Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Sales Data Classterization Analysis Using K-Means Method for Marketing Strategy Development

Mifta Almaripat[1*], Ahmad Faqih[2], Ade Rizki Rinaldy[3]

[1,2]Informatics Engineering, STMIK IKMI Cirebon
[3]Informatics Management, STMIK IKMI Cirebon
mitaalmarifat@gmail.com [1]*

Abstract

In the digital era, utilizing sales data is very important to support strategic decision making. This research aims to overcome the problems faced by 9Doors Store in optimizing marketing strategies and stock management. The main problem faced is the lack of in-depth analysis of existing sales data, which results in difficulties in formulating appropriate marketing strategies and efficient stock management. For this reason, this research applies the K-Means Clustering method to group products based on customer purchasing behavior characteristics. The data used includes product categories, selling prices, initial stock, number of products sold, and total sales obtained from 9Doors Store during the period March to September 2024. The method used in this research is Data Mining approach with K-Means algorithm, which is implemented using RapidMiner software. The data analysis process goes through Knowledge Discovery in Databases (KDD) stages, including data collection, data cleaning (preprocessing), data transformation, and data mining using K-Means. Cluster evaluation is done using Davies-Bouldin Index (DBI) to assess the quality of clustering results. The results of this study show that the division of sales data into three clusters provides optimal results with the lowest DBI value (0.106), which indicates efficient clustering. This finding identifies products with high, medium, and low sales levels, which can be used to formulate more targeted marketing strategies. With these results, Toko 9Doors can improve stock management and design more effective promotions based on better customer segmentation.

Keywords: *K-Means, clustering, sales data, marketing strategy, RapidMiner, Davies-Bouldin Index*

1. Introduction

K-Means algorithm is one of the most popular clustering algorithms due to its simplicity and efficiency. This research method uses a clustering data analysis approach using the K-Means Clustering algorithm based on the Davies Bouldin Index (DBI) value, and Measure types Numerical Measures with Euclidean Distance, and Manhattan Distance types, then iteration is also used in the clustering process. The K-Means Clustering algorithm was chosen because it is able to group objects based on similar characteristics in one cluster. Furthermore, accuracy testing is done with the Davies Bouldin Index (DBI) method to evaluate the best performance of clustering in producing optimal clusters. [1] The use of E-commerce can also provide benefits for companies such as expanding market places, reducing distribution costs and fast access to information, but competition in the business world must exist because those who sell goods there are still other stores that open the same business. This condition causes the owner of this shop to be required to find strategies that can increase store sales and marketing.

2. Research Methods

An easy way to comply with the paper formatting requirements is to use this document as a template and simply type your text into it.
Your paper must use a page size corresponding to A4 which is 21cm wide and 29.7cm long. The margins must be set as follows:

1. Problem Identification; Sales data at 9Doors Store will go through processing and cleaning before being analyzed using the K-Means Clustering algorithm. This algorithm serves to group clothing products based on sales patterns, by determining the optimal K value using the Davies Bouldin Index (DBI) [2] . This analysis aims to identify item category, item type, item type, initial stock, sold quantity, remaining stock, selling price, and total sales, based on sales data during a certain period. The results of this analysis will provide important information to formulate a more effective stock management strategy, so that Toko 9Doors can maximize profits while minimizing potential losses.
2. Data collection; Retrieved sales data from 9Doors with 418 datasets covering the period March to September 2024.
3. Data analysis; The analysis process is carried out to identify sales problems at the 9Doors Store. This analysis is done by evaluating sales data during a certain period. The sales data is then analyzed using the K-Means Clustering algorithm. This algorithm is used to

group products based on sales data characteristics, so that it can help 9Doors Store in understanding sales patterns and formulating more effective strategies[3].

4. Data Processing; 9Doors Store product sales data is processed and analyzed using the K-Means Clustering method in RapidMiner software[5]. This process aims to group products based on sales patterns. Data is processed systematically with RapidMiner software, resulting in product groupings that form the basis of the right stock management strategy to maximize 9Doors Store profits.
5. The KDD process includes systematic steps from start to finish, which can be generally described as follows:
   a. 9Doors store sales data containing attributes such as item category, item type, initial stock, quantity sold, final stock, selling price, and total sales during a certain period.
   b. Data Selection Selecting relevant data from the dataset, such as only those attributes that are directly related to sales pattern analysis.
   c. Pre-processing/Cleaning Cleaning the data from irrelevant or invalid values. This includes dealing with empty values or outliers and ensuring data consistency.
   d. Transformation Normalizing or transforming data into a format suitable for K-Means analysis, such as ensuring attributes have a uniform scale.
   e. Data Mining Using the K-Means algorithm to group data into clusters based on similar features.
   f. Evaluation Evaluate the clustering results with metrics such as Davies-Bouldin Index (DBI) to determine the best number of clusters.

