DAFTAR PUSTAKA

Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *24*, 32–58. https://doi.org/10.1016/j.accinf.2016.12.004

Bachtiar, F. A. (2018). Customer Segmentation Using Two-Step Mining Method Based on RFM Model. *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, 10–15. https://doi.org/10.1109/SIET.2018.8693173

Bramer, M. (2007). *Principles of Data Mining*. Springer London. https://doi.org/10.1007/978-1-84628-766-4

Cohen-addad, V., Kanade, V., Mallmann-trenn, F., & Mathieu, C. (2019). Hierarchical Clustering. *Journal of the ACM*, *66*(4), 1–42. https://doi.org/10.1145/3321386

Cristover, R., Toba, H., & Suteja, B. R. (2022). Segmentation and Formation of Customer Regression Model Based on Recency, Frequency and Monetary Analysis. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(2). https://doi.org/10.28932/jutisi.v8i2.5075

Ernawati, E., Baharin, S. S. K., & Kasmin, F. (2021). A review of data mining methods in RFM-based customer segmentation. *Journal of Physics: Conference Series*, *1869*(1). https://doi.org/10.1088/1742-6596/1869/1/012085

Eszergár-Kiss, D., & Caesar, B. (2017). Definition of user groups applying Ward's method. *Transportation Research Procedia*, *22*, 25–34. https://doi.org/10.1016/j.trpro.2017.03.004

Fang, C., & Liu, H. (2021). Research and Application of Improved Clustering Algorithm in Retail Customer Classification. *Symmetry*, *13*(10), 1789. https://doi.org/10.3390/sym13101789

Ghosal, A., Nandy, A., Das, A. K., Goswami, S., & Panday, M. (2020). *A Short Review on Different Clustering Techniques and Their Applications* (pp. 69–83). https://doi.org/10.1007/978-981-13-7403-6_9

Golalipour, K., Akbari, E., Hamidi, S. S., Lee, M., & Enayatifar, R. (2021). From clustering to clustering ensemble selection: A review. *Engineering Applications of Artificial Intelligence*, *104*, 104388. https://doi.org/10.1016/j.engappai.2021.104388

Hartama, D., Perdana Windarto, A., & Wanto, A. (2019). The Application of Data Mining in Determining Patterns of Interest of High School Graduates. *Journal of Physics: Conference Series*, *1339*(1), 012042. https://doi.org/10.1088/1742-6596/1339/1/012042

Hatami-Marbini, A., & Toloo, M. (2019). Data envelopment analysis models with ratio data: A revisit. *Computers & Industrial Engineering*, *133*, 331–338. https://doi.org/10.1016/j.cie.2019.04.041

Inayati, T., Johan Efendi, M., & Safika Dewi Ayu. (2022). Pengaruh Digital Marketing, Electronic Word of Mouth, dan Lifestyle terhadap Keputusan Pembelian pada Marketplace Shopee Indonesia. *Jurnal Teknologi Dan Manajemen Industri Terapan (JTMIT)*, *1*(3), 202–209.

KABASAKAL, İ. (2020). Güncellik Sıklık Parasallık Modeline Dayalı Müşteri Bölümlendirme: E-Perakende Sektöründe Bir Uygulama. *Bilişim Teknolojileri Dergisi*, 47–56. https://doi.org/10.17671/gazibtd.570866

Kadir, M., & Achyar, A. (2019, September 17). *Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id*. https://doi.org/10.4108/eai.1-4-2019.2287279

Li, M., Wang, Q., Shen, Y., & Zhu, T. (2021). Customer relationship management analysis of outpatients in a Chinese infectious disease hospital using drug-proportion recency-frequency-monetary model. *International Journal of Medical Informatics*, *147*, 104373. https://doi.org/10.1016/j.ijmedinf.2020.104373

Malik, M. M., Abdallah, S., & Ala'raj, M. (2018). Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Annals of Operations Research*, *270*(1–2), 287–312. https://doi.org/10.1007/s10479-016-2393-z

