Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Application of the K-Means Algorithm in the Segmentation Of 3kg Lpg Customers

Gina Selvia Ananda[1]*, Nana Suarna[2], Agus Bahtiar[3], Arif Rinaldi Dikananda [4], Faturrohman[5]

[1,2,3]*Informatics Enginneering, STMIK IKMI Cirebon, Indonesia*
[4,5]*Software Engineering, STMIK IKMI Cirebon, Indonesia author*
anandaginaselvia@gmail.com[1*], st_nana@yahoo.com[2], agusbahtiar038@gmail.com[3], rinaldi21crb@gmail.com[4], fathurrahman.ikmi@gmail.com[5]

Abstract

This research was motivated by PT Sumber Perkasa Mandiri's need to understand the purchasing patterns of 3 kg LPG gas customers more accurately in order to improve the effectiveness of its marketing strategy. The purpose of this study was to apply the K-Means Clustering algorithm to form customer segmentation based on transaction behavior. The method used is a quantitative approach with sales data analysis of 850 records through the stages of data selection, preprocessing, attribute transformation, and modeling using RapidMiner Studio. Model evaluation was carried out using the Davies-Bouldin Index to determine the optimal number of clusters. The results of the study show the formation of two main clusters, namely the premium customer cluster with high purchase frequency and high loyalty, and the low-activity customer cluster that only makes purchases when necessary. The best DBI value at K=2 of 0.057 indicates excellent cluster separation quality. These findings conclude that K-Means Clustering is effective in identifying differences in consumption behavior, and its implications provide a strategic basis for companies to design loyalty programs for high-value customers and more intensive promotions for low-activity customers.

Keywords: *K-Menas Clustering, Customer Segmentation, Sales Data Analysis, Marketing Strategi, Davies Bouldin Index*

1. Introduction

The distribution of subsidized 3 kg LPG in Indonesia still faces complex structural challenges, ranging from inaccuracy in targeting subsidy recipients, high fiscal burdens on the state, to regional distribution disparities. These conditions reflect the weak governance of energy subsidies, particularly in terms of distribution and supervision.[1] asserts that misdirected subsidy policies have the potential to create fiscal inefficiencies while deepening social inequality.

In addition, although 3 kg LPG is designed to support the acceleration of clean energy transition, research by [2] shows that the benefits of subsidies are actually enjoyed more by the middle and upper classes who have better access and purchasing power. This situation highlights the urgency of evaluating subsidized energy distribution policies through a more comprehensive analytical approach.

On the other hand, LPG distribution issues are also influenced by logistical and operational constraints. [3] reveal that factors such as distance to agents, transportation infrastructure, and distribution efficiency are the main causes of LPG scarcity and supply inequality. Dependence on conventional distribution processes causes significant variations in supply between regions, requiring a data-driven approach to map demand patterns more accurately. Advances in modern analytics also open up such opportunities, as described by Trive.

2. Literature Riview

Understanding customer segmentation in the context of subsidized energy distribution is rooted in grand theories of consumer behavior and energy usage pattern analysis, which emphasize that community characteristics and consumption behavior are greatly influenced by socioeconomic factors, accessibility, and market structure. In the context of 3kg LPG, segmentation is necessary to identify user groups based on purchasing patterns and frequency of use, in order to support more adaptive subsidy policies. Research in the energy sector shows that machine learning-based analytical approaches can improve the accuracy of subsidies and the effectiveness of distribution[1][4]. This concept is in line with the basic theory of data mining, which defines the process of discovering patterns in big data through unsupervised learning techniques, including clustering, which aims to produce homogeneous groups within clusters and heterogeneous groups between

clusters[5]. Within this framework, clustering-based segmentation becomes a relevant theoretical approach to understanding variations in subsidized energy consumption behavior.

