Journal of Applied and Research Computer Science and Information Systems
Vol. 2, No. 2, December 2024, pp. 191~202
ISSN: 2988-294X, DOI: 10.61098/jarcis.v2i2.188

Customer Segmentation of Cash Management System Using K-Means Clustering

Rizki Hesananda[1(✉)], Patria Apriliga[2]
[1]Universitas Siber Indonesia, Jakarta, Indonesia
[2] Bank Rakyat Indonesia, Jakarta, Indonesia
hessananda@cyber-univ.ac.id , patriaapriliga@gmail.com

Article Info

Article history:

Received October 27, 2024
Revised November 26, 2024
Accepted December 23, 2024

Keywords:

Cash Management System
Clustering
K-Means
Customer Segmentation
Banking Industry

ABSTRACT

The effective financial management is essential for running successful business operations. In the banking context, the Cash Management System (CMS) facilitates real-time, automated transaction management. PT Bank Rakyat Indonesia (Persero) Tbk., as one of Indonesia's largest banks, has implemented CMS since 2009. Despite its benefits, challenges persist, such as customer transactions outside regular working hours and difficulties in segmenting customers based on transaction volume and frequency. This study aims to address these issues by clustering BRI CMS users using the K-Means Clustering method, following the CRISP-DM framework. The research utilized transaction data of 2,727 users from January 2021 to April 2022. Data preparation involved cleaning anomalies and converting non-numeric values to numeric formats. Using the Elbow method, the optimal number of clusters was determined, resulting in three distinct user segments. The clustering revealed actionable insights, such as identifying high-value customers for targeted marketing and improving service strategies. This research offers a novel application of K-Means Clustering and CRISP-DM to CMS data management, contributing to better customer segmentation and strategic decision-making. These findings can help banks optimize resources, improve customer satisfaction, and enhance overall transaction efficiency.

Corresponding Author:

Rizki Hesananda
Universitas Siber Indonesia, Jakarta, Indonesia
Email: hessananda@cyber-univ.ac.id

1. INTRODUCTION

Effective and efficient financial management is the main key in running successful business operations [1]–[3]. In the context of banking, Cash Management System (CMS) is one of the services designed to assist customers, especially companies, in managing customer financial transactions in real-time and automatically [4] [5]. PT Bank Rakyat Indonesia (Persero) Tbk. as one of the largest banks in Indonesia, has launched CMS since October 2009, offering various benefits such as e-banking access, automated transactions, guaranteed security, and other excellent features [6] [7] [8]. However, although BRI's CMS offers various conveniences, there are challenges in managing CMS users. Many users conduct transactions outside the bank's working hours, which are between 08.00 to 15.00 WIB, from Monday to Friday. This causes obstacles in providing optimal service by bank officers, especially in terms of BRI's special CMS helpdesk [9] In addition, the marketing party (RM Dana) faces difficulties in grouping users based on the volume and number of customer transactions [10]. This grouping is important to determine customers who need follow-up to increase the volume and frequency of customer transactions [11] [12].

This research aims to cluster BRI CMS users based on the volume and number of transactions using the K-Means Clustering method. By clustering, banks can gain deeper insight into the characteristics of CMS users, so that they can develop more effective marketing strategies and improve services to customers [13][14][15]. To achieve these goals, this research uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) model. CRISP-DM is a process model that is widely used in data mining because it provides a clear and systematic framework [16][17][18]. This model includes six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [19][20][21]. By using the CRISP-DM, this research can be conducted in a structured and systematic manner, ensuring that each phase is analyzed and implemented properly.

The selection of clustering methods, specifically K-Means Clustering, is based on the need to effectively cluster CMS user data. K-Means is one of the most commonly used clustering algorithms due to its simplicity and ability to handle large datasets efficiently [22][23]. This algorithm works by dividing data into a number of clusters based on the distance between data, so as to group users who have similar characteristics. This research uses BRI CMS user transaction data from January 2021 to April 2022, which includes 6 columns and 2,727 rows of data. Through the steps in CRISP-DM, this research aims to identify groups of CMS users who have similar characteristics in terms of volume and number of transactions.

This research has several aspects of novelty. First, the use of the CRISP-DM model in the context of managing BRI CMS customers is an approach that has not been widely applied in previous studies. Second, the application of K-Means Clustering to group CMS users based on the volume and number of transactions is an innovation in identifying more targeted customer segments. This research is important nowadays because with the increasing digitalization and complexity of financial transactions, banks need to have a better strategy in managing customers [24][25]. By identifying user groups based on customer transaction behavior, banks can provide more personalized and efficient services, improve customer satisfaction, and optimize resources [26][27][28]. In addition, the results of this study can help marketing (RM Dana) in determining which customers need to be followed up, so as to increase the overall volume and number of transactions.

Overall, this research not only contributes to Bank Rakyat Indonesia, but also provides insights for other banking institutions that want to optimize the use of CMS and improve services to customers. Thus, this research has important value in developing a more effective and efficient customer management strategy in the digital era.

2. METHOD

This research uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) model which consists of six main stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment as shown in Figure 1. The following is an explanation of the steps in the CRISP-DM method applied in this research:.