Muthahharah, I., & Juhari, A. (2021). A Cluster Analysis with Complete Linkage and Ward's Method for Health Service Data in Makassar City. *Jurnal Varian*, *4*(2), 109–116. https://doi.org/10.30812/varian.v4i2.883

Newman, B. N., Strickler, J. G., O'Brien, C., Lui, T., & Lynch, M. (2019). Deconstructing perfectionism in college students: Patterns of behavior, emotion, and cognition. *Personality and Individual Differences*, *145*, 106–111. https://doi.org/10.1016/j.paid.2019.03.030

Nuseir, M., & Refae, G. el. (2022). The effect of digital marketing capabilities on business performance enhancement: Mediating the role of customer relationship management (CRM). *International Journal of Data and Network Science*, *6*(2), 295–304. https://doi.org/10.5267/j.ijdns.2022.1.008

Paramadina, M., Sudarmin, S., & Aidid, M. K. (2019). Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan). *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *1*(2), 22. https://doi.org/10.35580/variansiunm9357

Parikh, Y., & Abdelfattah, E. (2020). Clustering Algorithms and RFM Analysis Performed on Retail Transactions. *2020 11th IEEE Annual Ubiquitous Computing, Electronics*

and Mobile Communication Conference, UEMCON 2020, 0506–0511. https://doi.org/10.1109/UEMCON51285.2020.9298123

Ranganathan, P., & Gogtay, N. J. (2019). An Introduction to Statistics - Data Types, Distributions and Summarizing Data. *Indian Journal of Critical Care Medicine : Peer-Reviewed, Official Publication of Indian Society of Critical Care Medicine*, *23*(Suppl 2), S169–S170. https://doi.org/10.5005/jp-journals-10071-23198

Sabbir Hossain Shihab, Shyla Afroge, & Sadia Zaman Mishu. (2019). RFM Based Market Segmentation Approach Using Advanced K-Means and Agglomerative Clustering: A Comparative Study. *IEEE*.

Sangma, J. W., Sarkar, M., Pal, V., Agrawal, A., & Yogita. (2022). Hierarchical clustering for multiple nominal data streams with evolving behaviour. *Complex & Intelligent Systems*, *8*(2), 1737–1761. https://doi.org/10.1007/s40747-021-00634-0

Saura, J. R. (2021). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92–102. https://doi.org/10.1016/j.jik.2020.08.001

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. *Industrial Marketing Management*, *98*, 161–178. https://doi.org/10.1016/j.indmarman.2021.08.006

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 747–748. https://doi.org/10.1109/DSAA49011.2020.00096

Sheard, J. (2018). Quantitative data analysis. In *Research Methods* (pp. 429–452). Elsevier. https://doi.org/10.1016/B978-0-08-102220-7.00018-2

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Stormi, K., Lindholm, A., Laine, T., & Korhonen, T. (2020). RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers. *Journal of Management and Governance*, *24*(3), 623–653. https://doi.org/10.1007/s10997-018-9447-3

Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018a). Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study. *2018 IEEE 15th*

International Conference on E-Business Engineering (ICEBE), 119–126. https://doi.org/10.1109/ICEBE.2018.00027

Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018b). Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study. *2018 IEEE 15th International Conference on E-Business Engineering (ICEBE)*, 119–126. https://doi.org/10.1109/ICEBE.2018.00027

Verhulst, B., & Neale, M. C. (2021). Best Practices for Binary and Ordinal Data Analyses. *Behavior Genetics*, *51*(3), 204–214. https://doi.org/10.1007/s10519-020-10031-x

Widodo, E., Mashita, S. N., & Prasetyowati, Y. G. (2020). Perbandingan Metode Average Linkage, Complete Linkage, dan Ward'S pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia. *Faktor Exacta*, *13*(2), 81. https://doi.org/10.30998/faktorexacta.v13i2.6581

Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. *Jurnal Ilmiah SINUS*, *18*(1), 75. https://doi.org/10.30646/sinus.v18i1.448