The middle-range theory framework in this study is based on the K-Means Clustering method as the main algorithm in numerical data clustering. K-Means works by minimizing the distance between objects and cluster centers through an iterative process of centroid initialization, Euclidean distance calculation, and centroid updating until convergence is achieved. As it has evolved, various algorithmic modifications and improvements have been developed to address the classic weaknesses of K-Means, such as sensitivity to initialization and limitations in handling correlated data. Variations such as regularized K-Means [6], the integration of K-Means with Particle Swarm Optimization [7], and deep embedding approaches for nonlinear data representation [8][9] show that clustering models can be strengthened through computational optimization and deep learning integration. Methodologically, customer segmentation research generally follows comprehensive stages that include preprocessing, dimension reduction, clustering and optimization implementation, and evaluation using metrics such as the Davies-Bouldin Index and Silhouette Coefficient

Empirical findings from various studies show that the application of clustering in the energy sector and subsidy systems still faces a number of substantive challenges, including feature limitations, the absence of dataset standards, and issues of data quality and granularity [10]. This research gap is reinforced by studies showing that analysis of non-residential energy consumption remains minimal despite the increasing availability of data[11], while technical barriers such as limitations in smart meter infrastructure further reduce the effectiveness of feature engineering[12]. In the Indonesian context, K-Means-based research has begun to be adopted for the segmentation of LPG customers and related sectors, such as the grouping of LPG and Aqua customers [13], product sales analysis [14], regional poverty mapping [15], selection of outstanding students [16], and segmentation of internet voucher sales customers [17]. Overall, the literature shows that the use of clustering algorithms in subsidized energy distribution is still not optimal due to methodological and data limitations, so an adaptive approach that considers customer characteristics and the structure of the subsidized 3kg LPG supply chain is needed. This study aims to fill this gap through the application of K-Means Clustering, which is designed to address the limitations of existing methods.

3. Reach Method

This study employs a quantitative descriptive design with a data-mining approach to analyze transactional records of subsidized 3-kg LPG customers. The research was conducted using secondary data obtained from official LPG distribution points, capturing variables such as purchase frequency, refill volume, transaction value, and temporal purchasing patterns. Data collection took place at the distribution site during the documented operational period, and the dataset represents the complete population of recorded LPG customers, with the sample defined through total sampling to ensure full representation of consumption behavior. The study adopts an observational, retrospective strategy consistent with previous energy-sector clustering research that relies on historical transaction logs and smart-meter records (Okereke et al., 2023; Rahmadhan & Wasesa, 2022).

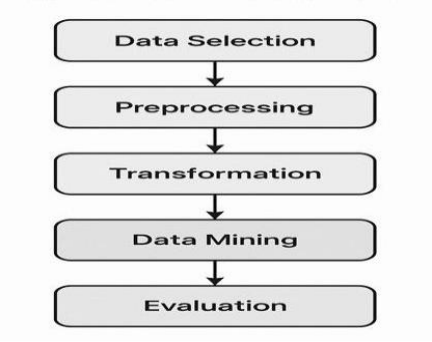


Fig. 1: Reach Procedure

Description of Research Methods Using KDD

1. Selection:
   Selecting transactional data of subsidized 3-kg LPG purchases sourced from official distributors, including variables such as refill frequency, number of cylinders purchased, transaction value, and purchase dates.
2. Preprocessing:
   Cleaning the dataset by removing duplicate entries, handling missing or inconsistent values, and filtering irrelevant attributes to ensure data quality before analysis.
3. Transformation:
   Converting categorical attributes (e.g., outlet name or region) into numerical form and normalizing all variables so they are suitable for the distance-based calculation required by the K-Means algorithm.
4. Data Mining:
   Applying the K-Means Clustering algorithm in RapidMiner to group LPG customers into clusters based on similarities in purchasing behavior, such as frequency, volume, and recency of transactions.
5. Evaluation:
   Assessing the resulting clusters using metrics such as the Davies–Bouldin Index and Silhouette Score, and interpreting whether the segmentation aligns with research objectives and reflects real customer consumption patterns.

3.1. Data Sources

The data sources for this study were obtained through a quantitative observational approach using secondary data from subsidized 3 kg LPG customer transaction records. Secondary data was chosen because it allows for objective analysis of energy consumption behavior based on digital traces generated by the official distribution system, thus eliminating dependence on respondent perceptions. In addition to providing broader data coverage, this data source also has a high level of accuracy because it is taken directly from the operational records of LPG agents and depots.

