J-Icon : Jurnal Informatika dan Komputer
Vol. 13 No. 1, March 2025, pp. 18~25
DOI: 10.35508/jicon.v13i1.12691

REVENUE SEGMENTATION FROM PAYMENT AGGREGATOR USING K-MEANS CLUSTERIZATION METHOD

Dimas Fahmi Suntoro[1*], Netty Fitriani[2], Arief Wibowo[3]

[1,2]Economic and Bussiness Faculty, Universitas Budi Luhur, Jakarta, Indonesia
[3]Information Technology Faculty, Universitas Budi Luhur, Jakarta, Indonesia
Email[1*]: 2231600632@student.budiluhur.ac.id
Email[2]: 2231600590@student.budiluhur.ac.id
Email[3]: arief.wibowo@budiluhur.ac.id

ABSTRACT

Customer habits during payment transactions can significantly influence the revenue of companies utilizing financial technology in selecting payment partners (payment aggregators). Customers are highly selective in choosing payment aggregators based on convenience, promotions, and benefits offered. This study aims to assist management in decision-making and business analysis by identifying patterns and similarities in revenue data for payment aggregators. Revenue grouping in payment aggregators is conducted using data mining techniques for clustering, employing the K-Means method. This method partitions revenue data into groups based on similarities. Testing with various cluster numbers, specifically k = 2 with DBI = 0.023; k = 3 with DBI = 0.209; and k = 4 with DBI = 0.116, and a maximum of ten iterations, revealed the best result with four clusters. The resulting clustering pattern categorizes payment types into four categories: Copper, Silver, Gold, and Platinum. The study findings indicate that payment methods such as CASH, MINIATM TRANSFER, and VA TRANSFER fall under the Platinum category, contributing the highest revenue at 10.78% of the total income. Based on this clustering pattern and the insights provided, companies should maintain features that support payment aggregators in the Platinum category, while for the Copper category, which generates the lowest revenue, business evaluations are needed to increase transaction frequency and value by adding merchants.

Keywords: Payment aggregator, Data mining, Clustering, RapidMiner

1. INTRODUCTION

Research Background

In the advancing digital era, data has become one of the most valuable assets for many organizations. Service providers often face challenges in categorizing the revenue generated through payment aggregators due to the large volume of transactions and diverse information. Implementing data mining techniques for clustering using the K-Means method can be an effective way to categorize revenue in payment aggregators [1]. Data mining for clustering aims to group data objects into clusters based on similarities in features or specific characteristics. The K-Means method is one of the most popular clustering methods widely applied across various domains. This method groups data into 'K' predetermined clusters based on the distance between data objects [2].

Previous research has analyzed customer data in payment aggregator transactions using the K-Means algorithm, resulting in clustering models of customer groups based on income and transaction amounts. This research utilized the WEKA application with the K-Means algorithm [3]. Another study analyzed sales transactions through bank payments using the K-Means algorithm combined with Particle Swarm Optimization, producing clusters of bank customer groups in bill payments [4].

In this study, the authors aim to apply the K-Means method to categorize revenue in payment aggregators. Payment aggregator revenue often originates from various sources such as e-commerce transactions, bill payments, money transfers, and other transactions. The K-Means method can uncover hidden patterns in revenue data, simplifying analysis and decision-making processes [1]. The objectives of this research include:

1. Identifying patterns and structures within revenue data in payment aggregators.
2. Evaluating the performance of the K-Means method in categorizing revenue in payment aggregators.
3. Providing valuable insights for companies to understand different revenue profiles.
4. This study aims to contribute to understanding the critical role of payment aggregators, the discovery of knowledge from data mining processes, the strength of clustering techniques, and the efficiency of

***) Corresponding Author**
Submitted : October 1, 2023
Accepted : October 24, 2024
Published : March 28, 2025

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

the K-Means algorithm. When combined, these components offer deep insights into revenue structures, business optimization, and customer-oriented strategies for companies.

2. METHODOLOGY

Payment Aggregator

A payment aggregator is an entity that provides electronic payment collection and processing services for merchants or sellers. It acts as an intermediary between customers, merchants, and financial institutions to facilitate online payment transactions [1].