Yang, L., Jia, G., Wei, F., Chang, W., Li, C., & Zhou, S. (2021). The CIPCA-BPNN Failure Prediction Method Based on Interval Data Compression and Dimension Reduction. *Applied Sciences*, *11*(8), 3448. https://doi.org/10.3390/app11083448

Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*, *61*. https://doi.org/10.1016/j.jretconser.2021.102588

Skripsi Ganjil 22/23

by Jhiro Faran

Submission date: 21-Dec-2022 02:42PM (UTC+0700)
Submission ID: 1985403528
File name: JHIRO_FARAN__197064416035__INFORMATIKA.docx (661.4K)
Word count: 7309
Character count: 47313

Skripsi Ganjil 22/23

ORIGINALITY REPORT

7%	6%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.its.ac.id Internet Source	1%
2	id.scribd.com Internet Source	<1%
3	docplayer.info Internet Source	<1%
4	repositori.kemdikbud.go.id Internet Source	<1%
5	digilib.uinsby.ac.id Internet Source	<1%
6	digilib.unimed.ac.id Internet Source	<1%
7	repository.upi.edu Internet Source	<1%
8	Submitted to Sriwijaya University Student Paper	<1%
9	openjournal.unpam.ac.id Internet Source	<1%

Combination of RFM's (Recency Frequency Monetary) Method and Agglomerative Ward's Method for Donors Segmentation

Jhiro Faran
Department of Informatics
Universitas Nasional
Jakarta, Indonesia
jhiraf07@gmail.com

Agung Triayudi
Department of Informatics
Universitas Nasional
Jakarta,Indonesia
agung.triayudi@gmail.com

Rima Tamara Aldisa
Department of Informatics
Universitas Nasional
Jakarta,Indonesia
rimatamaraa@gmail.com

***Abstract*— Customer segmentation is a technique to improve the company's relationship with its customers. The RFM method is often used when segmenting customers. The RFM (Recency Frequency and Monetary) method is used to determine the character of each customer based on the transaction data made. Each variable is then given a weight to determine the final score. Final scores are used to divide customers into several segments. This research will replace the scoring procedure with RFM-based cluster analysis. Ward's Method is a Hierarchical Agglomerative Clustering technique that has high memory and time complexity. The cluster technique has the advantage of maximizing the homogeneity of group members and minimizing variance within the group. Customers will be segmented using this clustering algorithm and RFM. The data used is 10070 data and is divided into 2 datasets. Data is partitioned to reduce memory usage and outlier handling. Outlier data has 991 data and normally distributed data has 9079 data. With a maximum evaluation result of 0.936 in the distribution of 2 clusters for outlier data and a maximum evaluation result of 0.345 in the distribution of 3 clusters for normal distribution data. It follows that outlier data has significant heterogeneity between groups, while data with a normal distribution has low heterogeneity.**

Keywords*— *RFM, Ward's Method, Digital Marketing, Hierarchical Agglomerative Clustering

I. INTRODUCTION

Since the creation of both fields, numerous organizations have utilized digital marketing (DM) and data science (DS) to enhance their profits. Many organizations use artificial intelligence for business analysis to uncover opportunities, minimize risks, and aid in decision making [1]. The integration of data science with customer relationship management (CRM) provides an understanding of customer characteristics, enabling organizations to determine the most effective approach [2].

Customer segmentation is a critical step in the process of identifying customers [3]. Its segmentation aims to help the industry approach consumers in the most efficient manner, based on their characteristics and attributes [4]. Moreover, Zakat institutions categorize their donors through segmentation, dividing them into various groups. This group was formed based on the dependability of donor behavior [5].

One often used segmentation is the Recency Frequency Monetary (RFM) analysis method. The RFM model focuses on consumer habits and the relationship between Frequency and Monetary. The RFM analysis process is divided into three stages: preparing the dataset for RFM format, segmenting the data using clustering, and analyzing the clustering results [6]. Clustering has three major goals: (1) to get a usable knowledge of the data, (2) to determine the degree of similarity, and (3) to organize and analyze the data from clustering analysis. [7].

