JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Optimalization of Grouping Models on Sales Transaction Data in the Josi.Id Store Using the K-Means Algorithm

Resda Dayanti*[1], Rudi Kurniawan[2], Tati Suprapti[3]
[123]STMIK IKMI Cirebon, Indonesia
[1*]resdadayanti524@gmail.com, [2]rudi226ikmi@gmail.com, [3]tatisuprapti112004@gmail.com

Abstract: This study aims to optimize the K-Means algorithm to improve the clustering model of fashion goods sales transaction data at the josi.id store over a period of seven months. One of the main challenges is the lack of understanding of the characteristics of sales transaction data at the josi.id store, as well as the difficulty in identifying products that cause spikes on big days. With the K-Means clustering method used to group data, the optimal K value, attributes that affect the Davies Bouldin index (DBI) value. The analysis of the results shows that the key attribute that affects the k value is the TYPE OF ITEM with K = 3 as the optimal value, has the lowest DBI value of 0.258 compared to other cluster configurations. With the characteristics of cluster 0 (429 items) showing dominant sales during the Eid season. Cluster 1 (343 items) shows high sales during the holiday period. Cluster 2 (309 items) has stable sales during weekdays. These results show good separation and uniformity of clusters in each cluster. The attribute of ITEM TYPE, based on the characteristics of each cluster is Bracket clothes products show the highest total sales of up to 7 million, supported by traffic (love feature) that is often viewed. Blouses have total sales of under 2 million, while dresses show great variation with total sales between 1 and more than 3 million. Skirts have a more diverse sales distribution, with transactions reaching 3 million. which includes categories such as Dresses, bracket clothes, Tops, and Skirts, plays an important role in grouping sales transaction data, especially for seasonal products such as during Eid.
Keywords: K-Means; Davies-Bouldin Index; Data Clustering; Sales Transaction Data; Seasonal Products;

1. INTRODUCING

In the Rapid development of information technology has brought about major changes in various sectors, including business and technology. Clustering techniques in driving sales growth through personalized experiences are one of the important applications for finding hidden patterns in complex data. In the fashion sector, clustering helps analyze sales transactions by grouping customers based on behavior, preferences, and demographics, enabling more personalized and effective marketing strategies [1]. Algorithms such as K-Means and hierarchical clustering support the identification of deep transaction data clusters. Integration of big data with clustering also facilitates trend prediction, dynamic price adjustment, and efficient stock management, which contributes to customer satisfaction and increased sales [2].

Josi.id Store is one of the business actors that sells clothing to date. The problems faced by Josi.id Store are real examples of the challenges faced by many sellers today. One of the main challenges for Josi.id Store is the lack of understanding of the characteristics of sales transaction data, as well as the difficulty in identifying products that cause spikes in demand on big days. As a result, even though the stock of goods is increased in large quantities to achieve more profit, the high volume of stock and demand causes disruptions in the shipping process, so that not all goods can be delivered on time.

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Therefore, applications using the K-Means algorithm are relatively simple and easy to implement [3]. By implementing the K-means algorithm, the Josi.id store can analyze traffic to find out which products are frequently viewed or liked (with the love feature). These products are the ones that need to be prioritized for stock. In this way, obstacles such as unnecessary or excessive stock accumulation can be overcome, so that stock management becomes more efficient and shipping is more organized. In addition, this algorithm also helps predict demand trends based on traffic data, so that stores can be better prepared for spikes in demand on big days without difficulty in shipping.

Previous studies are in line with Rosida et al. on Clustering of HIV/AIDS Disease in West Java Using the K-Means Clustering Algorithm from k-2 to k-20, and it can be seen that the cluster that approaches 0 is k-2, with a DBI value of 0.414 [4]. The research conducted by Hutagalung et al., on Grouping ENT Disease Data Using the K-Means Clustering Algorithm, was able to group patients well, and evaluation using Davies Bouldin (DBI) produced a value of 0.90, which indicates that the quality of the resulting cluster is quite good [5].