The functions of a payment aggregator include receiving payments from customers on behalf of merchants or sellers and collecting payment information such as credit card details, digital payment methods, or bank transfers. Payment aggregators also process received payments by transmitting instructions to relevant financial institutions to transfer funds from customer accounts to merchant accounts.

One of the responsibilities of a payment aggregator is to ensure transaction security through the implementation of appropriate encryption protocols and security measures to protect sensitive customer data. Payment aggregators also monitor transactions to detect potential fraud or suspicious activities that could harm customers or merchants. Additionally, they provide transaction reports and financial information to merchants for reconciliation and financial reporting purposes. Payment aggregators integrate payment systems with e-commerce applications or merchant systems to facilitate smooth payment processes [1].

Several business models for payment aggregators involve charging additional fees on transactions to generate profit, commonly known as markup fees, which may be a percentage of the transaction amount or a fixed rate. Payment aggregators may also charge service fees to merchants as compensation for using their platform or infrastructure. In some cases, payment aggregators and merchants may have revenue-sharing agreements, where the aggregator receives a certain percentage of the revenue generated by the merchant.

The advantages of using payment aggregators include easy integration facilitated by the provision of application programming interfaces (APIs) that simplify integration with merchant systems, enabling them to quickly accept online payments. Payment methods are diverse, as payment aggregators provide access to various electronic payment methods such as credit cards, bank transfers, digital wallets, or carrier billing, thus expanding the potential customer base.

Data Mining

Data mining is the extraction of valuable knowledge or information from large and complex datasets [5]. The primary goal of data mining is to uncover hidden patterns, relationships, or trends within the data that can be used to make better decisions, understand consumer behavior, optimize business processes, and enhance overall performance [6]. The process begins with data pre-processing, which involves cleaning and transforming raw data before the data mining process. This stage includes handling missing values, addressing incomplete or invalid data, and normalizing data when necessary [7]. Exploratory Data Analysis (EDA) is then conducted as an initial step to understand the data, involving visualization, descriptive statistics, and exploratory analysis to identify initial patterns and trends [7].

Several data mining methods are utilized, such as clustering, classification, regression, association, and anomaly detection. Clustering groups data objects based on similarities in specific attributes, while classification creates models to predict labels or classes for new objects based on previously labelled observations. Regression is used to model the relationship between dependent and independent variables, enabling predictions of continuous values and understanding the impact of independent variables. Association uncovers relationships or associations between items in the data, and anomaly detection identifies unusual patterns or objects that deviate from the general data trends, helping detect rare events, potential fraud, or system disruptions [8].

The models generated through data mining are evaluated for reliability using appropriate metrics, such as accuracy, precision, recall, F1 score, or area under the curve (AUC) [6]. The results are then interpreted and visualized to understand the patterns, relationships, or trends identified, with graphs, charts, or other visualizations effectively conveying information to stakeholders [7]. Ultimately, the insights gained from data mining are used to make better decisions, develop business strategies, optimize operational processes, and identify new opportunities, providing valuable solutions to business challenges and leveraging emerging opportunities [8].

Clustering

Clustering is a data mining technique aimed at grouping data objects with similarities into clusters. Conversely, data objects without similarities are placed in different clusters. Clustering is used to identify patterns, structures, or relationships in data without class labels or prior supervision [9]. Several clustering methods are commonly applied in data analysis, including K-Means, hierarchical clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Gaussian Mixture Models (GMM) [10].

The K-Means method is among the most popular clustering techniques, dividing data into k predetermined clusters based on the distance between data objects. Hierarchical clustering builds a dendrogram to represent hierarchical relationships among data objects, using either agglomerative (starting with individual objects and sequentially merging them) or divisive (starting with one large group and splitting it sequentially) approaches. DBSCAN identifies clusters based on data density, grouping objects close in parameter space while considering isolated objects as noise or anomalies. GMM assumes that data within each cluster is generated from a Gaussian distribution and estimates the parameters of the Gaussian distribution for each cluster to predict the likelihood of a data object belonging to a specific cluster.

Evaluating clustering quality often involves metrics such as the Sum of Squared Errors (SSE), which measures the extent of data objects' proximity to their cluster centers, aiming to minimize the total squared error. The silhouette coefficient assesses how similar objects within the same cluster are compared to those in other clusters, ranging from -1 to 1, with higher positive values indicating better clustering quality. The Calinski-Harabasz index measures how well clustering separates clusters with low dispersion within each cluster.