2.1. Data Mining

Data Mining and Knowledge Discovery in Databases (KDD) for 9Doors represent a systematic approach to transforming raw sales data into strategic business insights through a comprehensive six-stage process[6]. The methodology begins with meticulous data selection, gathering transactional information from point-of-sale systems, online platforms, and customer management records, followed by rigorous preprocessing to clean and standardize the data, ensuring high-quality input for analysis. During the transformation stage, raw data is converted into meaningful formats, creating derived variables like purchase frequency, average transaction value, and profit margins, which enable deeper understanding of sales dynamics[7]. The core data mining phase employs advanced techniques such as classification to segment customers, clustering to identify purchasing patterns, association rule mining to discover product relationships, and predictive modeling to forecast sales trends, ultimately uncovering hidden opportunities within the 9Doors sales ecosystem. Pattern evaluation critically assesses the statistical significance and business relevance of discovered insights, filtering out noise and focusing on actionable intelligence. The final knowledge presentation stage visualizes findings through interactive dashboards and comprehensive reports, translating complex data analyses into clear, strategic recommendations that can drive targeted marketing efforts, optimize product offerings, and enhance overall business performance. By leveraging this data-driven approach, 9Doors can gain a competitive edge, make informed decisions, understand customer behaviors more deeply, and develop personalized strategies that not only increase sales but also improve customer satisfaction and long-term business growth[8].

2.2. Clustering

Clustering is the process of dividing a collection of objects into groups based on internal similarities and external difference [8].

2.3. K-Means Algorithm

K-Means algorithm is an Unsupervised Learning method with an iterative process, where the dataset is grouped into k predetermined number of non-overlapping clusters or subgroups. The algorithm attempts to keep the points in the cluster close to each other, while the clusters themselves are in different spaces. The goal of this algorithm is to allocate data points to clusters so that the sum of the squared distances between the data points and the cluster centroids is minimal. At this point, the centroid of the cluster is the average value of the data points in the cluster.

3. Result And Discussion

3.1. Sample Dataset

Tabel 1: Sample Dataset

No	Size		Kategori	Jenis Barang	Tipe Barang	Stok Awal	Terjual	Stok Akhir	Harga Jual	Total Penjualan	Status
1.	820	2100	Pintu	A LAMINATE UNO	Flush Membrane	50	11	39	1027700	11304700	Tersedia
2.	820	2100	Pintu	A LAMINATE UNO	Solid Honeycomb Board	50	15	35	1236100	18541500	Tersedia
3.	820	2100	Pintu	A LAMINATE UNO	Solid Wood	50	15	35	2200800	33012000	Tersedia
4.	820	2100	Pintu	B. LAMINATE DUO	Flush Membrane	50	16	34	1074400	17190400	Tidak Tersedia
…	…	…	…	….	…	…	…	…	…	…	…
418.	700	2000	Jendela	Jendela Mati	D-26	10	0	10	1609700	0	Tersedia

3.2. Data Selection

At the data selection stage, the data that has been collected will be selected for analysis needs. The initial dataset has 11 attributes. To filter the attributes to be used, the Select Attributes operator is used in the RapidMiner application. Of the 11 attributes in the 9Doors

Store sales dataset, only 6 attributes were selected through this filtering process: Sales Price, Initial Stock, Final Stock, Sold, Item Type, Sales Price, Total Sales. The Select Attributes operator view can be seen in Figure 3.1.

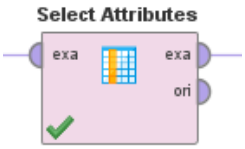

Fig. 1: Operator Select Attributes

Tabel 2: Parameter Select Attributes

Parameter	Description
Type	Include Attributes
Attributes filter type	A subset
Select Attributes	Harga Jual, Stok Awal, Stok Akhir, Terjual, Tipe Barang, Harga Jual, Total Penjualan

3.3. Data Preprocessing

Pre-processing or data cleaning is the first step in data analysis that aims to prepare and clean the data that has been selected, ensuring that the data is of good quality before being analyzed[9]. This stage is very important to overcome noise, inconsistencies, or other problems that can affect the results of the analysis. In this dataset, there are no missing values, so the cleaning process does not require the use of additional operators to fill in or delete missing data[10]. The results of this check are shown in Figure 3.2, which confirms that there is no missing data, so the data is ready to be used in the next stage of analysis.