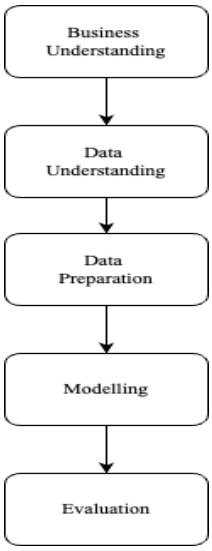

Figure 1. Research Flow Diagram

1. Business Understanding

At this stage, the main objective is to understand the business context and determine the research objectives. In this study, researchers found that many users of BRI's Cash Management System (CMS) make transactions outside of the bank's working hours, which are between 08.00 to 15.00 WIB from Monday to

Friday. This causes obstacles in optimal service by bank officers. The marketing party (RM Dana) also has difficulty in categorizing users based on the volume and number of transactions. Therefore, the purpose of this research is to group BRI CMS users based on the volume and number of transactions using the K-Means Clustering method to assist banks in determining customers who need to be followed up to increase the volume and frequency of customer transactions.

2. Data Understanding

This stage involves data collection and an in-depth understanding of the data to be used. The data used in this research is BRI CMS user transaction data from January 2021 to April 2022. At this stage, researchers conducted interviews with the bank to explore the research customer dataset. It was found that there are several companies with anomalous volume data, namely companies that have volume amounts above the average of other companies. These anomalies need to be identified and analyzed further to ensure the relevance of the cluster to be formed.

3. Data Preparation

This stage includes all the activities required to build the final dataset that will be processed in the modeling stage. The data selected are the relevant attributes of company ID, transaction amount, and transaction volume. The next step is to ensure that the dataset has no anomalous data. It was found that there was string data in the volume attribute which should only contain numeric data, so the string data was converted to numeric data. In addition, anomalous data that had been identified in the previous stage was removed from the dataset to ensure the accuracy of the clustering results.

4. Modelling

At this stage, modeling technique using the K-Means algorithm is selected and applied a clustering. K-Means Clustering is an algorithm that determines a set of k clusters and assigns each instance to create one cluster. Clusters consist of similar instances based on a measure of distance between customers. The algorithm starts with k points chosen as the centroids of k potential clusters. These starting points can be randomly drawn from the input samples or determined by the k-means heuristic. Next, all samples are assigned to the nearest cluster, and the cluster centroid is recalculated by averaging all samples in a cluster. This process is repeated until the centroid no longer moves or the maximum optimization step is reached. In this research, In this study, the best number of clusters is determined using the elbow method.

5. Evaluation

After the model is built, an evaluation phase is conducted to ensure that the model meets the business and research objectives. The evaluation is done by analyzing the clustering results and ensuring that the resulting groupings are in accordance with the research objectives. The clustering results are analyzed using the Davies-Bouldin Index (DBI) to determine whether the clusters formed are good for use as data segmentation. By following the CRISP-DM stages until the evaluation stage, this research is expected to provide systematic and good results in grouping BRI CMS users, so that it can assist banks in developing more effective and efficient strategies for managing customers.

3. RESULTS AND DISCUSSION

After conducting according to the method mentioned in the previous section, the following are the results and analysis of this research:

3.1. Business Understanding

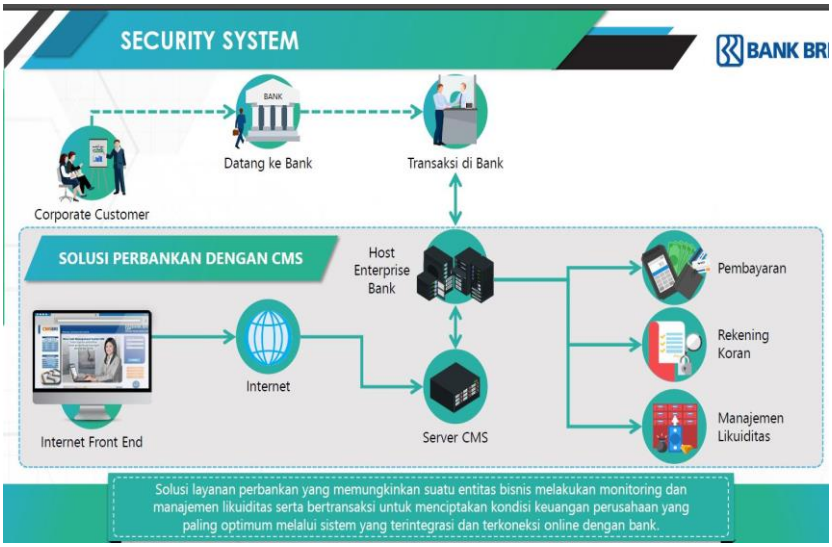

Figure 2. Cash Management System Workflow

At the business understanding stage, it was found that many users of BRI's Cash Management System (CMS) made transactions outside the bank's working hours, namely between 08.00 to 15.00 WIB from Monday to Friday. In Figure 2 contain the workflow of Cash Management System. This resulted in many users not receiving optimal service by bank officers. In addition, the marketing department (RM Dana) has difficulty in grouping users based on the volume and number of transactions. This study aims to group BRI CMS users based on the volume and number of transactions using the K-Means Clustering method to assist banks in identifying customers who need to be followed up to increase the volume and frequency of customer transactions.