Clustering has various applications across different fields. In customer analysis, it is used to identify customer profiles based on shopping behavior, preferences, or demographic characteristics. In market segmentation, clustering helps define distinct market segments based on consumer preferences, needs, or demographics. In bioinformatics, it is applied to group genetic samples or biological data to uncover patterns or related groups. Document clustering organizes documents by topics, content, or attribute similarities, while anomaly detection identifies unusual data, such as fraudulent activities or system disruptions.

Research Methodology

The research methodology employed in this study is clustering using the K-Means method. K-Means is one of the most popular clustering methods in data analysis. It is used to group data objects into clusters based on similarities in specific features or characteristics [11]. The general steps of the K-Means method are as follows [12]. First, during initialization, the number of clusters (k) to be formed is determined, and 'k' initial points are randomly selected as the initial cluster centers. Next, classification involves calculating the distance between each data object and the existing cluster centers. Following this, assignment allocates each data object to the cluster with the nearest center, where the distance can be measured using metrics such as Euclidean or Manhattan distance. Subsequently, cluster center updates are performed by recalculating the centers for each cluster, taking the average of all data objects assigned to that cluster. This process iterates, repeating the classification and assignment steps, until no changes occur in the data object assignments or until a predefined stopping condition is met, such as a fixed number of iterations or convergence of cluster centers. Finally, the clustering results are evaluated using metrics like the Sum of Squared Errors (SSE) to assess how closely data objects in a cluster are to their respective cluster centers, ensuring the quality of the clustering.

The advantages of the K-Means method include its relatively easy implementation and fast data processing capabilities. It is efficient when applied to large datasets and produces good results for clusters that are spherical in shape and have uniform sizes. However, K-Means also has certain limitations. It is dependent on random initialization, which can affect the clustering results. Additionally, the method is sensitive to outliers, which may influence the positions of cluster centers. K-Means is suitable only for numerical data and is not appropriate for categorical data or data with attributes of varying scales. The framework applied in this study is illustrated in Figure 1.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

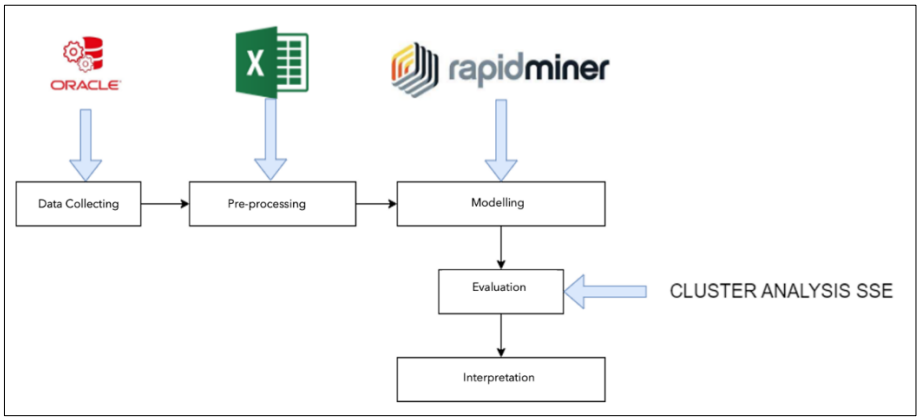

Figure 1. Research Framework

The steps in data processing using the K-Means method are as follows:

1. Data collection involves gathering revenue data from the payment aggregators that serve as the research objects. This data includes transaction counts (Trx Count), transaction amounts (Trx Amount), transaction types (Trx Type), and other relevant attributes [13].
2. Data preprocessing is conducted before the data can be used in the clustering process. This stage includes cleaning the data, handling missing values, and normalizing the data if necessary. An overview of the preprocessed payment aggregator data to be analyzed in this study is presented in Table 1.