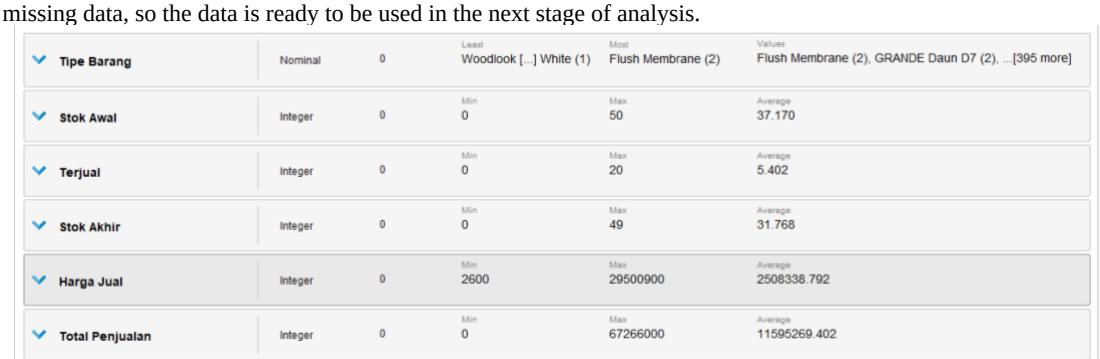

Fig. 2: Result Preprocessing

3.4. Data Transformation

The transformation process is needed in data mining to change the type of data to suit the required analysis[11]. in this study, the categorization of attributes in the dataset can be done using the Set Role operator in the RapidMiner application, according to the analysis needs. In this study, the attribute “Tipe Barang.” is categorized as “ID”. The Set Role operator display can be seen in Figure 3.3.

Fig. 3: Operator Set Role

Tabel 3: Parameter Set Role

Parameter	Description	
Set Role	Edit List :	
	Attribute Name :	Target role :
	Tipe Barang	Id

3.4.1. Data Mining

In the data mining stage, the clustering technique implemented is the K-Means Clustering algorithm using the Clustering operator. This operator is the main modeling operator to produce dataset clustering. The clustering operator display can be seen in Figure 3.5.

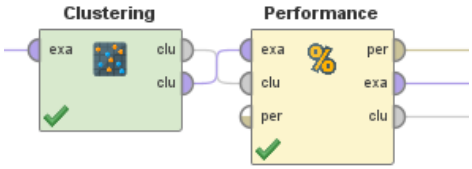

Fig. 4: Operator Clustering

Tabel. 4: Parameter Clustering

Parameter	Isi
K	2-9
Measure Types	Numerical Measures, Bregman Divergences, Mixed Measures
Max Optimazation	100

4. Results And Discussions

4.1. Result

This research is based on the comparison of Davies Bouldin Index (DBI) value with K-Means Clustering method, using various types of measurements such as Numerical Measure, Mixed Measure, and Bregman Measures with Euclidean Distance type[12]. The testing process is done through iterations to determine the best performance in data processing, starting with the value of K = 2 to K = 9. The analysis showed that the cluster with the closest DBI value to 0 was found at K=3, with a value of 0.106. This value indicates a fairly good clustering performance in forming optimal clusters. In addition, Mixed Measure and Numerical Measure parameters using Euclidean Distance also produce the best clustering with a DBI value close to 0, which is 0.108. Based on these results, it can be concluded that dividing the data into 3 clusters provides a better level of accuracy. The DBI value of 0.106

Tabel 5: Testing Results Based on Davies Bouldin Index (DBI)

Cluster	Measure Type	Davies Bouldin Index
2	MixedMeasures	0.112
	NumericalMeasures	0.112
	Bregman Measure	0.112
3	MixedMeasures	0.108
	NumericalMeasures	0.108
	Bregman Measure	0.106
4	MixedMeasures	0.111
	NumericalMeasures	0.111
	Bregman Measure	0.111
5	MixedMeasures	0.116
	NumericalMeasures	0.116
	Bregman Measure	0.113
6	MixedMeasures	0.122
	NumericalMeasures	0.122
	Bregman Measure	0.110
7	MixedMeasures	0.130
	NumericalMeasures	0.130
	Bregman Measure	0.111
8	MixedMeasures	0.139
	NumericalMeasures	0.139
	Bregman Measure	0.124
9	MixedMeasures	0.130

	NumericalMeasures	0.130
	Bregman Measure	0.130

4.2. Discussions

This is a text of acknowledgements. Do not forget people who have assisted you on your work. Do not exaggerate with thanks. If your work has been paid by a Grant, mention the Grant name and number here.