Because there are many variables and a volume of data, clustering will be harder to select the group. This condition will raise the computation cost and impact the algorithm's consistency with the data [8]. This paper uses HAC (Hierarchical Agglomerative Clustering) to perform clustering. Hierarchical clustering could handle a variety of data, but development must be efficient because the technique for computing is critical in particular areas [9]. The time complexity of hierarchical clustering is comparable to $O(N^3)$, and the space complexity is comparable to $O(N^2)$. The value of N is the number of records. Ward's method is one of the agglomerative algorithms. The algorithm has high accuracy but is very sensitive to outliers. Many simulations have been carried out on several datasets and Ward's method has significant performance over the others [10].

According to research conducted by Yash Parikh and Eman Abdelfattah, an agglomerative clustering combination has a longer execution time than K-means and DBSCAN [11]. According to research by Sabir Hossain Shihab, Shyla Afroge, and Sadia Zaman Mishu, combines RFM analysis methods between k-means, HAC, and advanced k-means. Based on this research, the speed of HAC is slower than the others but has the advantage that there is no need to declare the number of clusters that must be created [12].

II. THEORETICAL BACKGROUND

A. RFM

RFM is a three-variable calculation model based on customer transactions. "Recency" is when the customer last made a transaction. "Frequency" is how often the customer establishes a transaction. "Monetary" refers to the amount customers spend on products and services [13].

1) *Recency, the following equation is used to calculate the recency value:*

$$R = Dl - Dt_i \quad (1)$$

Where "*Dl*" is the reference transaction calculation date, and "*Dt*" is the latest transaction date for each customer.

2) *Frequency, the following equation is used to calculate the frequency value*:

$$F = \sum_{i=0}^{n} Nt_i \quad (II)$$

Where "*Nt*" is the total number of customer transactions for each consumer.

1) *Monetary, the following equation is used to calculate the monetary value*:

$$M = \sum_{i=0}^{n} At_i \quad (2)$$

Where "At" is the total amount each customer spends throughout the transaction for each consumer.

B. Hierarchical Agglomerative Clustering

HAC is called "Bottom Up" because the system scales from a small to a big cluster. The HAC work system organizes all samples into clusters, which are then examined for similarities and joined to generate larger clusters. The fundamental idea of a hierarchical cluster:

- Search for similarities with the nearest neighbor and cluster systems.
- Using distance calculations.
- A dendrogram is a hierarchical structure in the shape of a tree (such as a decision tree).

C. Ward's Method

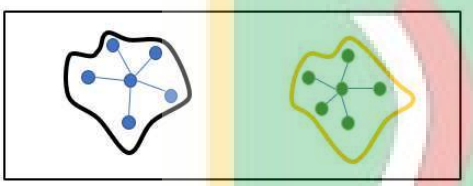

Fig. 1. Ward's Method

They are designed to maximize homogeneity and decrease variance mistakes within a group. This cluster technique does not have a definitive procedure for determining the number of clusters [14].

1) Sum of Square Error (SSE) to calculate the distance between clusters. SSE can only be computed when a cluster has several members.

$$SSE = \sum_{j=1}^{p} \left(\sum_{i=1}^{n} X_{ij}^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_{ij} \right)^2 \right) \quad (3)$$

Where "X_{ij}" is a member of each cluster, "p" is the number of clusters counted and "n" is the number of objects in a cluster.

D. Silhouette Coefficient

The silhouette coefficient is calculated using the average intra-cluster distance "a" and the nearest mean-cluster distance "b" for each data point [15]. The silhouette score description:

- If the silhouette value is close to 1, the cluster distribution is said to be good.
- If the silhouette value is close to 0, some cluster members are wrong in the cluster placement.
- If the silhouette value is close to -1, the cluster members are in the wrong cluster placement.