3.2. Data Understanding

In the data understanding stage, a dataset containing 6 columns and 2,727 rows of BRI CMS user data from January 2021 to April 2022 was analyzed. The columns in the dataset include user ID, company name, industry type, transaction date, transaction amount, transaction volume, and several other attributes that describe the user's profile and transaction activity. The data can be seen in Table 1 as follows:

Table 1. CMS User Dataset

No	Corporate Name	Corporate ID	Brach Unit Recommender	Volume	FBI	TRX
1	CN-1	CID-1	BUR-1	4.925.335.300	-	800
2	CN-2	CID-2	BUR-2	3.880.430.175	-	687
3	CN-3	CID-3	BUR-3	-	-	-
4	CN-4	CID-4	BUR-4	6.319.158.230	-	136
5	CN-5	CID-5	BUR-5	-	200.000	1
…	---	---	---	---	---	---
3635	CN-3635	CID -3635	BUR -3635	-	1.000.000	5
3636	CN-3636	CID -3636	BUR -3636	-	-	-
3637	CN-3637	CID -3637	BUR -3637	-	3.400.00	17
3638	CN-3638	CID -3638	BUR -3638	-	-	-
3639	CN-3639	CID -3639	BUR -3639	2.871.206.369	2.429.000	106

Description:

1. Corporate Name : Active Company Name
2. Corporate_ID : ID Perusahaan Aktif
3. Brach Unit Recommender : Brach Unit Recommender
4. Volume : User Transaction Volume (IDR)
5. FBI : Fee Based Income
6. Trx : Total User Transactions (IDR)

Active Company Name is the name given to an active and operating company. The company has a Corporate_ID that serves as a unique identification to distinguish the company from others. The Initiating Work Unit (Recommending KC) is the work unit responsible for the development and implementation of the

company's strategy. They play an important role in ensuring the success of the company's operations. User Transaction Volume (IDR) is the number of transactions conducted by users in IDR. This figure shows the transaction activity carried out by users and provides an overview of the company's revenue potential. Fee Based Income (FBI) is the revenue earned from fee-based user transactions. This revenue comes from fees charged to users for each transaction made. Total User Transactions (Rupiah) is the number of transactions made by users in rupiah. This figure provides an overview of the number of transactions made by users and provides information about the company's revenue potential.

In the initial analysis, it was found that some companies have anomalous volume data, i.e. companies that have volume amounts above the average of other companies. These anomalies were identified as they could affect the clustering results and needed to be addressed at the data processing stage.

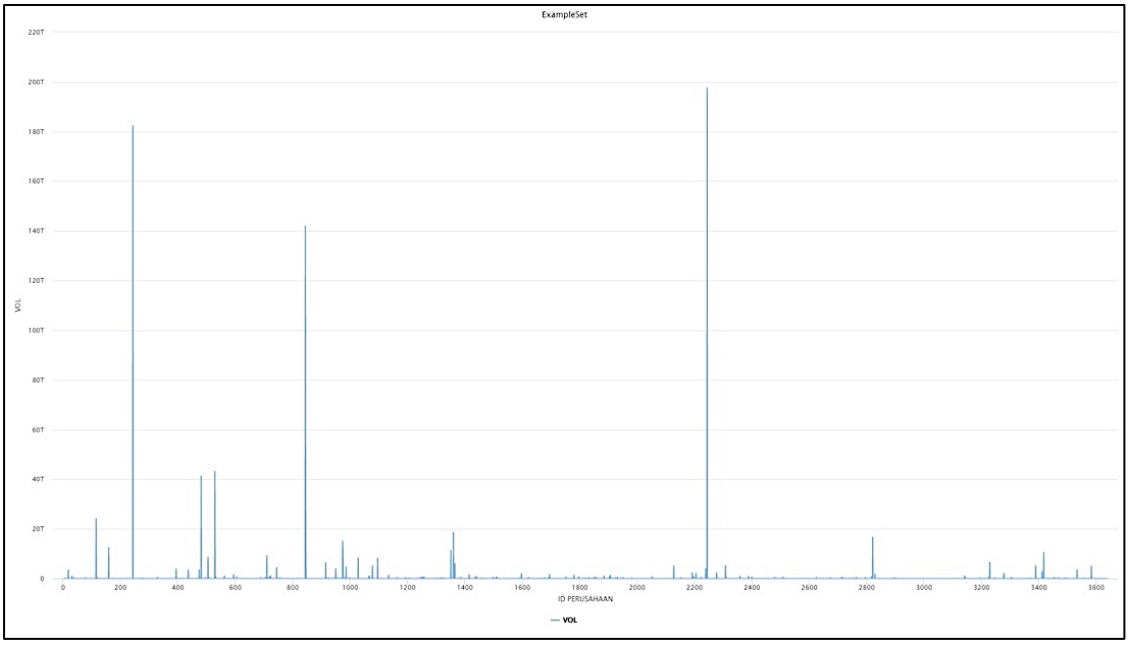

Figure 3. Volume Column Visualization

Figure 3 shows that some companies have anomalous volume data. The anomaly in question is that the company has a volume above the average of other companies. This can cause the cluster that will be formed by the k-means algorithm to be irrelevant, therefore in the data preparation stage the anomalous data needs to be removed from the dataset.