3.2. Population and Sample

The population in this study comprises all customers of subsidized 3-kg LPG cylinders recorded within the official distribution system during the study period. This population represents the entirety of transactional activities carried out by households or small enterprises that routinely purchase subsidized LPG through authorized agents and distribution outlets. Given that the research aims to analyze actual consumption behavior reflected in digital transaction records, the use of this complete population ensures comprehensive coverage of the consumption patterns under investigation.
The sample used in the analysis is determined through a *total sampling* approach, whereby all available transaction records are included as analytical units. This technique is appropriate because the dataset reflects the full set of customers served by the distributor, allowing the segmentation model to capture the heterogeneity of purchasing behaviors without sampling bias. The adoption of total sampling also aligns with the objective of producing a data-driven clustering model that accurately represents the empirical characteristics of the LPG 3-kg customer base.

3.3 Data Selection

Reading the sales transaction dataset using the Read CSV operator to read data from a CSV file containing 850 records. Attribute Selection Using the Select Attributes operator to select relevant attributes in the dataset. Label Assignment Using the Set Role operator to assign labels to the data to be used.

3.4 Data Analysis

The data analysis technique in this study uses a quantitative approach based on data mining by applying the K-Means Clustering algorithm to subsidized 3 kg LPG customer transaction data. This analysis aims to identify consumption patterns to produce accurate and relevant customer segmentation for distribution needs. The analysis process was conducted using RapidMiner Studio through a series of systematic stages covering data processing, cluster formation, and evaluation of the quality of the resulting clusters.

4. Results and Discussion

The data used in this study consisted of 801 transactions of 3 kg LPG sales recorded during the period of January 1 to December 31, 2024. The analysis was conducted to implement the K-Means algorithm in grouping customers based on purchasing patterns as a basis for formulating more effective distribution and sales strategies. The analysis process included data selection, a pre-processing stage to ensure data quality and suitability, and evaluation of the clustering results. Through this approach, customers were grouped based on similarities in consumption behavior so that PT Sumber Perkasa Mandiri could design more appropriate service strategies for each segment formed.

4.1. Data Selection

The initial stage in the Knowledge Discovery in Databases (KDD) process begins with data selection to ensure that the information used is relevant to the research objectives. In this study, data on 3 kg LPG sales transactions from PT Sumber Perkasa Mandiri was collected covering the period from January 1 to December 31, 2024. This data was used as the basis for grouping consumers based on their purchasing patterns.

Row No.	Id Registrasi	Nama Pang...	Kota	Kecamatan	Kelurahan	Alamat	Total Pemb...	Harga	jumlah
1	3270000000...	Tono	KABUPATEN...	TENGAH TANI	DESA MEGU...	SETU KULO...	736	16000	11776000
2	3270000000...	Wawan	KABUPATEN...	TALUN	DESA TRUS...	BLOK KEBO...	2511	16000	40176000
3	3220000000...	Tono	KABUPATEN...	KAPETAKAN	DESA TEGAL...	BLOK III ALB...	185	16000	2960000
4	3240000000...	Tika	KABUPATEN...	KALIWEDI	DESA TEGAL...	BLOK ASINA...	109	16000	1744000
5	3360000000...	Rudianto	KABUPATEN...	PANGURAGAN	DESA GESIK	BLOK KARA...	234	16000	3744000
6	3500000000...	Wawan	KABUPATEN...	SUMBER	DESA MEGU...	SETU KULO...	808	16000	12928000
7	3170000000...	Desi	KABUPATEN...	KALIWEDI	KELURAHAN...	BLOK SIKLIN...	135	16000	2160000
8	3750000000...	Anton	KABUPATEN...	KLANGENAN	KELURAHAN...	BLOK SETU ...	309	16000	4944000
9	3680000000...	Abdi	KABUPATEN...	DEPOK	DESA GOMB...	BLOK KALIA...	188	16000	3008000
10	3520000000...	Rina	KABUPATEN...	KALIWEDI	KELURAHAN...	JL. RAYA FA...	290	16000	4640000
11	3480000000...	Rizki	KABUPATEN...	KAPETAKAN	DESA ASTAP...	ASTAPADA B...	2245	16000	35920000
12	3520000000...	Aldo	KABUPATEN...	GUNUNG JATI	DESA MEGU...	BLOK WARE...	2389	16000	38224000
13	3480000000...	Nabila	KABUPATEN...	PANGURAGAN	DESA GEMP...	BLOK TENG...	1033	16000	16528000
14	3490000000...	Yuni	KABUPATEN...	DEPOK	DESA TEGAL...	BLOK ASINA...	95	16000	1520000
15	3350000000...	Edi	KABUPATEN...	JAMBLANG	KELURAHAN...	BLOK PEKU...	2293	16000	36688000
16	3150000000...	Fitri	KABUPATEN...	GEMPOL	KELURAHAN...	JL. FATAHILL...	2129	16000	34064000