The main objective of this study is to optimize the sales transaction data grouping model at the Josi.id store using the K-Means algorithm, to identify the characteristics of sales transaction data that can improve business strategies. can analyze produced the best cluster at K = 3 with a Davies-Bouldin Index (DBI) value of 0.258, indicating good cluster separation. The TYPE OF ITEM attribute plays a significant role in creating homogeneous clusters, supporting the analysis of transaction patterns and data distribution. Clustering produces: Cluster 0: Dominant sales during Eid (429 items). Cluster 1: Sales increase on holidays (343 items). Cluster 2: Stable pattern on weekdays (309 items). Tunic Clothes recorded the highest sales (up to 7 million), tops under 2 million, while dresses and skirts have variations of up to 3 million. This clustering supports decision making based on (love features) which results in strategic inventory and marketing.

With this research, the data mining approach is applied to analyze sales transaction data using KDD (Knowledge Discovery in Databases) with the K-Means Algorithm which can significantly improve the clustering model for fashion goods transaction data on online platforms such as Josi.id to utilize historical sales data, businesses can uncover consumer preferences and optimize inventory and marketing strategies [6].

The implications of this research, if the research objectives are achieved, are that the K-Means algorithm can be an effective tool in grouping complex transaction data [7]. This opens up opportunities for other researchers to develop better data analysis methods or combine K-Means with other algorithms to improve the accuracy and efficiency of data analysis in the E-commerce sector [8].

2. RESEARCH METHODOLOGY

The research method used in this study is a quantitative approach, aimed at analyzing sales transaction data systematically and measurably to obtain objective and accurate results. Statistical and computational techniques are used to evaluate the performance and accuracy of the K-Means clustering algorithm. Experimental techniques are employed to assist in determining the most effective parameter settings for the K-Means algorithm.

2.1. Knowledge Discovery in Database (KDD) Analysis Technique

As for analyzing data in the application of data mining, it uses the Knowledge Discovery in Databases (KDD) [9] stage process which consists of Data Selection, Data Preprocessing, Data transformation, Data mining, Evaluation / Interpretation. Which is a non-trivial process to find new patterns that are useful and easy to understand [10].

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

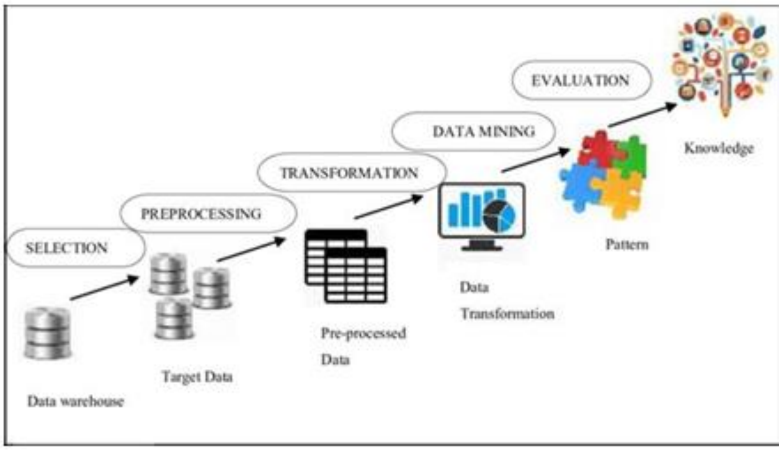

Figure 1. Knowledge Discovery in Databases

1. Data *Selection,* At the stage of selecting or selecting data used in research. Therefore, the process of selecting available data is appropriate and supports the analysis process used in research.
2. Data *Preprocessing ,* At the Processing Stage is an important step in data processing which aims to overcome missing data and clean up duplicate data and prepare data before analysis or modeling.
3. Data *Transformation,* At this stage, the data change process is carried out, before the data can be processed using the K-Means algorithm.
4. Data *Mining* is the process of extracting important information or patterns from large and complex datasets. This process involves the application of statistical methods, machine learning algorithms, or other data analysis techniques.
5. *Interpreation / Evaluation,* namely the process of analyzing and understanding the evaluation results of the model or algorithm used in a study or data analysis.