III. METHODOLOGY

The focus of this research was limited to align with its objectives, thereby restricting its scope. The focus of this research will be on exploring donor segmentation strategies utilizing the RFM and Agglomerative Ward's methods.

The collected data consists of transactions conducted directly between January 2019 and January 2022. There are 277,442 rows and ten columns in the datasets utilized.

Google Colaboratory is used to develop lines of code for the experiment. Google created Colab to execute the Python programming language. Python version 3.7.15 is the programming language used.

A. Design System

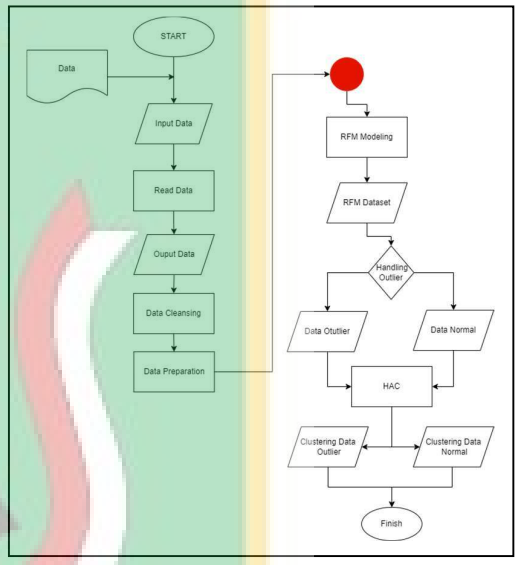

Fig. 2. Design system

The system created for segmentation uses RFM and Ward's Method.

B. Data Preparation

In the data preparation stage, checks data types, cleans data from missing values, checks data duplication, normalizes, and performs feature engineering. The purpose of this process is to improve the performance and efficiency of the model.

C. Modeling

At this stage, the donors are segmented through a process in which the cleaned and organized dataset is transformed using the RFM (Recency, Frequency, Monetary) technique prior to the clustering phase.

1) RFM Method

At this point, each donor has a unique identifier to distinguish them from other donors. Thus, the following attributes are recruited:

- Donor_id
- Date
- Donation_amount

In order to calculate the Recency value, it is necessary to merge the date of the donor's most recent transaction with the date of the most recent data update. The data shows that the donor's last transaction took place on January 17, 2022. Therefore, the end date is January 18, 2022, one day after the final day of the dataset. By using the Recency algorithm, the donor's ID is used to obtain the results for each individual donor.

To calculate the value of F (Frequency), the sum of the number of occurrences of each "Donor id" in the transaction dataset is used. The resulting value represents the Frequency for each donor.

To determine the value of M (Monetary), the total "Donation amount" for each donor is calculated using their donor id and the monetary formula. The monetary worth of each donation is derived from the computation.

2) RFM Correlation

To determine which features are correlated or interrelated, the author utilized Spearman and Pearson correlation methods.

3) RFM Data Selection

This selection is necessary in order to keep the dataset from being too large, enabling the clustering processing to be carried out. Given that HAC has high time and memory complexity levels, data reduction is required. Based on the descriptive analysis of the data, donors with high Recency values and low-Frequency Monetary values have not conducted transactions recently and can thus be eliminated

4) Handling Outlier

Ward's Method has a limitation when performing clustering if the data distribution is not normal, which can impact the model negatively and cause suboptimal results. Typically, dealing with outliers can be done by removing the outliers or through further analysis. Outliers in the analyzed dataset refer to donors with high Frequency and Monetary values, which can be considered as "loyal" donors. Therefore, a new dataset will be created to separate these outlier data from the normally distributed data.

5) Dataset

Furthermore, the dataset containing outliers will be referred to as "HighRFM" and the dataset containing normal distributions will be referred to as "MinRFM".