ExampleSet (850 examples,0 special attributes,9 regular attributes)

Fig. 3: Data Set

In the data selection stage in RapidMiner, data on migrant workers in Cirebon Regency was imported from an Excel file (xlsx) using the Read Excel operator, which reads and loads data in spreadsheet format. This operator was selected based on the file type, as other options such as Read CSV.

.

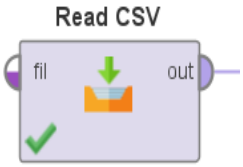
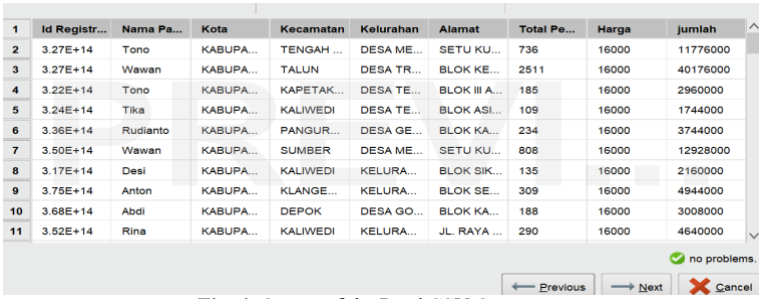


Fig. 3: Operator Read Cvs

After selecting the data to be entered using the Read Excel data import function, the selected data will be displayed. The attributes in the displayed data are nine attributes, namely Registration ID, Base Name, Subdistrict, Village, Address, Total Purchases, Price, and Quantity. Figure



	Id Registr...	Nama Pa...	Kota	Kecamatan	Kelurahan	Alamat	Total Pe...	Harga	jumlah
1	3.27E+14	Tono	KABUPA...	TENGAH ...	DESA ME...	SETU KU...	736	16000	11776000
2	3.27E+14	Wawan	KABUPA...	TALUN	DESA TR...	BLOK KE...	2511	16000	40176000
3	3.22E+14	Tono	KABUPA...	KAPETAK...	DESA TE...	BLOK III A...	185	16000	2960000
4	3.24E+14	Tika	KABUPA...	KALIWEDI	DESA TE...	BLOK ASI...	109	16000	1744000
5	3.36E+14	Rudianto	KABUPA...	PANGUR...	DESA GE...	BLOK KA...	234	16000	3744000
6	3.50E+14	Wawan	KABUPA...	SUMBER	DESA ME...	SETU KU...	808	16000	12928000
7	3.17E+14	Desi	KABUPA...	KALIWEDI	KELURA...	BLOK SIK...	135	16000	2160000
8	3.75E+14	Anton	KABUPA...	KLANGE...	KELURA...	BLOK SE...	309	16000	4944000
9	3.68E+14	Abdi	KABUPA...	DEPOK	DESA GO...	BLOK KA...	188	16000	3008000
10	3.52E+14	Rina	KABUPA...	KALIWEDI	KELURA...	JL. RAYA ...	290	16000	4640000

no problems.
Previous Next Cancel

Fig. 4: Output of the Read CSV Operator.

After the data was successfully imported, the *data selection* stage was carried out by eliminating irrelevant attributes through the *exclude attributes* process to ensure that only important variables were retained. Next, the *role* of each attribute is determined so that the data structure meets the needs of the analysis and can be processed optimally in the next modeling stage.