Table 6: Optimization Results and Cluster Evaluation Results

Number of Clusters (K)	DBI	Stability Centroid	Description
2	0.198	Unstable	Cluster division is not clear
3	0.106	Stable	Best clustering, minimal overlap
4	0.145	Stable	Not as good as K=3
5	0.160	Unstable	Clustering is more homogeneous, but overlap increases

5. Conclussion

The effectiveness of the K-Means Method is proven to be effective in grouping sales data based on certain characteristics. The clustering results show that the data can be divided into three optimal clusters with a Davies-Bouldin Index (DBI) value of 0.106, which indicates that the cluster division has good quality. This clustering provides deep insights into the sales patterns of each product group at 9Doors Store, making it easier to make decisions related to marketing and stock management[13].

Research Contribution, The results of this study show that the K-Means algorithm not only helps understand sales patterns but also becomes a practical tool to support data-based decision making[14]. Clustering provides a basis for customizing marketing strategies based on market needs, thereby improving operational efficiency and business competitiveness[15].

References

[1] [Al-Fahmi, B. M., Rahmawati, E., & Sagirani, T. (2023). Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *9*(2), 141–149. https://doi.org/10.25077/teknosi.v9i2.2023.141-149

[2] Alifa, T. T., & , Rini Astuti, F. M. B. (2024). Implementasi Data Mining Menggunakan Algoritma K- Means Clustering Dalam Analisis Penjualan Produk. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 602–607. https://doi.org/10.36040/jati.v8i1.8340

[3] Arsyad, A. T., Nurlatifah, H., & Sunarmo, S. (2024). Usulan Strategi Pemasaran Universitas Al Azhar Indonesia Menggunakan Metode K-Means Clustering. *JURNAL Al-AZHAR INDONESIA SERI SAINS DAN TEKNOLOGI*, *9*(1), 40. https://doi.org/10.36722/sst.v9i1.2699

[4] Ashari, I. F., Banjarnahor, R., Farida, D. R., Aisyah, S. P., Dewi, A. P., & Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*, *6*(1), 07–15. https://doi.org/10.30871/jaic.v6i1.3485

[5] Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, *1*(1), 1–5. https://doi.org/10.46576/djtechno.v1i1.964

[6] Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, *10*(2), 133–146. https://doi.org/10.34010/jamika.v10i2.2683

[7] [Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021). Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari. *Jurnal Sains Teknologi Transportasi Maritim*, *3*(1), 1–7. https://doi.org/10.51578/j.sitektransmar.v3i1.30

[8] Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(1), 606–620. https://doi.org/10.35957/jatisi.v9i1.1516

[9] [Idham, I., Rosika, H., & Yuliadi, Y. (2024). Implementasi Rapidminer Untuk Clestering Data Penjualan Pakaian Menggunakan Metode K-Means. *JUTECH : Journal Education and Technology*, *5*(1), 221–231. https://doi.org/10.31932/jutech.v5i1.3642

[10] Kristianto, W. W., & Rudianto, C. (2020). Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki). *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, *5*, 90–98.

[11] Mar'atun Sholihah, O., Suarna, N., Dwilestari, G., & R, N. (2023). Implementasi Metode K-Means Clustering Untuk Menganalisa Penerima Bantuan Di Desa Palasah. *Jurnal Informatika Dan Teknologi Informasi*, *2*(1), 154–160. https://doi.org/10.56854/jt.v2i1.123

[12] Maryam, S., Astuti, R., & Basysyar, F. M. (2023). *OPTIMALISASI JUMLAH CLUSTER DATA SEKOLAH DASAR (SD) MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING*. *7*(6), 3640–3646.

[13] Noviati, N., Mulyawan, M., Kurnia, D. A., & Rinaldi, A. R. (2022). Clustering Data Penjualan Produk Makanan pada Toko Toserba Yogya Siliwangi dengan Menggunakan Metode K-Means. *MEANS (Media Informasi Analisa Dan Sistem)*, *7*(1), 77–84. https://doi.org/10.54367/means.v7i1.1850

[14] Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 849–855. https://doi.org/10.36040/jati.v6i2.5755

[15] Sulaiman, M., Yudistira, R., Narasati, R., & Herdiana, R. (2024). Penerapan Data Mining dengan Metode Clustering untuk menentukan Strategi Peningkatan Penjualan Berdasarkan Data Transaksi. Jurnal Informatika Dan Rekayasa Perangkat Lunak, 6(1).