6) Modeling

In the following steps, the datasets containing outliers and normal distribution data will be subjected to the clustering process. This process will be carried out through iteration to determine the optimal number of clusters and yield the best evaluation results. This study utilizes the Sklearn library for clustering, with the following parameters:

- n_clusters, determining the number of clusters to be formed.
- linkage, specifying the Agglomerative algorithm to be used, in this case, Ward's method
- Affinity, using Euclidean to measure the distance between each point
- Compute_distance, for visualization of the dendrogram.

D. Evaluation

The evaluation was performed by calculating the silhouette score on the dataset. This research conducted the evaluation to determine if the machine learning model attains the research objectives and serves as a measurable benchmark for the machine learning results..

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. RFM Dataset

RFM data consist of 10070 lines and four attributes.

Table 1 RFM Dataset

Dataset	Number of Attributes	Number of Instances
RFM	4	10070

The dataset will be processed using Ward's Method to cluster the data.

B. RFM Correlation

Table 2 RFM Correlation

Correlation	Attribute	Recency	Frequency	Monetary
Pearson	Recency	1	-0.072	-0.04
	Frequency	-0.072	1	**0.6**
	Monetary	-0.04	**0.6**	1
Spearman	Recency	1	-0.26	-0,16
	Frequency	-0,26	1	**0.55**
	Monetary	-0.16	**0.55**	1

The table indicates that the correlation between Frequency and Monetary is positive, while the correlation between Recency and the other variables is negative.

C. Handling Outlier

10070 data are utilized following data reduction. Performing an outlier test will make isolating the dataset simpler. Outlier measurements can be seen in figure 3.

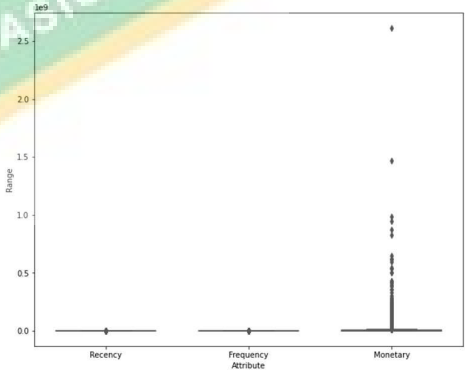

Fig. 3. Monetary Outlier

Based on the outlier test value on the dataset with Monetary dataset attributes, it is divided into 2 data with details in table 3.

Table 3 RFM Selection Data

Dataset	Number of Attributes	Number of Instance
HighRFM	4	991
MinRFM	4	9079
Total		10070

Each dataset will undergo modeling using Ward's Method and evaluations will be conducted on each dataset

D. The Descriptive Analysis

1) "HighRFM" Analysis

The descriptive analysis carried out includes the average, minimum, and maximum values, which can be seen in the table below.

Table 4 "HighRFM" descriptive

Index	Recency	Frequency	Monetary
Mean	123.856	58.009	52352320.345
Minimum	0	1	12125000.00
Maksimum	689	3990	2613667000.00

2) "MinRFM" Analysis

The descriptive analysis carried out includes the average, minimum, and maximum values, which can be seen in the table below.

Table 5 "MinRFM" descriptive

Index	Recency	Frequency	Monetary
Mean	95.022	8.021	3342231.14
Minimum	0	1	28000.00
Maksimum	696	149	12100000.00

E. Clustering Result and Evaluation Score

The advantage of Hierarchical Agglomerative Clustering is that we do not need to declare how many clusters we want to create. Thus, the author performs iterations for cluster division and will choose the best cluster evaluation results

1) "HighRFM" Dataset

Table 6 HighRFM cluster result

Dataset	Cluster	Silhouette Score
"HighRFM"	2	0.936
	3	0.588
	4	0.591
	5	0.591
	6	0.605
	7	0.609
	8	0.572

The above dataset shows the cluster division and its evaluation values. Based on the table, the 2-cluster division has the best silhouette score. Therefore, the author will use this result for further analysis.