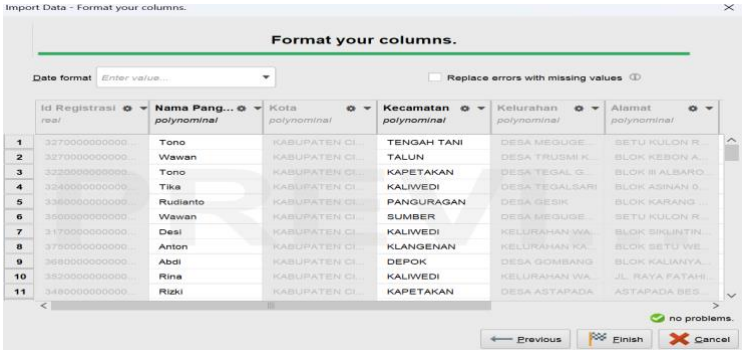


Fig. 5: Result Selection Data

The dataset used contains data on 3 kg LPG sales at PT Sumber Perkasa Mandiri in Cirebon Regency for the period January 1–December 31, 2024. The data includes serial numbers, customer IDs, customer names, administrative locations (city, subdistrict, village), total purchases, prices, and number of transactions. The K-Means algorithm was applied to group customers based on their purchasing patterns, thereby supporting the formulation of more targeted and efficient marketing strategies for the company.

4.2. Data preprocessing

This operator assigns roles to each attribute in the dataset according to its function in the clustering process. For example, certain attributes may be set as labels or identifiers, while others may be designated as the basis for clustering. By defining the role of each attribute, the clustering process can proceed in a more targeted and effective manner.

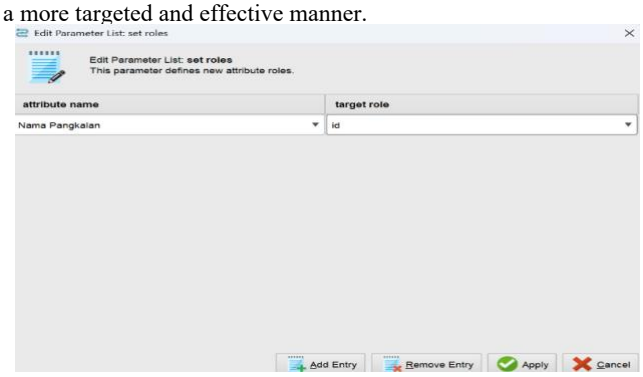


Fig. 6: Set Role Process.

4.3. Transformation

The transformation stage is carried out to convert polynomial attributes into numerical ones so that the data can be processed using distance-based calculations. This process is done through the Nominal to Numerical operator so that all variables are suitable for use in the K-Means algorithm.

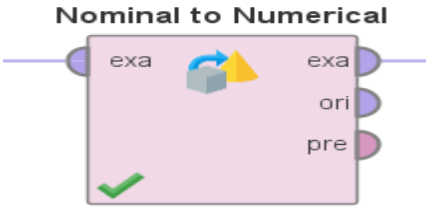
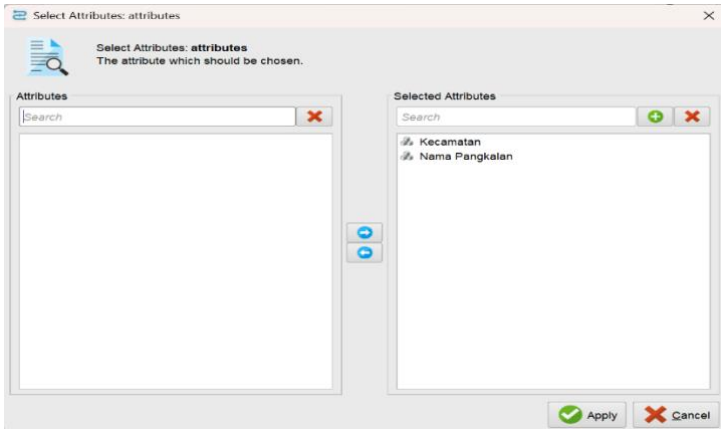


Fig. 7: Nominal to Numerical Operator.

Demonstrates the use of the Nominal to Numerical operator with specific parameter settings to convert the nominal operators "Base name" and "District" into numerical form as shown.