Table 1. Pre-processing Results

No.	Payment Aggregator	Transaction Count (Freq)	Transaction Amount (Rp Million)
1.	Atome	1,659,186	1,250,010
2.	Cash	1,127,369	658,955
3.	CIMB QR	35,944	3,462
4.	Credit	2,463,870	15,174,600
5.	Debit	2,379,841	1,186,290
6.	Doku QRIS	5,534	1,355
7.	Ecomm	62,007	227,201
8.	Eligibility	98,761	8,911
9.	Emoney	864,954	1,929
10.	Gopay QR	2,409	1,764
11.	Installment	4,613	26,050
12.	Kredivo QR	114,520	64,852
13.	Miniatm Transfer	495,737	728,845
14.	Nobu QR	223,348	42,514
15.	Nobu Qr Dynamic	56,393	8,525
16.	Ovo Push To Pay	183,983	26,274
17.	Shopeepay QR	641,564	68,756
18.	Tcash QR	753,715	54,458
19.	VA Transfer	80,178	806,619
20.	Other	20,159	43,963

The implementation of the K-Means method will be applied to the pre-processed revenue data. Parameters such as the number of clusters will be determined based on the needs of the analysis. The clustering results will be evaluated using appropriate metrics, such as the Sum of Squared Errors (SSE) or cluster validity indices. The clustering results will then be analysed to identify patterns and revenue profiles. This research is expected to contribute to the payment industry by enhancing the understanding of revenue within payment aggregators. Insights gained from data mining clustering techniques can help payment aggregators optimize business strategies, improve decision-making processes, and better meet customer needs.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

3. RESULTS AND DISCUSSION

The critical stages of this study include Data Collection, Data Pre-processing, Determination of K Value, and Implementation of the K-Means Algorithm. Each phase is essential in uncovering hidden patterns, optimizing strategies, and making informed decisions.

1. Data Collection
   Data was collected from the database and required cleaning. The data presented in Table 1 underwent further inspection and processing to create the dataset. This dataset includes necessary information such as transaction status, transaction dates (month and year), payment aggregator names, transaction frequency, and transaction amounts.
2. Data Pre-processing
   At this stage, after obtaining a clean dataset, the relevant data was identified, including the collection time, the transaction status used, and the conversion of transaction types (payment aggregator names) into numerical values to allow processing in RapidMiner. The data included only successful transaction statuses recorded between January 2019 and June 2023.
3. Determination of K Value
   In the K-Means method, 'K' refers to the number of clusters to be formed in the data. The value of 'K' was predetermined as an input parameter for the K-Means algorithm. Each cluster aimed to represent distinct groups or categories in the data. In this study, the authors used 'K' as four, dividing the dataset into four distinct categories based on attribute or characteristic similarities. These categories are represented in Table 2.

Table 2. Clustering Categories

Cluster	Category
Cluster_0	Copper
Cluster_1	Silver
Cluster_2	Gold
Cluster_3	Platinum

Criteria for Determining Cluster Categories:

1. Cluster_0 (Copper): Transaction nominal values range from IDR 1.355 million to IDR 227.201 million, with transaction frequencies between 4,613 and 864,954.
2. Cluster_1 (Silver): Transaction nominal value is IDR 15,174.600 million, with a transaction frequency of 2,463,870.
3. Cluster_2 (Gold): Transaction nominal values range from IDR 1,250.010 million to IDR 1,186.290 million, with transaction frequencies between 1,659,186 and 2,379,841.
4. Cluster_3 (Platinum): Transaction nominal values range from IDR 658.955 million to IDR 806.619 million, with transaction frequencies between 80,178 and 1,127,369.

Implementation of the K-Means Algorithm

The study evaluated K values of 2, 3, and 4 to identify the optimal clustering configuration. At this stage, the focus was on the use of 'K' as four, which resulted in the formation of four clusters. In this context, RapidMiner was used to cluster the obtained data. The process diagram for clustering with RapidMiner is shown in Figure 2.