2.2. K-Means Algorithm

Algorithm K-Means one of the methods clustering (grouping) the most popular in unsupervised learning [10]. This algorithm aims to divide data into a number of groups (clusters) based on the similarity or proximity between data. To group data into K clusters based on proximity to the centroid [8]. This algorithm begins by determining the number of clusters, randomly selecting centroids, and automatically grouping data without requiring initial information about the target class [11]. Refers to the systematic process of discovering useful knowledge from large data sets. This process is widely used in data mining projects and involves several stages that guide the transformation of raw data into actionable insight [12]. Here is the formula:

$$d(x_i, c_i) = \sqrt{\sum_{i=1}^{N} (x_i - c_i)} \quad (1)$$

Where,
x_i = Data distance
c_i = Centroid Point

2.3. Davies Bouldin Index (DBI)

In this study, data mining techniques were carried out using the k-means algorithm. The k-means algorithm is one of the data mining algorithms that is widely used in clustering research. This

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

study will apply the Davies Bouldin Index (DBI) as one way to determine the most optimal number of clusters [13]. shows that in terms of grouping insurance products offered by national companies, the K-Means technique achieves a better Davies Boldin Index (DBI) value than the K-Medoids method. In addition, a web-based vehicle fleet clustering application was developed using the K-Means study method because of its high relevance based on the DBI value [14]. Here are the formulas:

1. Sum of Square Within-cluster (SSW)

To find out the cohesion in a cluster i is by calculating the value of the SSW cluster. Cohesion is defined to obtain the Sum of Square Within cluster value.

$$SS = W_i \frac{1}{m_i} \sum_{j=}^{m_i} d\ (xj, cj) \tag{2}$$

W_i : Weight or importance of the cluster i.

m_i : Number of data points in cluster i.

$d\ (x_i , c_j)$:Distance between a data point xj, and the centroid cj of cluster i.

2. Sum of Square Between-cluster (SSB)

The SSW calculation aims to determine the separation between clusters. Equation (3) is used to calculate the Sum of Square Between cluster value.

$$SSB_{i,j} = \sum_{i=1}^{Nc} |C| \cdot d(x_i, c_i) \tag{3}$$

N_C :Total number of clusters.

C : The size or cardinality of a cluster.

$d(x_i, c_i)$: Distance between the cluster centroid and data points.

3. Ratio

The ratio aims to determine the comparative value between the two clusters i and cluster to-j. To calculate the ratio value owned by each cluster, to calculate the ratio, the following equation (4) is used.

$$R_{i,j} = \frac{SSW_I + SSW_J}{SSB_{\boldsymbol{i,j}}} \tag{4}$$

Resda Dayanti: *Corresponding Author

Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

SSW_I, SSW_J : Sum of square within-cluster for clusters i.

$SSB_{i,j}$: Sum of square between clusters i and j.

4. Davies Bouldin Index

The ratio value obtained previously can be used to calculate the Davies Bouldin index (DBI) value using the following equation (5) as follows.

$$DBI=\frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}\, R_{i,j} \tag{5}$$

k : The total number of clusters

$R_{i,j}$: the ratio that compares the within-clusters scatter(compactness) and between-cluster separation for clusters i and j.

$max_{i\neq j}\ R_{i,j}$:The maximum ratio for cluster i when compared to any other cluster $j(i\neq j)$.

Experimental techniques are used to help determine the most effective parameter settings for the K-Means algorithm, such as: the optimal number of clusters (k) using the Davies Bouldin Index (DBI) method, attributes that affect the DBI value, and the characteristics of each cluster. Figure 3.1 on Research Methods.

3. RESULTS AND DISCUSSION

3.1. Result

The results of the research in this discussion will describe the optimization process of the clustering model with the Davies Bouldin Index (DBI) value which will affect the formation of the number of Clusters, attributes that affect the Davies Bouldin (DBI) value, characteristics of each cluster in sales transaction data at the Josi.id store using the K-Means algorithm using machine learning, namely RapidMiner (AI Studio 2024.1.0).

3.1.1. Data Understanding

The data obtained from Josi.id Store was collected through direct interview techniques with the store owner using WhatsApp online chat messages. This interview process aims to obtain accurate and relevant primary information related to operations, customer satisfaction, and sales performance. The raw data is then systematically summarized using Microsoft Excel to facilitate further analysis.