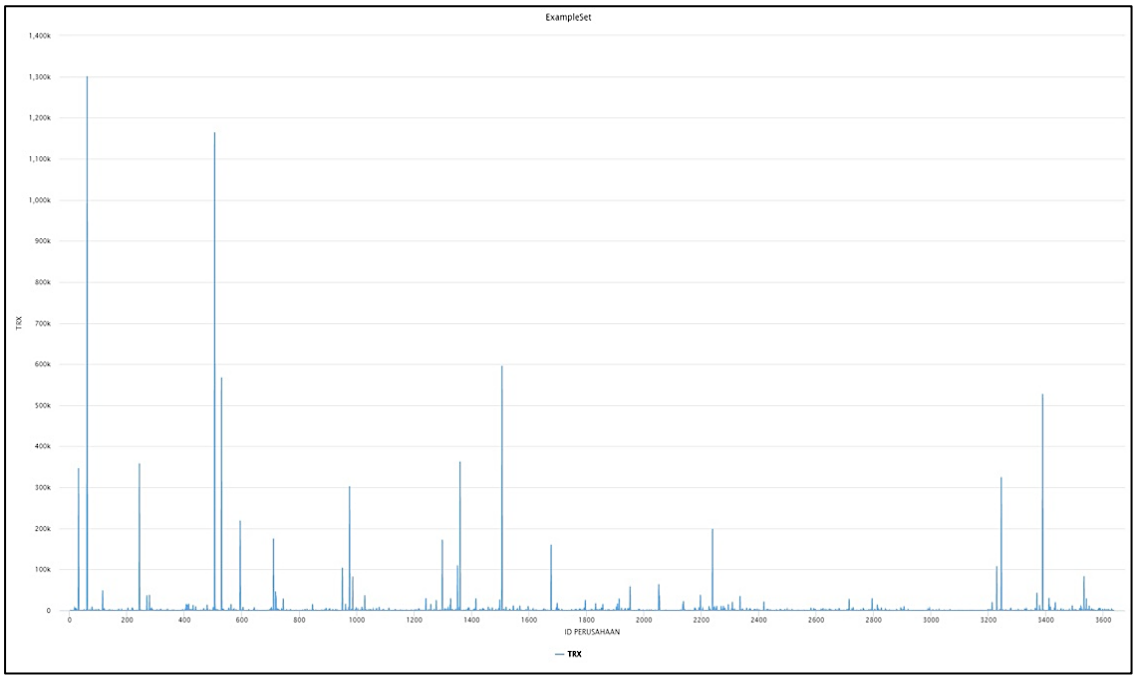

Figure 4. Transaction Column Visualization

Figure 4 shows that a number of companies have abnormal transaction data. Companies have more transactions than average. This may cause the clusters created by the k-means algorithm to become irrelevant. As a result, such anomalies should be removed from the dataset at the data preparation stage. Anomalies occur because some companies conduct a very large number of transactions when compared to the average transactions of other companies.

3.3. Data Preparation

In the data processing stage, from the 6 available attributes, relevant attributes were selected for clustering analysis, namely company ID, number of transactions, and transaction volume. Data with anomalies were identified and removed to ensure the accuracy of the clustering results. The data can be seen in table 2 as follows:

Table 2. Dataset after Attribute Selection

No	Corporate ID	Volume	TRX
1	CID-1	4.925.335.300	800
2	CID-2	3.880.430.175	687
3	CID-3	0	0
4	CID-4	6.319.158.230	136
5	CID-5	0	1
---	---	---	---
6	CID -3635	0	5
7	CID -3636	0	0
8	CID -3637	0	17
9	CID -3638	0	0
10	CID -3639	2.871.206.369	106

In addition, volume attributes that had string data were converted to numeric data to facilitate the clustering process. The final dataset used for modeling consists of three main columns: Company ID, transaction amount, and transaction volume. In Figure 5 and Figure 6 are the visual differences of the data before and after the anomalous data is removed. The anomalous data removed are companies that have a transaction volume of more than 2,000,000 and have more than 500 transactions. In Figure 5, we can see a diagram of data that has not been normalized:

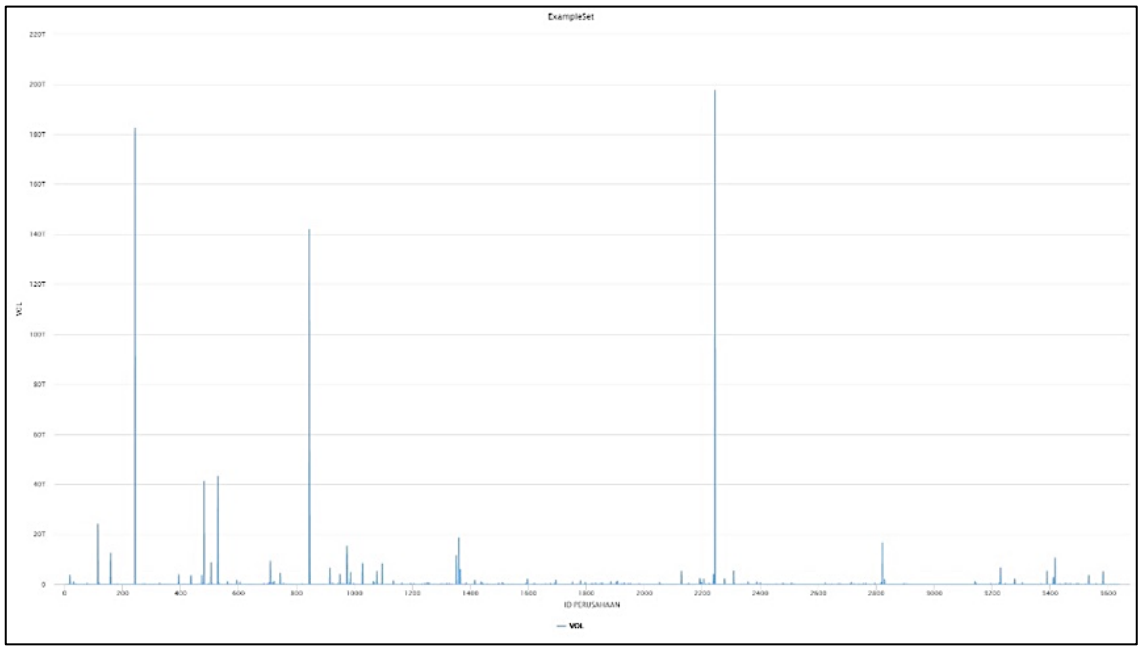

Figure 5. Dataset Visualization Before Deleted Anomaly

After removing the anomalous data, the distribution of volume and transaction amounts in the dataset becomes more normalized. This is important to get an overview of the division of clusters that are more in line with customer segmentation in the field. The remaining volume distribution in the current dataset is volume from 0 to 2,000,000 rupiah and the number of transactions is 0 to 500 transactions. Visualization of the dataset that has removed the anomaly is shown in Figure 6.

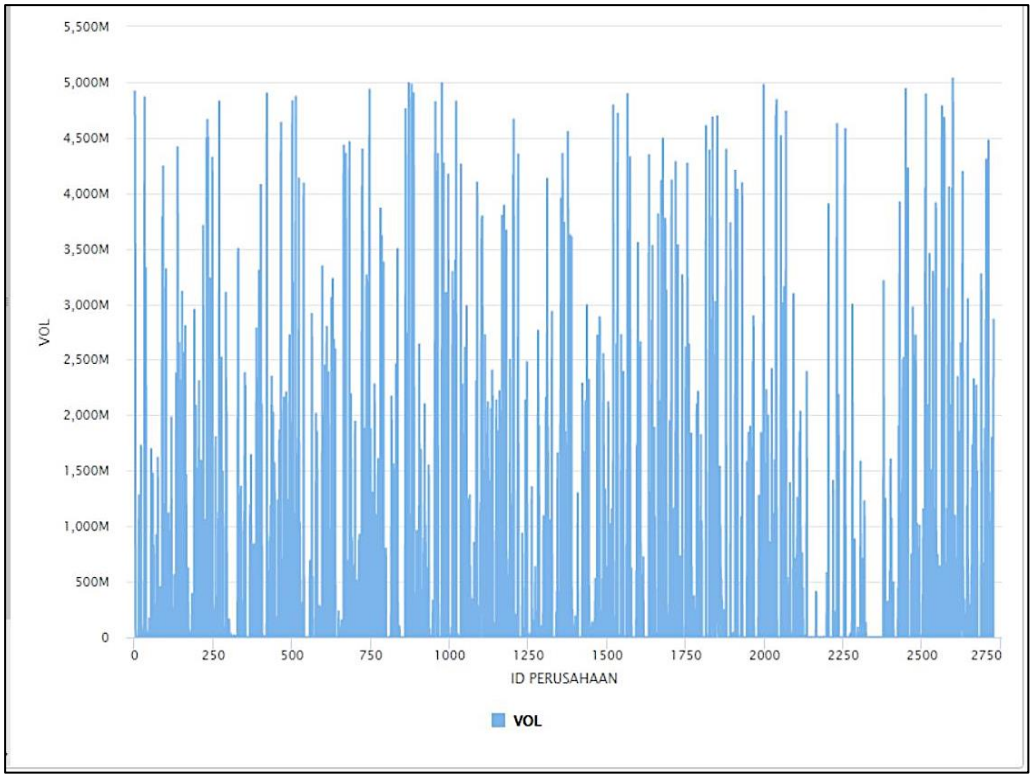

Figure 6. Dataset Visualization After Deleted Anomaly

3.4. Modelling

Modeling is a crucial phase in the data mining process, where the insights derived from data preparation and exploration are used to develop predictive or descriptive models. The goal of this stage is to apply appropriate techniques that can accurately represent the underlying patterns within the data. In this study, the focus is on applying the K-Means Clustering algorithm to segment companies based on specific characteristics. The selection of the optimal number of clusters is critical to ensure meaningful and actionable groupings. By leveraging tools such as the Elbow method, we aim to identify the best cluster configuration that reflects the natural divisions within the data. The subsequent sections detail the modeling approach, the rationale behind the choice of clusters, and the interpretation of the resulting segments.