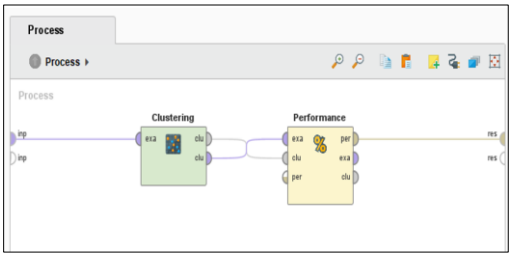

Figure 2. Modelling in Rapidminer Application

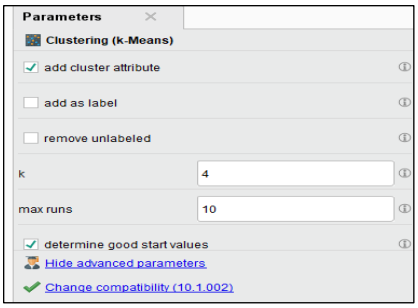

Figure 3. Rapidminer's Configuration

The dataset was already in CSV file format, containing data as agreed upon. In the clustering icon of RapidMiner, the researchers set 'K' to four with a maximum of ten iterations. The configuration of the clustering icon in RapidMiner is illustrated in Figure 3.

Previously, the researchers also used 'K' values of two and three to compare the results obtained with 'K' set to four, two, and three. The comparison of data obtained during testing with RapidMiner is shown in Table 3.

Table 3. Comparison of Testing Results with 'K' Using RapidMiner

K Value	Davies Bouldin Index
2	0.023
3	0.209
4	0.116

The Davies-Bouldin Index (DBI) is one method used to determine the optimal number of clusters. This approach aims to maximize the distance between clusters while minimizing the distance between objects within the same cluster. The ideal number of clusters corresponds to the minimum DBI value. As shown in Table 3, the researchers conducted three tests with k = 2, 3, and 4, using a maximum of ten iterations. The DBI results indicate that the smallest value was obtained for k = 2 (DBI = 0.023), while the largest value was for k = 3 (DBI = 0.209). Although a DBI value closer to zero is considered better, the researchers did not choose k = 2 because it resulted in too few categories. Therefore, k = 4, with a DBI of 0.116, was selected. Consequently, clustering with RapidMiner produced four categories.

The clustering process with RapidMiner resulted in four categories: cluster_0 (CL_0), cluster_1 (CL_1), cluster_2 (CL_2), and cluster_3 (CL_3) with details are presented in Table 4.

Table 4. Clustering Results from RapidMiner

Transaction Count (Freq)	Transaction Amount (Rp Million)	Transaction Type	Cluster
1,659,186	1,250,010	1	Cl_2
1,127,369	658,955	2	Cl_3
35,944	3,462	3	Cl_0
2,463,870	15,174,600	4	Cl_1
2,379,841	1,186,290	5	Cl_2
5,534	1,355	6	Cl_0
62,007	227,201	7	Cl_0
98,761	8,911	8	Cl_0
864,954	1,929	9	Cl_0
24,09	1,764	10	Cl_0
4,613	26,050	11	Cl_0
114,52	64,852	12	Cl_0
495,737	728,845	13	Cl_3
223,348	42,514	14	Cl_0
56,393	8,525	15	Cl_0
20,159	43,963	16	Cl_0
183,983	26,274	17	Cl_0
641,564	68,756	18	Cl_0
753,715	54,458	19	Cl_0
80,178	806,619	20	Cl_3

The transaction type codes based on payment aggregators are presented in Table 5.

Table 5. Transaction Type Codes by Payment Aggregator

Transaction Type	Aggregator
1	Atome
2	Cash
3	CIMB QR
4	Credit
5	Debit
6	Doku QRIS
7	Ecomm
8	Eligibility
9	Emoney
10	Gopay QR
11	Installment
12	Kredivo QR

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

13	Miniatm Transfer
14	Nobu QR
15	Nobu Qr Dynamic
16	Other
17	Ovo Push To Pay
18	Shopeepay QR
19	Tcash QR
20	VA Transfer

The following outlines the clustering results obtained through RapidMiner:

1. **Cluster 0 (Copper):** Represents the category with the lowest transaction frequency and value. Payment aggregators in this cluster include CIMB QR, DOKU QRIS, ECOMM, ELIGIBILITY, EMONEY, GOPAY QR, INSTALLMENT, KREDIVO QR, NOBU QR, NOBU QR DYNAMIC, OTHER, OVO PUSH TO PAY, SHOPEEPAY QR, and TCASH QR.
2. **Cluster 1 (Silver):** Represents the category with the fewest payment aggregators, primarily including CREDIT.
3. **Cluster 2 (Gold):** Represents the category that includes payment aggregators such as ATOME and DEBIT.
4. **Cluster 3 (Platinum):** Represents the category with the highest transaction frequency and value. Payment aggregators in this cluster include CASH, MINIATM TRANSFER, and VA TRANSFER.