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Tanggal	NO	Nama Barang	Jenis Barang	Jumlah	Harga	Total
01/04/2023	1	ABRINA DRESS	*DRESS*	15	Rp 166.760	Rp 2.501.400
02/04/2023	2	ONE SET CIARA	DRESS	15	Rp 226.160	Rp 3.392.400
03/04/2023	3	SKRIT CARDIGAN	DRESS	1	Rp 127.261	Rp 127.261
….	….	………	….	….	……	……
31/10/2023	1082	ABRINA DRESS	DRESS	9	Rp 166.760	Rp 1.500.840
……..	…….	……	…….	……..	…….	………
31/10/2023	1087	ONE SET AYISHA MELAYU	TUNIC	29	Rp 254.320	Rp 7.375.280

3.1.2. Data Selection

This data stage is carried out on the View of the operator retrieve from the Microsoft Excel file by selecting a total of 1,087 data to 1,084 data on the attributes No., Date, Item Name, Item Type, Quantity, Price, Total. To ensure that the selected is the type of item and no. Can be seen in Figure 4.1 as follows:

Figure 2. Operator Retrieve From File

parameters used in the Retrieve From File operator, specifically listing a parameter called Repository Entry and its associated values, which include real Josi.id sales data. Displayed in Table 2 as follows:

Table 2. Parameter Operators

Number	Parameter	File
1	Repository entry	REAL SALES TRANSACTION DATA

The dataset used in this study is the fashion sales transaction dataset from the Josi.id store contained in excel data. And the results of the retrieve from file in Table 3 as follows:

Table 3. Hasil Operators on Retrieve From File

NO	Cluster	TYPES OF GOODS	INFORMATION
1	Cluster_2	0	DRESS
2	Cluster_2	0	DRESS
3	Cluster_2	0	DRESS
… ….	………	…	…..
1083	Cluster_1	1	KURUNG CLOTHES
1084	Cluster_1	1	KURUNG CLOTHES

The Select Attributes operator works on the "exa" connection, meaning to refer to a specific tool or system, to be able to be part of a data selection or extraction tool. In Figure 4.2, the select attributes operator is as follows:

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

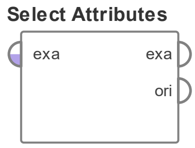

Figure 3. Select attributes operator

Showing Figure 4.3 In the analysis process in RapidMiner shows the data processing model pattern as follows:

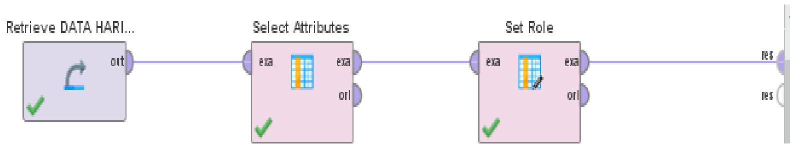

Figure 4. Preprocessing step process model

Figure 4.4 After applying the Select Attributes operator model process, the dataset will consist of the results of the selected attributes according to the settings for the TYPE OF GOODS and NO. attributes. The results are shown in the illustration in Table 4 as follows:

Table 4. Select attributes results

NO	CLUSTER	TYPES OF GOODS	INFORMATION
1	Cluster_1	0	DRESS
2	Cluster_1	0	DRESS
3	Cluster_1	0	DRESS
… ….	…….	…	…..
1083	Cluster_2	1	TUNIC CLOTHES
1084	Cluster_2	1	TUNIC CLOTHES

3.1.3. Data Preprocessing

The preprocessing stage is an important step in data processing that aims to clean up duplicate data and overcome missing data and prepare data before analysis or modeling. One of the operators often used in this stage is "Replace Missing Values". Shows the operator in Figure 4. 4.

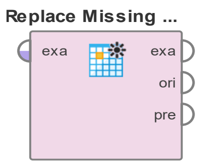

Figure 5. Operator Replace Missing Values

The result of using the "exa" connection on the Replace Missing Values Operator shows the attributes of a dataset after processing. So, the illustration of the operator results in Table 5 as follows.

Table 5. Result of the Replace operator missing values

ATTRIBUTE	ATTRIBUTE TYPE	MISSING VALUE
NO	Integer	0
TYPES OF GOODS	Polynomial	0
…….	……	0
PRICE	Integer	..
NAME OF GOODS	Integer	0

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Showing Figure 4.5 In the analysis process in RapidMiner shows the data processing model pattern as follows:

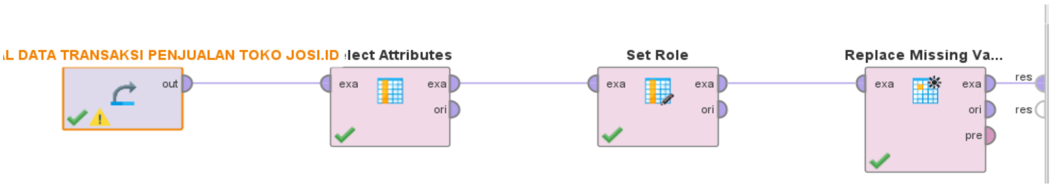

Figure 6. Preprocessing step process model

3.1.4. Data Selection

The stage of converting raw data into a format more suitable for analysis using the Nominal to Numerical and Date to Numerical operators. Stages in the crucial steps to prepare the dataset to suit the needs of analysis and clustering using the K-Means algorithm.