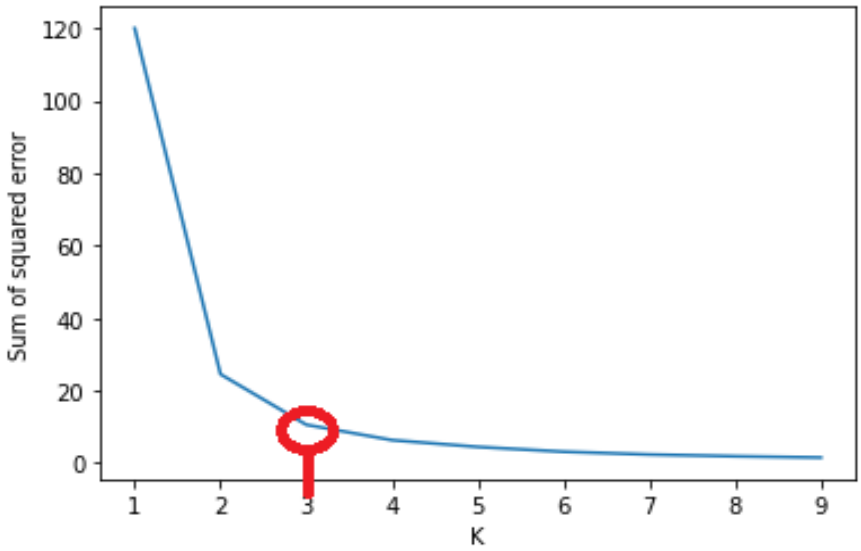

Figure 7. Elbow Method for Determine K

Based on several trials, the number of clusters of 3 was selected as showed in Figure 7. This indicates that the company segmentation is divided into 3 characteristics represented by 3 clusters. The distribution in cluster 0 is 2,274 companies, in cluster 1 there are 163 companies and in cluster 2 there are 220 companies. This clustering result can be seen in Figure 8.

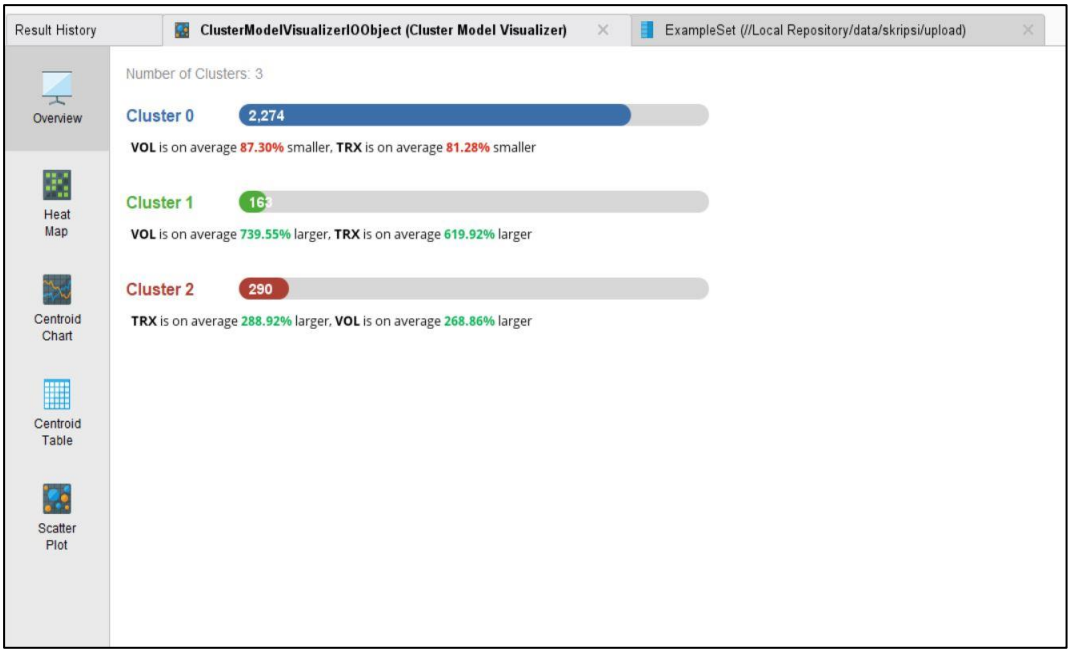

Figure 8. K-Means Cluster Result

The K-Means algorithm was used to determine three clusters based on the volume and number of transactions. The selection of the number of clusters (k = 3) is based on the analysis that this division will produce the optimal group. The following are the results of clustering using K-Means Clustering:
1. Cluster 0 (Blue Color): This cluster has a total transaction volume of 59,286,231.989 with 17,012 transactions.
2. Cluster 1 (Red Color): This cluster has a total transaction volume of 3,918,738,416,798 with 654,276 transactions.
3. Cluster 2 (Green Color): This cluster has a total transaction volume of 1,721,706,603,697 with 353,455 transactions.