4. CONCLUSION AND SUGGESTION

This study demonstrates that CASH, MINIATM TRANSFER, and VA TRANSFER represent the payment aggregators generating the highest revenue. It is therefore recommended to prioritize maintaining the availability of applications managing these features. Additionally, for categories with the lowest revenue, a business evaluation is needed to consider adding merchants to increase transaction frequency and value.

The optimal cluster configuration was achieved with K=2, yielding a Davies-Bouldin Index (DBI) of 0.023. Based on these findings, it is recommended to implement business management strategies such as targeted marketing, customer experience optimization, availability management, competitive analysis, and product and service innovation. These strategies aim to drive sustainable business growth and enhance customer satisfaction within the payment aggregator industry.

REFERENCES

[1] N. L. P. Handayani and P. F. Soeparan, "Peran Sistem Pembayaran Digital Dalam Revitalisasi UMKM," Transform. J. Econ. Bus. Manag., vol. 1, no. 3, pp. 20–32, 2022, doi: 10.56444/transformasi.v1i3.425

[2] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis Pada Data Mining: Studi Kasus Algoritma K-Means Clustering," J. Teknoif Tek. Inform. Inst. Teknol. Padang, vol. 10, no. 2, pp. 52–61, 2022. [online]. Available: https://teknoif.itp.ac.id/index.php/teknoif/article/view/707.

[3] D. A. Lestari, E. D. Purnamasari, and B. Setiawan, "Pengaruh payment gateway terhadap kinerja keuangan UMKM," J. Bisnis, Manajemen, dan Ekon., vol. 1, no. 1, pp. 1–10, 2020, doi: 10.47747/jbme.v1i1.20.

[4] F. Mar'i and A. A. Supianto, "Clustering Credit Card Holder Berdasarkan Pembayaran Tagihan Menggunakan Improved K-Means dengan Particle Swarm Optimization," J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 6, p. 737, 2018, doi: 10.25126/jtiik.201856858.

[5] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung), Reg. Dev. Ind. Heal. Sci. Technol. Art Life, pp. 394–403, 2019, [Online]. Available: https://ptki.ac.id/jurnal/index.php/readystar/article/view/82.

[6] Yuli Mardi, "Data Mining: Classification Using the C4.5 Algorithm," J. Edik Inform., vol. 2, no. 2, pp. 213–219, 2019.

[7] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan," InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan), vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

[8] N. F. Adani, A. F. Boy, and R. Syahputra, "Implementasi data mining untuk pengelompokan data penjualan berdasarkan pola pembelian menggunakan algoritma K-Means clustering pada Toko Syihan," J. Cyber Tech, vol. x. No.x, no. x, pp. 1–11, 2019, doi:10.53513/jct.v2i5.4648.

[9] A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering dengan RFM Model dan Algoritma K-means," e-Proceeding Eng., vol. 8, no. 2, pp. 2876–2883, 2021. [Online]. Available: openlibrarypublications.telkomuniversity.ac.id /index.php/engineering/article/view/14687.

[10] W. M. P. Dhuhita, "Clustering Using the K-Means Method to Determine Nutritional Status of Toddlers," J. Inform., vol. 15, no. 2, pp. 160–174, 2015, doi: jurnal.darmajaya.ac.id/index. php/JurnalInformatika/article/view/598

[11] R. Nainggolan and F. A. T. Tobing, "Analisis Cluster Dengan Menggunakan K-Means Untuk Pengelompokkan Online Customer Reviews (OCR) Pada Online Marketplace," Method. J. Tek. Inform. dan Sist. Inf., vol. 6, no. 1, pp. 1–5, 2020, doi: 10.46880/mtk.v6i1.246.

[12] B. S. Purnomo and P. T. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan di Kota Yogyakarta Menggunakan Metode K-Means," J. Comput. Sci. Technol., vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.

[13] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, "Penerapan algoritma K-Means dalam menentukan tingkat penyebaran pandemi COVID-19 di Indonesia," J. Teknol. Inf., vol. 4, no. 1, pp. 128–132, 2020, doi: 10.36294/jurti.v4i1.1266.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)