, Harapan Indah Scooter, to analyze its sales performance. Furthermore, unlike Sharyanto and Lestari (2022), this study does not employ customer segmentation based on RFM data, focusing solely on product sales and inventory optimization instead of customer behavior [12]. This study shares several similarities with previous research in applying K-Means clustering to sales data, but it also offers a fresh perspective by focusing on a specific product within a specialized market. By targeting scooter spare parts, the methodology employed here adds to the broader use of clustering techniques in retail and inventory management.

5. Conclusion

Based on the results of data analysis and processing, it can be concluded that the K-Means algorithm has proven effective in identifying sales patterns of scooter spare part products. By using this method, researchers have succeeded in grouping products based on their sales levels. Three sales groups were found after the data was analyzed, namely Cluster 0 for products with low sales, Cluster 1 for products with medium sales, and Cluster 2 for products with high sales. Cluster 0 consists of 27 items, Cluster 1 consists of 7 items, and Cluster 2 consists of only 1 item. Cluster 0 includes products with low sales, while Cluster 1 and 2 include products with medium and high sales. The results of this analysis provide a better understanding of the distribution of scooter spare part product sales and provide useful information for Harapan Indah Scooter to increase sales and plan more effective marketing strategies.

References

[1] Davenport, T. H., Harris, J. G., Jones, G. L., Lemon, K. N., Norton, D., & McCallister, M. B. (2007). The dark side of customer analytics. *Harvard Business Review, 85*(5), 37.

[2] Chopra, S., Meindl, P., & Kalra, D. V. (2007). *Supply chain management by Pearson*. Pearson Education India.

[3] Bahtiar, A., & Ali, I. (2023). Transformasi strategi penjualan Batik Cirebon dengan pendekatan analisis pengelompokan K-Means. *KOPERTIP: Scientific Journal of Informatics Management and Computer, 7*(1), 1-7. https://doi.org/10.32485/kopertip.v7i1.309

[4] Ariati, I., Norsa, R. N., Akhsan, L., & Heikal, J. (2023). Segmentasi pelanggan menggunakan K-Means clustering studi kasus pelanggan UHT Milk Greenfield. *Cerdika: Jurnal Ilmiah Indonesia, 3*(7), 729-743. https://doi.org/10.59141/cerdika.v3i7.639

[5] Yulianti, Y., Utami, D. Y., Hikmah, N., & Hasan, F. N. (2019). Penerapan data mining menggunakan algoritma K-Means untuk mengetahui minat customer di toko hijab. *Jurnal Pilar Nusa Mandiri, 15*(2), 241-246. https://doi.org/10.33480/pilar.v15i2.650

[6] Prasetya, A., Salkiawati, R., & Alexander, A. D. (2023). Analisis cluster K-Means dengan metode elbow untuk menentukan pola penjualan produk traffic room Summarecon Mal Bekasi. *Journal of Students' Research in Computer Science, 4*(1), 105-118. https://doi.org/10.31599/5xve0445

[7] Anggraeni, Y., & Handayani, P. (2023). Penerapan metode K-Means untuk menentukan penjualan tiket renang pada Splash Swimming Pool & Gym. *Jurnal Informatika dan Teknologi Informasi, 2*(1), 173-181. https://doi.org/10.56854/jt.v2i1.167

[8] Nurajizah, S., & Salbinda, A. (2021). Penerapan data mining metode K-Means clustering untuk analisa penjualan pada toko fashion hijab Banten. *Jurnal Teknik Komputer AMIK BSI, 7*(2). https://doi.org/10.31294/jtk.v4i2

[9] Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi RapidMiner dengan metode clustering K-Means untuk analisa penjualan pada toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika), 7*(1), 101-104. https://doi.org/10.36040/jati.v7i1.6094

[10] Fahlevi, M. R., Putri, D. R. D., & Syahrin, E. (2023). Analisis pengelompokan data pelelangan barang dengan metode K-Means clustering. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika), 8*(1), 53-61. http://dx.doi.org/10.30645/jurasik.v8i1.541

[11] Pambudi, W. T., & Witanti, A. (2021). Penerapan algoritma K-Means clustering untuk menganalisis penjualan pada toko Ayu Collection berbasis web. *Jurnal Informatika Universitas Pamulang, 6*(3), 645-650. https://dx.doi.org/10.32493/informatika.v6i3.12380

[12] Sharyanto, S., & Lestari, D. (2022). Penerapan data mining untuk menentukan segmentasi pelanggan dengan menggunakan algoritma K-Means dan model RFM pada e-commerce. *JURIKOM (Jurnal Riset Komputer), 9*(4), 866-871. https://doi.org/10.30865/jurikom.v9i4.4525

[13] Surapati, U., & Jannah, M. (2024). Penerapan data mining menggunakan metode K-Means untuk mengetahui minat customer dalam pembelian merchandise Kpop. *Jurnal Sains dan Teknologi, 5*(3), 875-884. https://doi.org/10.55338/saintek.v5i3.2739

[14] Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). Penerapan data mining pengelompokan menu makanan dan minuman berdasarkan tingkat penjualan menggunakan metode K-Means. *Journal of Applied Informatics and Computing, 6*(1), 40-46. https://doi.org/10.30871/jaic.v6i1.3824A

[15] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi algoritma K-Means clustering dalam penetuan siswa kelas unggulan. *Jurnal Tekno Kompak, 15*(2), 25-36. https://doi.org/10.33365/jtk.v15i2.1162

[16] Arvio, Y. (2023). Pemodelan segmentasi transaksi jual beli produk menggunakan pendekatan model K-Means dan subtractive clustering studi kasus survey pada beberapa cabang optik retail. *PETIR, 16*(1). https://doi.org/10.33322/petir.v16i1.1897

[17] Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi data mining untuk menentukan tingkat penjualan paket data Telkomsel menggunakan metode K-Means clustering. *Jurnal Ilmiah Teknologi dan Rekayasa, 25*(1), 76-88. http://dx.doi.org/10.35760/tr.2020.v25i1.2677

[18] Hutabarat, S. M., & Sindar, A. (2019). Data mining penjualan suku cadang sepeda motor menggunakan algoritma K-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), 2*(2), 126-132. https://doi.org/10.32672/jnkti.v2i2.1555

[19] Darmi, Y., & Setiawan, A. (2017). Penerapan metode clustering K-Means dalam pengelompokan penjualan produk. *Jurnal Media Infotama, 12*(2). https://doi.org/10.37676/jmi.v12i2.418

[20] Montero, Z. (2022). Customer grouping for customer relationship management optimization with the K-Means algorithm. *Journal of Computer Science and Information Technology*, 98-105. https://doi.org/10.35134/jcsitech.v8i4.46

[21] Ridzki, M. (2023). K-Means algorithm method for clustering best-selling product data at XYZ grocery stores. *International Journal of Social Service and Research, 3*(12), 3354-3367. https://doi.org/10.46799/ijssr.v3i12.652

[22] Wardhani, R. K. P., & Anjani, A. (2023). Sales Product Clustering Using RFM Calculation Model And K-Means Algorithm on Primskystore. *Technium Romanian Journal of Applied Sciences and Technology, 16*, 176-182. https://doi.org/10.47577/technium.v16i.9978.