1. Nominal to numerical Use of the Nominal to numerical operator Used to change the type of polynomial or Nominal attributes to numeric types. in Figure 4.6 as follows:

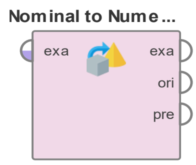

Figure 7. Nominal to numerical operators

Numeric format makes data manipulation easier, by using the operator from polynomial to numeric. The illustration in the table shows the results of Table 5 as follows:

Table 6. Nominal to Numerical Results

NO	CLUSTER	TYPES OF GOODS	INFORMATION
1	Cluster_1	0	DRESS
2	Cluster_1	0	DRESS
3	Cluster_1	0	DRESS
… ….		…	…..
1083	Cluster_2	1	TUNIC CLOTHES
1084	Cluster_2	1	TUNIC CLOTHES

2. Date to Numerical is one of the stages in data transformation, fashion sales transaction data from Date, Month, Year. With the aim of converting date-based attributes into a numeric format that is more suitable for analysis. Illustration of the date to numerical operator in Table 4. 7 as follows:

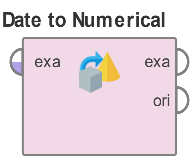

Figure 8. Date to Numerical Operator

Numeric format makes it easier to manipulate data, such as calculating duration between dates, identifying patterns based on time. Illustration in Table 7 as follows:

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Table 7. Date to Numerical Results

NO	CLUSTER	DATE	INFORMATION
1	Cluster_2	0	02/04/2023
2	Cluster_2	0	02/04/2023
3	Cluster_2	0	02/04/2023
… ….		…	…..
1083	Cluster_1	1	10/31/2023
1084	Cluster_1	1	10/31/2023

Showing Figure 4.9 In the analysis process on RapidMiner shows the following model transformation pattern:

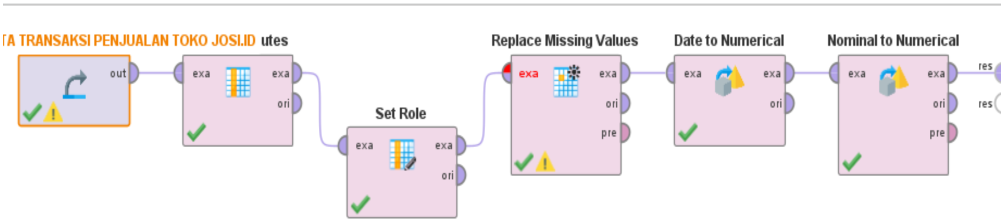

Figure 9. Transformation step process model

3.1.5. Data Transformation

At the stage of the K-Means algorithm is used to perform the clustering process, so that it allows the formation of product groups based on specific characteristics and the level of sales popularity by using the K-Means Clustering operator. The following is the K-Means operator in Figure 4.10.

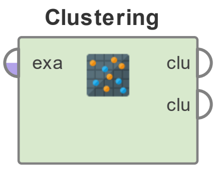

Figure 10. K-Means Operator

Such an iterative approach allows evaluating the performance of the algorithm on different numbers of clusters, resulting in more accurate and relevant data segmentation.

Table 8. K-Means Clustering Results

CLUSTERING	NUMBER OF CLUSTER	QUANTITY
2	Cluster 0: 429 items Cluster 1: 652 items	0.456
3	Cluster 0: 429 items Cluster 1: 343 items Cluster 2: 309 items	0.258

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Based on Binary Representation, the Cluster Distance Performance operator is used to evaluate the quality of the clusters formed and determine the most optimal number of Clusters (K). This operator calculates the average distance between each data point and its cluster centroid. The smaller the average distance, the better the quality of the resulting Cluster, on the attributes TYPE OF ITEM and NO of the cluster members is the implementation of the K-Means clustering model cluster In the experimental results, the DBI value using can be measured in binary numbers with the smallest value obtained is the K value = 3 with the Davies Bouldin (DBI) value of 0.258. This value is proven to be the best DBI value. Illustration of the operator performance in Figure 4. 11.