Fig, 8: Correction Set Role

Describes the attribute selection process in the Nominal to Numerical Transformation stage, where categorical variables are converted into numerical format so they can be used in machine learning analysis or modeling. At this stage, attributes such as District and Base Name are selected for conversion using the unique integers method, so that each category is represented by a unique numerical value. This transformation is an important part of the preprocessing process to ensure that nominal attributes can be processed effectively in distance-based algorithms such as K-Means.

Row No.	Nama Pang...	cluster	Kecamatan ...
1	Tono	cluster_0	1
2	Wawan	cluster_0	0
3	Tono	cluster_0	0
4	Tika	cluster_0	0
5	Rudianto	cluster_0	0
6	Wawan	cluster_0	0
7	Desi	cluster_0	0
8	Anton	cluster_0	0
9	Abdi	cluster_0	0
10	Rina	cluster_0	0
11	Rizki	cluster_0	0
12	Aldo	cluster_0	0
13	Nabila	cluster_0	0
14	Yuni	cluster_0	0

Fig. 9: Numeric Representation of the Name Column.

The image shows the results of converting the Base Name and Subdistrict attributes using the Nominal to Numerical operator, where categorical values have been converted into numerical representations. This transformation is necessary so that the data can be further processed in quantitative analysis and machine learning algorithms that require numerical input. After conversion, each category in the District attribute is represented by a unique number, thereby improving data compatibility for statistical calculations, pattern analysis, and distance-based clustering such as K-Means.

4.4. Data Minig

This study applies the K-Means algorithm to segment 3 kg LPG customers at PT Sumber Perkasa Mandiri based on transaction patterns. After going through the preprocessing and data transformation stages, the K-Means algorithm is used to group customers based on similarities in purchase frequency and number of transactions. The results of the grouping produce several customer segments that can be used as a basis for formulating more targeted marketing strategies that are in line with market characteristics.

Fig. 10: K-Means Algorithm.

4.5 Evaluation

This study uses the Cluster Distance Performance operator to evaluate the quality of clustering results. This operator calculates the Davies–Bouldin Index (DBI) and the average distance between data and centroid, enabling objective assessment of the proximity and separation of clusters produced by the K-Means algorithm.

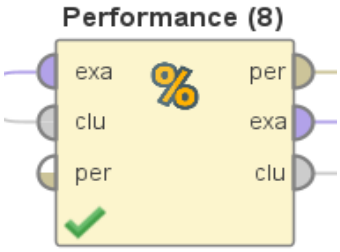


Fig. 11: Cluster Distance Performance Operator.

The clustering process is displayed as a series of stages involving the addition of the Multiply operator to produce variations in the number of clusters. Each K scenario, ranging from K = 2 to K = 10, is evaluated using the Performance operator to assess the quality of the grouping. This evaluation aims to determine the number of clusters that best suits the characteristics of the data. During the analysis stage, the quality of each cluster is measured using the Davies–Bouldin Index (DBI) metric, where the DBI value for each variation of K is presented in

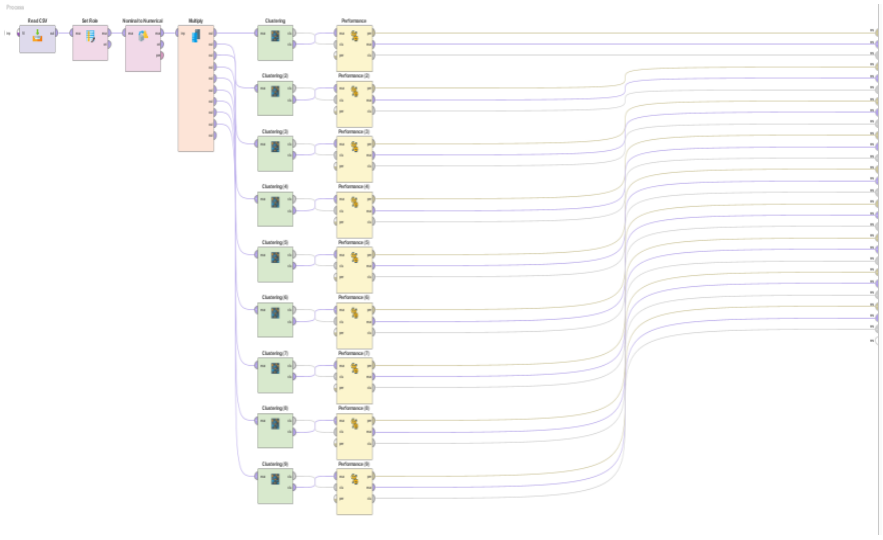


Fig. 12: Sub Process Operator.