2) "MinRFM" Dataset

Table 7 MinRFM cluster result

Dataset	Cluster	Silhouette Score
"MinRFM"	2	0.316
	3	0.345
	4	0.319
	5	0.328
	6	0.297
	7	0.250
	8	0.265

The above dataset shows the cluster division and its evaluation values. Based on the table, the 3-cluster division has the best silhouette score. Therefore, the author will use this result for further analysis.

F. Cluster Analysis

This analysis is conducted to determine the number of members in each cluster and to analyze each cluster. The analysis is related to the attributes of the cluster.

1) "HighRFM" Dataset

Based on the evaluation results, the best cluster division for the "HighRFM" dataset is 2 clusters. The following is the breakdown of the member division.

Table 7 Member of Group – "HighRFM"

Dataset	Cluster	Group	Member
HighRFM	2	1	989
		2	2

The author conducted a descriptive analysis on the "HighRFM" cluster and obtained the following results:

- The average Recency value is 123.486 for cluster 1 and 307 for cluster 2.
- The average Frequency value is 53.58 for cluster 1 and 2248 for cluster 2.
- The average Monetary value is 48,329,164.67 for cluster 1 and 2,041,802,800 for cluster 2.

2) "MinRFM" Dataset

Based on the evaluation results, the best cluster division for the "MinRFM" dataset is 3 clusters. The following is the breakdown of the member division.

Table 8 Member of Group – "MinRFM"

Dataset	Cluster	Group	Member
MinRFM	3	1	3649
		2	5024
		3	406

The author conducted a descriptive analysis on the "HighRFM" cluster and obtained the following results:

- The average Recency value for Cluster 1 is 76.404, for Cluster 2 it is 77.875, and for Cluster 3 it is 474.549.
- The average Frequency value for Cluster 1 is 12.541, for Cluster 2 it is 4.128, and for Cluster 3 it is 15.586.
- The average Monetary value for Cluster 1 is 5,323,888.42, for Cluster 2 it is 1,834,111.88, and for Cluster 3 it is 4,193,767.43.

V. CONCLUSIONS

The Ward's Method has limitations, one of which is not suitable for non-normal distributed data. Therefore, the author segmented the data so that the loyal donor data can be analyzed further. Based on the descriptive analysis of Monetary values, the dataset containing outlier data has higher results than the dataset containing normally distributed data. The drawback of the outlier dataset is the high average Recency value, indicating donors who have been making transactions for a long time.

In the clustering process with the same iteration and cluster division in both datasets, the dataset containing the outliers received an average score of 0.641 and the dataset containing the normally distributed data received a score of 0.303. This result indicates that the data containing outliers has a high variance among its groups. In the outlier dataset, group two has the highest average "Frequency" and "Monetary" result from group one and other groups from the other dataset.

The author concludes that this research successfully replaces the RFM scoring process or the conventional process. The conventional process has limitations in providing weights to each of its variables. With the clustering process, the author was able to group donors with the Ward's Method. By minimizing the variance in group members, it makes the group have a high homogeneity among its members and a high variance among groups.

REFERENCES

[1] J. R. Saura, "Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics," *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 92–102, Apr. 2021, doi: 10.1016/j.jik.2020.08.001.

[2] M. Nuseir and G. el Refae, "The effect of digital marketing capabilities on business performance enhancement: Mediating the role of customer relationship management (CRM)," *International Journal of Data and Network Science*, vol. 6, no. 2, pp. 295–304, 2022, doi: 10.5267/j.ijdns.2022.1.008.

[3] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, and R. Rahmani, "Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, Oct. 2018, pp. 119–126. doi: 10.1109/ICEBE.2018.00027.

[4] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *Journal of Retailing and Consumer Services*, vol. 61, Jul. 2021, doi: 10.1016/j.jretconser.2021.102588.

[5] C. Fang and H. Liu, "Research and Application of Improved Clustering Algorithm in Retail Customer Classification," *Symmetry (Basel)*, vol. 13, no. 10, p. 1789, Sep. 2021, doi: 10.3390/sym13101789.