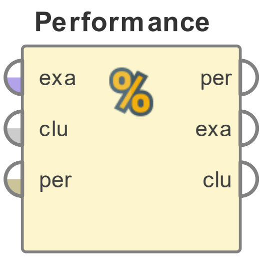

Figure 11. Operator Distance performance

For each value of k, the Davies-Bouldin index is calculated to determine the value of k that produces the best cluster. The following is a display of Table 4 as follows:

Table 8. Clustering results using binary numbers

CLUSTERING	DAVIES BOULDIN
2	0.371
2	0.453
2	0.371
2	0.456
3	0.975
3	0.520
3	0.274
3	0.258

The smaller the DBI value, the better the clustering quality, indicating higher compactness within clusters and clearer separation between different clusters. A DBI of 0.258 demonstrates better clustering performance compared to other values, such as 0.456, at a specific k. This value proves that the method used can identify optimal patterns in data clustering. A smaller Davies-Bouldin Index (DBI) indicates greater separation between clusters and higher compactness within clusters.

3.1.6. Interpretation / Evaluation

The table shows that the integration of information on commodity transaction sales, assessed through the Davies-Bouldin Index (DBI), reveals that values close to zero occur between experimental clusters 2 and 3, with cluster 3 producing an optimal k value of 0.258 when taking into account the attribute TYPE OF ITEM. Cluster 0 with 429 items has a centroid value of 2.611, which means that the average type of goods is dominated by the group of goods with the largest value. Cluster 1 with 343 items and a centroid value of 0 has an average type of goods with no dominant attribute, Cluster 2 with a value of 309 and a centroid of 1 has an average type of goods that dominates.

This cluster model distribution explains the degree of variability in the volume of data within each cluster and facilitates the evaluation of the efficacy of the cluster depiction in the clustering procedure performed. An illustration of the results of the cluster model is shown in Table 9.

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Table 9. Cluster model operator results

NO	CLUSTER MODEL	TYPES OF GOODS
1	Cluster 0	429 items
2	Cluster 1	343 items
3	Cluster 2	309 items

The following is an illustration based on the Performance Vector result on RapidMiner in Table 10 as follows:

Table 10. Performance Vector operator results

NO	PERFORMANCE VEKTOR	VALUE
1	Avg. within centroid distance	0.094
2	Avg. within centroid distance_cluster_0	0.238
3	Avg. within centroid distance_cluster_1	0.000
4	Avg. within centroid distance_cluster_2	0.000
5	Davies Bouldin	0.258

Shows a value of 0.094 which indicates that in general the data in the cluster is quite close to their respective centroids, indicating a good level of cluster compactness. Cluster 0 has an average distance of 0.238, which is higher than other clusters, indicating that the data in this cluster is more spread out. The process of analyzing Josi.id Store sales data is carried out using the clustering method to group data based on certain patterns. Illustration of the final results of the analysis based on clustering analysis using K-Means in Figure 4. 12.

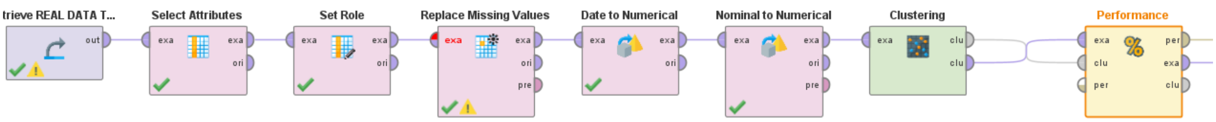

Figure 12. Final model implementation on RapidMiner

The Davies-Bouldin Index (DBI) value of 0.258 indicates a good distribution of cluster quality, because a smaller DBI indicates a more optimal clustering. At K = 3, with scatter / bubble plot visualization at this value indicates effective cluster separation and high internal consistency. The visualization results on the scatter plot in the Figure 4.13.