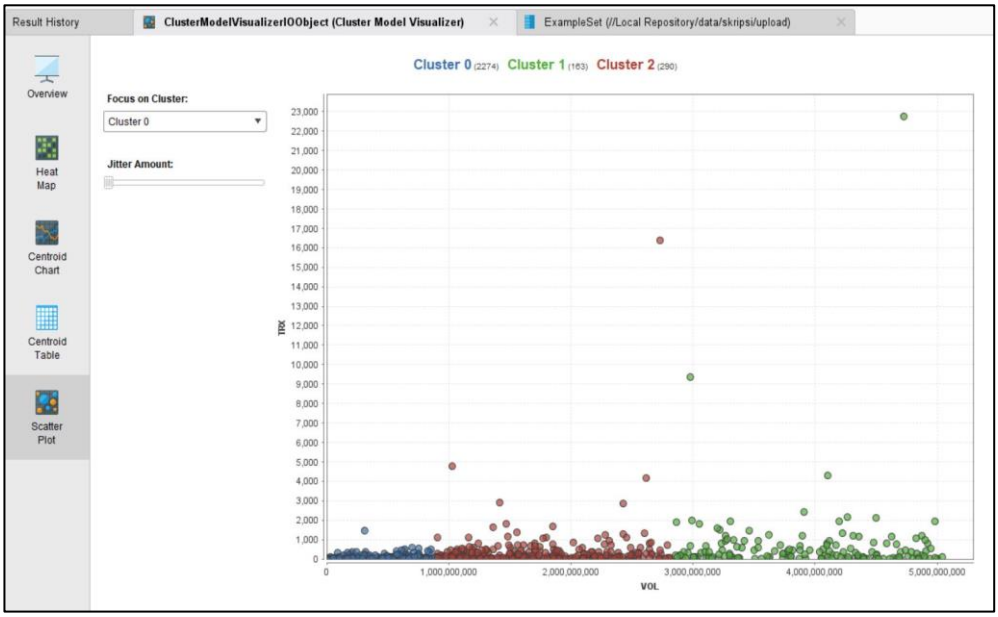

Figure 9. Scatter Plot Cluster Result

The results of this clustering are depicted in the scatter plot in Figure 9, where cluster 0 tends to have a low volume and number of transactions, cluster 2 has a medium volume with a relatively high number of transactions, and cluster 1 has the highest volume and number of transactions.

The interpretation of the results of the clustering that has been done is that companies that are members of cluster 0, are companies that do a lot of transactions and volume, so RM does not need to do any follow-up; just maintain the existing performance. Next is the company in cluster 1, this company is a company that needs to be followed up by RM because it has problems in the volume or number of transactions using CRM, there could be competition from other banks or difficulties in using CRM applications. The last is the company in cluster 2, the company in this cluster has very minimal transactions and the transaction volume is also very low. RM can follow up with a more extra approach such as visiting the company directly or offering more attractive benefits so that the company can increase the number and volume of transactions using BRI CRM.

3.5. Evaluation

the results section, the evaluation of the clustering model is carried out through a detailed comparison of the Davies-Bouldin Index (DBI) across various numbers of clusters. The DBI serves as a key metric to assess the compactness and separation of clusters, providing insight into the quality of the clustering configuration. By comparing the DBI values for different cluster numbers, we can identify the optimal clustering solution that best represents the underlying structure of the data. The comparison of DBI showed in Table 3.

Table 3. Comparison of DBI for Each K Numbers

K (Number of Clusters)	DBI
1	N/A
2	0,651
3	0,303
4	0,358
5	0,454
6	0,509
7	0,556
8	0,603
9	0,704
10	0,757

DBI is not applicable in K=1 because there's only one cluster. K=3 is the lowest DBI (0,303), indicating this is the best clustering choice. K=2, 4, 5, and others DBI increases as the number of clusters deviates from the optimal number, indicating worse clustering quality. Table 3 shows that as the number of clusters increases or decreases from the optimal value (K=3), the DBI value increases, reflecting less optimal clustering.

4. CONCLUSION

This study successfully clusters users of the Cash Management System (CMS) of Bank Rakyat Indonesia (BRI) using the K-Means algorithm based on the volume and number of transactions. Based on evaluation, 3 Cluster is the most optimal Cluster for the model; Its have 0,303 in DBI Score. The clustering results identified three main groups: users with high transaction volume and frequency, users with medium volume and high transaction frequency, and users with low transaction volume and frequency. The findings provide valuable insights for BRI to optimize marketing strategies and improve customer service. The application of the K-Means Clustering method in this study shows that data-driven customer segmentation can assist banks in identifying customer groups that need more attention, so as to increase transactions and volume through more targeted services. In addition, this research also shows that the use of the CRISP-DM approach in data analysis provides a systematic and effective framework for dealing with complex business problems.

However, this study has some limitations. First, the data used only covers the period from January 2021 to April 2022, so the results may not fully illustrate long-term transaction patterns. Second, this study only uses two main variables, namely transaction volume and amount, which may not cover all important aspects of customer behavior. Third, anomalous data omitted from the dataset may affect the clustering results and the potential insights that can be gained. For future research, it is recommended that using data with a longer period to get a more comprehensive picture of customer transaction patterns. Adding other relevant variables, such as transaction type, transaction time, and frequency of service use, to produce more detailed segmentation. Using other clustering methods or combining several methods to compare results and improve segmentation accuracy. Conduct further analysis of the factors that influence changes in customer transaction patterns in order to provide more appropriate recommendations for improving bank services. With these suggestions, it is hoped that future research can make a greater contribution to understanding and improving customer management in the banking sector.

REFERENCES

[1] N. A. Helfasari, R. R. Gamayuni, and U. Syaipudin, "Cashless Banking and Financial Performance of Bank Rakyat Indonesia," in *ICEBE 2020: Proceedings of the First International Conference of Economics, Business & Entrepreneurship, ICEBE 2020, 1st October 2020, Tangerang, Indonesia*, 2021, p. 22.

[2] J. Carine, D. Hidayat, and E. Fernando, "Factor Analysis of Influence of Purchase Intention on the Sociolla Platform," *J. E-bus. Manag. Sci.*, vol. 1, no. 1, pp. 11–28, 2023, doi: 10.61098/jems.v1i1.40.

[3] S. A. Pratminingsih, E. Astuty, and V. Mardiansyah, "The influence of electronic word of mouth and brand image on buying decision," *J. E-bus. Manag. Sci.*, vol. 11, no. 3 Special Issue, pp. 995–1002, 2019.

[4] R. Rusmiyatun, A. Probowati, and M. A. Islam, "Cash Management System, Is It Increase Transparancy?," in *International Conference On Digital Advanced Tourism Management And Technology*, 2023, vol. 1, no. 2, pp. 679–684.