4. Evaluation

The results of grouping migrant workers based on subdistricts and countries of destination in Cirebon Regency were analyzed using RapidMiner with a variation in the number of clusters from K = 2 to K = 10. The Davies–Bouldin Index (DBI) value for each K is shown in the table to identify the most optimal cluster. Based on the evaluation results, the lowest DBI value was obtained at K = 2 with a score of 0.057, so K = 2 was determined to be the most appropriate number of clusters.

Table 1: Davies-Bouldin Index (DBI) for Each Cluster

Nilai K	DBI
2	0.057
3	0.293
4	0.244
5	0.342
6	0.331
7	0.356
8	0.282
9	0.315
10	0.348

\

The cluster model results show that cluster 2 has a total of 844 items, and cluster 1 contains 6 items, bringing the total number of items to 850.

Cluster Model

```
Cluster 0: 844 items
Cluster 1: 6 items
Total number of items: 850
```

Fig. 13: Cluster Model K = 2

The results of the K-Means model performance evaluation are displayed through the Performance Vector, which measures the average distance within clusters (within-centroid distance) and the Davies–Bouldin index (DBI) as an indicator of cluster separation quality. The overall average distance was recorded at 4,269,323,594,872.39,8 with an average distance per cluster of 1,590,031,531.100 for Cluster 0, 407003805188521. 700 for Cluster 1, and 3230215664173586.500 for Cluster 2. The DBI value obtained was 0.057, indicating that the clusters formed had a good level of separation and a compact structure. Thus, the K-Means model with K = 2 is considered to produce optimal segmentation for analyzing 3 kg LPG purchase patterns at PT Perkasa Mandiri.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 426932359487239.800
Avg. within centroid distance_cluster_0: 407003805188521.700
Avg. within centroid distance_cluster_1: 3230215664173586.500
Davies Bouldin: 0.057
```

Fig. 14: Performance Vector.

Based on the within-centroid distance metric and the Davies–Bouldin Index (DBI) value, the quality of the clustering results can be assessed objectively. The within-centroid distance metric shows the average proximity of data in each cluster to its centroid, thus reflecting the degree of cluster density. Meanwhile, a DBI value of 0.057 indicates excellent separation between clusters. Considering these two indicators, it can be concluded that the applied clustering model produces a dense and well-separated cluster structure.

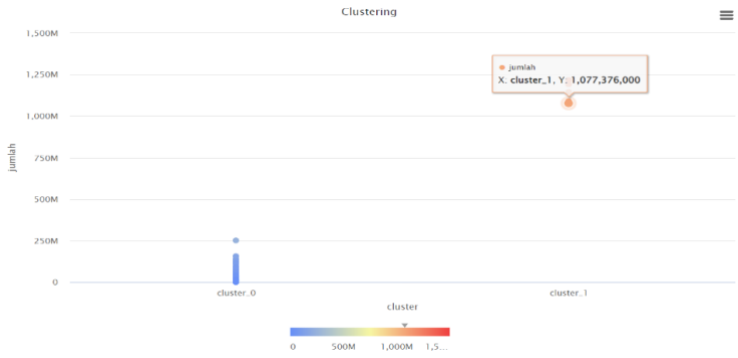


Fig. 15: Visualizations Cluster

5. Conclusion

Based on the results of applying the K-Means Clustering algorithm in segmenting 3 kg LPG purchases at PT Sumber Perkasa Mandiri, it can be concluded that this method is able to provide a more structured understanding of customer consumption patterns. The clustering process produced several clusters with different characteristics, which then became the basis for formulating more targeted marketing strategies. The first cluster represents customers with high purchase rates and strong loyalty, while the other clusters represent customer groups with low engagement and more sporadic purchasing patterns.

These results show that K-Means-based segmentation can be used to design more precise strategies, such as loyalty programs for premium customers, selective promotions for the middle segment, and a more intensive acquisition approach for clusters with low engagement levels. In addition to producing clearer segmentation, this analysis also helps identify specific needs and challenges in each cluster, ranging from service quality improvements to more effective promotional strategies. Thus, the use of K-Means Clustering has been proven to contribute to increased distribution effectiveness, market share expansion, and overall customer satisfaction.