[6] F. A. Bachtiar, "Customer Segmentation Using Two-Step Mining Method Based on RFM Model," in *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, Nov. 2018, pp. 10–15. doi: 10.1109/SIET.2018.8693173.

[7] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," 2020, pp. 69–83. doi: 10.1007/978-981-13-7403-6_9.

[8] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[9] V. Cohen-addad, V. Kanade, F. Mallmann-trenn, and C. Mathieu, "Hierarchical Clustering," *Journal of the ACM*, vol. 66, no. 4, pp. 1–42, Aug. 2019, doi: 10.1145/3321386.

[10] D. Eszergár-Kiss and B. Caesar, "Definition of user groups applying Ward's method," *Transportation Research Procedia*, vol. 22, pp. 25–34, 2017, doi: 10.1016/j.trpro.2017.03.004.

[11] Y. Parikh and E. Abdelfattah, "Clustering Algorithms and RFM Analysis Performed on Retail Transactions," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2020*, Oct. 2020, pp. 0506–0511. doi: 10.1109/UEMCON51285.2020.9298123.

[12] Sabbir Hossain Shihab, Shyla Afroge, and Sadia Zaman Mishu, "RFM Based Market Segmentation Approach Using Advanced K-Means and Agglomerative Clustering: A Comparative Study," *IEEE*, Feb. 2019.

[13] K. Stormi, A. Lindholm, T. Laine, and T. Korhonen, "RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers," *Journal of Management and Governance*, vol. 24, no. 3, pp. 623–653, Sep. 2020, doi: 10.1007/s10997-018-9447-3.

[14] B. N. Newman, J. G. Strickler, C. O'Brien, T. Lui, and M. Lynch, "Deconstructing perfectionism in college students: Patterns of behavior, emotion, and cognition," *Pers Individ Dif*, vol. 145, pp. 106–111, Jul. 2019, doi: 10.1016/j.paid.2019.03.030.

[15] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, Oct. 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.

ICCoSITE

*The 2023 International Conference on **Computer Science, Information Technology, and Engineering***

Ref. : ICCoSITE/LoA/I/2023/102
Date : February 3, 2023

LETTER OF ACCEPTANCE

2023 International Conference on Computer Science, Information Technology, & Engineering

Virtual Conference (Online) 16 February 2023

Paper	:	1570876371 **Combination of RFM's (Recency Frequency Monetary) Method and Agglomerative Ward's Method for Donors Segmentation**
Authors	:	Jhiro Faran, Agung Triayudi and Rima Tamara Aldisa (Universitas Nasional, Indonesia)

Dear Author(s),

We are pleased to inform you that your paper has been **accepted to be presented** in the 2023 International Conference on Computer Science, Information Technology, & Engineering (ICCoSITE) which organized by **Bina Insani University** with our partner **APTIKOM, Universitas Indraprasta PGRI**, **Universitas Nasional, Institut Teknologi Garut, STMIK ROSMA, Universitas Majalengka,** and **IEEE Indonesia Section**.

The ICCoSITE will be conducted as a virtual conference on 16 February 2023. The accepted and presented paper has been strictly undergone the peer-reviewed process and will be submitted for uploading to the IEEEXplore Digital Library and it will be normally indexed in SCOPUS database.

This the detail of catalog number IEEE Xplore Library

1) XPLORE COMPLIANT CFP23DI9-ART ISBN: 979-8-3503-2095-4
2) USB CFP23DI9-USB ISBN: 979-8-3503-2094-7

Please understand perfectly that the publication of papers in the IEEEXplore Digital Library will take AT LEAST 1 – 4 months from the conference date. By registering to this conference, authors and co-authors will be deemed to agree, understood, and accept all IEEE Official Conference Terms and Conditions at https://www.ieee.org/web/conferences/organizers/pubs/conference_publications.html

Congratulations on the acceptance of your paper and thank you for your interest. We look forward to seeing you at the conference soon.

General Chair,

BINA INSANI
INTERNATIONAL CONFERENCE

Ahmad Chusyairi, S.Kom., M.Kom.