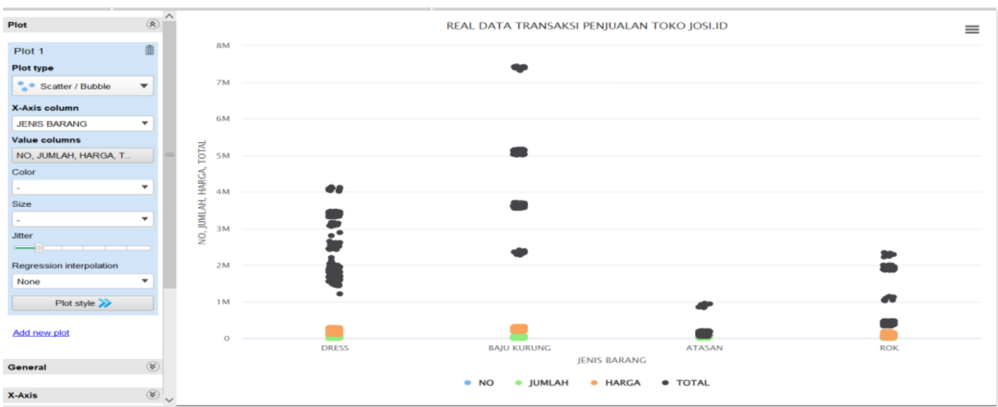

Figure 13. Scatter Plot

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

The distribution graph of Josi.id Store sales data shows:
1. Tunic Clothes : Highest sales (7 million), utilizing the love feature.
2. Top: Lowest sales (under 2 million).
3. Dress: Significant variation (1–3 million), leveraging traffic.
4. Skirt: Sales vary (up to 3 million), also taking advantage of the love feature.

3.2. Discussion

3.2.1. Optimal K value based on Davies Bouldin Indeks (DBI)

Ananda et al.'s research focuses on data on the percentage of children who have birth certificates in Indonesia, with an average of 84.71%. From the clustering process, 4 clusters were found with different member compositions, namely 1 member in cluster 1, 19 members in cluster 2, 13 members in cluster 3, and 1 member in cluster 4[15].The main attributes include QUANTITY, PRICE, DATE, ITEM TYPE, ITEM NAME, and AMOUNT. The K-Means algorithm successfully grouped sales transactions at the Josi.id Store into 3 clusters with a Davies-Bouldin Index (DBI) value = 0.258, indicating good clustering quality. he results of this study are in line with the research of

3.2.2. Best Cluster based on value Davies Bouldin Indeks (DBI)

The distribution of the number of items in each cluster shows significant variation, with Cluster 0 (429 items - Eid day), Cluster 1 (343 items - holidays), and Cluster 2 (309 items - weekdays). This result is in line with the research of Ibnu fikri Fauzi et al., who used DBI to determine the optimal number of clusters, where K=3 gave the best result with DBI = 0.406, supporting the effectiveness of clustering in data analysis[13].

3.2.3. Characteristics of each cluster in the Principal Component Attributes

Item type determines transaction grouping, with dresses, baju kurung, tops, and skirts showing seasonal trends. Sales increase significantly during eid, while weekdays are more stable. K-Means (K=3, DBI=0.258) forms Cluster 0 (Lebaran - 429 items), Cluster 1 (Holidays - 343 items), and Cluster 2 (Weekdays - 309 items). Dresses are the favorite with the highest sales, while tops and skirts are stable, especially during new model launches. sukirman et al.'s research confirms the role of product type in forming homogeneous clusters, with DBI = 0.181 in K-Means Clustering. This result indicates good clustering quality, where products are grouped based on their level of popularity[16].

5. CONCLUSION

The implementation of the K-Means Algorithm on the fashion sales data of Josi.id Store for seven months produced the best cluster at K = 3 with a Davies-Bouldin Index (DBI) value of 0.258, indicating good cluster separation. The TYPE OF ITEM attribute plays a significant role in creating homogeneous clusters, supporting the analysis of transaction patterns and data distribution. Clustering produces: Cluster 0: Dominant sales during Eid (429 items). Cluster 1: Sales increase on holidays (343 items). Cluster 2: Stable pattern on weekdays (309 items). Tunic Clothes recorded the highest sales (up to 7 million), tops under 2 million, while dresses and skirts have variations of up to 3 million. This clustering supports decision making based on (love features) which results in strategic inventory and marketing.

6. ACKNOWLEDGEMENT

I would like to thank Allah SWT for His blessings, always supporting me in any case. And to the owner of the Josi ID shop who has given me the opportunity to collect data that aims to be used as research material. And thanks to the parents of the supervising lecturer, friends in the struggle, along with the ranks of STMIK IKMI CIREBON who have provided input, encouragement and direction in completing this research.