References

[1] R. K. Gobel, "Equity and efficiency: An examination of Indonesia's energy subsidy policy and pathways to inclusive reform," *Sustainability*, vol. 16, no. 1, p. 407, 2024, doi: 10.3390/su16010407.

[2] R. Kalli, P. R. Jena, and S. Managi, "Subsidized LPG scheme and the shift to cleaner household energy use: Evidence from a tribal community of eastern India," *Sustainability*, vol. 14, no. 4, p. 2450, 2022, doi: 10.3390/su14042450.

[3] N. A. Pambudi *et al.*, "Renewable energy in Indonesia: Current status, potential, and future development," *Sustainability*, vol. 15, no. 3, p. 2342, 2023, doi: 10.3390/su15032342.

[4] S. K. Trivedi, A. D. Roy, P. Kumar, D. Jena, and A. Sinha, "Prediction of consumers refill frequency of LPG: A study using explainable machine learning," *Heliyon*, vol. 10, no. 1, p. e23466, 2024, doi: 10.1016/j.heliyon.2023.e23466.

[5] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[6] H.-H. Zhao, X.-C. Luo, R. Ma, and X. Lu, "An extended regularized K-means clustering approach for high-dimensional customer segmentation with correlated variables," *IEEE Access*, vol. 9, pp. 48405–48412, 2021, doi: 10.1109/ACCESS.2021.3067499.

[7] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Appl. Soft Comput.*, vol. 113, p. 107924, 2021, doi: 10.1016/j.asoc.2021.107924.

[8] S. P. Nguyen, "Deep customer segmentation with applications to a Vietnamese supermarkets' data," *Soft Comput.*, vol. 25, pp. 7785–7793, 2021, doi: 10.1007/s00500-021-05796-0.

[9] K. Tabianan, S. Velu, and V. Ravi, "K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022, doi: 10.3390/su14127243.

[10] C. A. Gomea, "Customer clustering and reliability metrics in power distribution systems: A holistic approach," *Appl. Energy Syst.*, vol. 45, no. 2, pp. 178–193, 2023.

[11] L. J. Olivia, B. Rahayudi, and I. Cholissodin, "Optimasi Rute Distribusi Gas Lpg 3 Kg dengan Integrasi Algoritma K- Means dan Ant Colony Optimization pada Multiple Travelling Salesman Problem," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, pp. 947–955, 2022.

[12] J. J. Moreno Escobar, O. Morales Matamoros, R. Tejeida Padilla, I. Lina Reyes, and H. Quintana Espinosa, "A comprehensive review on smart grids: Challenges and opportunities," *Sensors*, vol. 21, no. 21, p. 6978, 2021, doi: 10.3390/s21216978.

[13] E. R. Ningrum, A. Sanwidi, R. Akbarita, and M. N. H. Qomaruddin, "Optimasi Rute Pendistribusian Gas Elpiji Menggunakan Algoritma Floyd Warshall Dan Algoritma Greedy," *J. Ilm. Mat. Dan Terap.*, vol. 20, no. 1, pp. 1–14, 2023, doi: 10.22487/2540766x.2023.v20.i1.15568.

[14] I. Rosyadi, H. H. Kusumawardhani, F. A. Artanto, A. Alwan, A. Hardani, and F. Nafilaturrosyidah, "Clustering K-Means Dalam Pengelompokan PenjualanProduk Pada RTO Group," *Teknomatika*, vol. 13, no. 02, p. 3, 2023.

[15] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.

[16] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.

[17] Fahmi Naufal, Martanto, Arif Rinaldi Dikananda, and D. Rohman, "Analisis Segmentasi Pelanggan Voucher Wifi Dengan Metode K-Means," *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 81–90, 2025, doi: 10.51401/jinteks.v7i1.5169.