[5] A. Fadilah, A. Sokarina, and I. P. Lenap, "Revealing the Effectiveness of Government Intern Control

Systems in Cash Management for Fraud Prevention," *ALEXANDRIA (Journal Econ. Business, Entrep.*, vol. 4, no. 2, pp. 55–59, 2023.
[6] J. Jamaludin *et al.*, *Transformasi Digital Era Disrupsi Industri 4.0*. Yayasan Kita Menulis, 2022.
[7] G. E. Paparang, M. Wullur, and F. J. Tumewu, "ANALISIS PELAKSANAAN STANDAR MUTU PELAYANAN DI BANK RAKYAT INDONESIA (BRI) KANTOR CABANG," *Neraca J. Ekon. Manaj. dan Akunt.*, vol. 2, no. 4, pp. 134–149, 2024.
[8] R. Hesananda and E. Y. Agustian, *Generasi Z dan Data Mining: Panduan Klasifikasi Pinjaman Bank sebagai Data Analis Keuangan*. Penerbit NEM, 2024.
[9] T. Ismawanto, R. G. Setianegara, and S. Rahmani, "Pengaruh Kualitas Pelayanan Dan Kinerja Karyawan Frontliner Terhadap Kepuasan Nasabah PT Bank Rakyat Indonesia (Persero), Tbk Kantor Cabang Balikpapan Sudirman Unit Klandasan," *J. Bisnis Dan Kewirausahaan*, vol. 16, no. 1, pp. 1–11, 2020.
[10] T. A. Tristanto, N. Nugraha, I. Waspada, M. Mayasari, and P. S. Kurniati, "Sustainability performance impact of corporate performance in Indonesia banking," *J. East. Eur. Cent. Asian Res.*, vol. 10, no. 4, pp. 668–678, 2023.
[11] F. Johnson and D. Pastory, "The Influence of Cash Management on Financial Performance of Private Schools in Tanzania," in *Applied Research Conference in Africa*, 2022, pp. 854–865.
[12] P. C. Oranefo and E. Bennee, "CASH MANAGEMENT: AN EMPRICAL STUDY ON FINANCIAL PERFORMANCE OF SELECTED BANKS IN NIGERIA," *Adv. J. Manag. Account. Financ.*, vol. 8, no. 4, 2023.
[13] K. Golalipour, E. Akbari, S. S. Hamidi, M. Lee, and R. Enayatifar, "From clustering to clustering ensemble selection: A review," *Eng. Appl. Artif. Intell.*, vol. 104, p. 104388, 2021.
[14] T. M. Ghazal, "Performances of k-means clustering algorithm with different distance metrics," *Intell. Autom. Soft Comput.*, vol. 30, no. 2, pp. 735–742, 2021.
[15] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, 2022.
[16] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021.
[17] J. S. Saltz, "CRISP-DM for data science: strengths, weaknesses and potential next steps," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 2337–2344.
[18] R. Hesananda, *Algoritma Klasifikasi Bibit Terbaik untuk Tanaman Keladi Tikus*. Penerbit NEM, 2021.
[19] M. North, *Data Mining for the Masses, Third Edition*. 2018.
[20] V. Plotnikova, M. Dumas, and F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," *Data Knowl. Eng.*, vol. 139, p. 102013, 2022.
[21] S. Peker and Ö. Kart, "Transactional data-based customer segmentation applying CRISP-DM methodology: A systematic review," *J. Data, Inf. Manag.*, vol. 5, no. 1, pp. 1–21, 2023.
[22] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023.
[23] Y. Ren *et al.*, "Deep clustering: A comprehensive survey," *arXiv Prepr. arXiv2210.04142*, 2022.
[24] E. Ernawati, S. S. K. Baharin, and F. Kasmin, "A review of data mining methods in RFM-based customer segmentation," in *Journal of Physics: Conference Series*, 2021, vol. 1869, no. 1, p. 12085.
[25] M. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Inf. Syst. E-bus. Manag.*, vol. 21, no. 3, pp. 527–570, 2023.
[26] 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.
[27] J. Joung and H. Kim, "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews," *Int. J. Inf. Manage.*, vol. 70, p. 102641, 2023.
[28] H. Komatsu and O. Kimura, "Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types," *Energy Build.*, vol. 283, p. 112831, 2023.

.

BIOGRAPHIES OF AUTHORS

Rizki Hesananda – Corresponding Author

Rizki Hesananda is an IT industry practitioner and a computer science lecturer actively involved in teaching, research, and community service. With experience as a Web Developer across various sectors including freelancing, corporate companies, ministries, and start-ups, I have developed a deep understanding of different organizational IT needs. Early in my career, driven by curiosity about real-world IT, I pursued a Master's degree and have since worked on over 50 websites. My passion for learning and teaching has led me to explore emerging fields like Artificial Intelligence, particularly Data Mining and Computer Vision. I believe that staying updated and continuously enhancing one's skills are crucial in the rapidly evolving IT landscape.

Patria Apriliga - Author

Patria Apriliga has been a Funding Officer at Bank Rakyat Indonesia since 2017, bringing extensive experience in managing relationships with funding customers. With a bachelor's degree in Information Technology, Patria has leveraged her academic background to conduct research in data mining, which supports her work in the banking industry. Her expertise in both IT and banking enables her to effectively navigate and optimize financial operations, making her a valuable asset in the field.

.