7. REFERENCES

[1] B. KANDEMİR, ‘A Methodology for Clustering Items with Seasonal and Non-seasonal Demand Patterns for

Resda Dayanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 1, March 2025, Page 84-96
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i1.21

Inventory Management', *J. Adv. Res. Nat. Appl. Sci.*, vol. 8, no. 4, pp. 753–761, 2022,
[2] L. Xue, 'Cluster analysis-based big data mining method of e-commerce consumer behaviour', *Int. J. Web Based Communities*, vol. 19, no. 1, pp. 53–63, 2023, doi: 10.1504/IJWBC.2023.10048319.
[3] Y. Hartati, S. Defit, and G. W. Nurcahyo, 'Best Seed Clustering Using K-Means Algorithm in Increasing Sales', *J. Inform. Ekon. Bisnis*, vol. 3, pp. 4–10, 2021.
[4] W. Rosida and Y. A. Wijaya, 'Clustering of HIV/AIDS Disease in West Java Using K-Means Clustering Algorithm', *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 306–315, 2023.
[5] M. I. Hutagalung and Sriani, 'Grouping of ENT Disease Data Using K-Means Clustering Algorithm', vol. 4, no. October, pp. 1568–1577, 2024.
[6] G. Triyandana, L. A. Putri, and Y. Umaidah, Data Mining Application of Food and Beverage Menu Grouping Based on Sales Level Using the K-Means Method', *J. Appl. Informatics Comput.*, 2022, doi: 10.30871/jaic.v6i1.3824.
[7] M. Dahria, R. Gunawan, and Z. Lubis, 'Implementation of K-Means for the Best Product Grouping of PT. Koko Pelli', Semin. Nas. Science and …, pp. 495–498, 2019, [Online].
[8] A. R. Khorunisa and M. Fadilah, The Influence of Shopee Marketing Strategy on Consumer Shopping Behavior in Indonesia', *Civ. Off. Empir. Stud. Soc. Sci.*, vol. 3, no. 1, pp. 1–8, 2024, doi: 10.53754/civilofficium.v3i1.519.
[9] W. Purba, W. Siawin, and . H., 'Implementation of Data Mining for Grouping and Predicting Potentially Layoff Employees Using the K-Means Clustering Algorithm', *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019.
[10] Idham, H. Rosika, and Yuliadi, CLOTHING USING THE K-MEANS METHOD In the digital era, technological developments are increasingly developing, especially sales data management is becoming increasingly crucial for businesses to understand consumer behavior, identify several groups or clusters that have', vol. 5, pp. 221–231, 2024.
[11] A. S. Permana and R. Sanjaya, Application of Data Mining in Grouping Popular Grocery Items Using K-Means Clustering (Case Study: Gunung Bumi Store)', J. Educator. Tambusai, vol. 7, no. 1, pp. 2419–2424, 2023.
[12] S. Lestari, Muhatri, A. R. Fachrezi, M.Agung.Sutrisno, and M. Geovany, 'Implementation of the K-Means Clustering Algorithm on Sales of Specs' Brand Futsal Shoes, vol. 2, no. 3, pp. 50–61, 2024.
[13] I. F. Fauzi, M. Gito Resmi, and T. I. Hermanto, 'Determining the Optimal Number of Clusters Using the Davies Bouldin Index in the K-Means Algorithm to Determine Disease Groups', vol. 7, no. 2, pp. 2598–8069, 2023.
[14] H. Syukron, M. Fauzi Fayyad, F. Junita Fauzan, Y. Ikhsani, and U. Rizkya Gurning, 'Comparison of K-Means K-Medoids and Fuzzy C-Means for Customer Data Clustering with LRFM Model', *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 76–83, 2022.
[15] D. Ananda, S. Rohimah, B. Susilo, D. Wulan, and A. Mustofa, ' Implementation of K-Means in Grouping Birth Certificate Data in Indonesia', *SENTIMAS Semin. Nas. Penelit. dan Pengabdi. Masy.*, pp. 66–71, 2022, [Online].
[16] M. I. Sukirman, T. H. Pudjiantoro, I. Santikarama, J. Informatika, and F. Sains, 'Sales Clustering of Clothing Line Products Using the K-Means Clustering Method for Implementing the Up-Selling Concept', pp. 56–62, 2021.

Resda Dayanti: *Corresponding Author