Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13024
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Internet Service Provider User Customer Lifetime Segmentation Analysis using RFM and K-Means Algorithm

Muhammad Febrian Rachmadhan Amri [1)*], Mohamad Hafidhul Umam [2)], Arief Wibowo [3)], I Made Satrya Ramayu [4)]
[1,4)] Digital Business Study Program, Institut Desain dan Bisnis Bali, Indonesia
[2,3)] Faculty of Information Technology, Universitas Budi Luhur, Indonesia
[1)] febrian.rachmadhan@gmail.com, [2)] mohammadhafidhul@gmail.com,
[3)] arief.wibowo@budiluhur.ac.id, [4)] satrya.ramayu@idbbali.ac.id

Submitted : Sep 19, 2023 | **Accepted** : Oct 15, 2023 | **Published** : Jan 1, 2024

Abstract: The characteristics of each customer can be segmented using RFM (Recency, Frequency, Monetary) which means customer's last transaction time, number of customer transactions, and amount of money spent. The Lifetime and K-Means methods are used to perform the process of clustering or grouping customers based on segmentation through RFM. The results will be divided into 4 clusters namely Gold, Silver, Platinum and Diamond. The results of clustering are visualized with graphs and cluster tables containing the results of segmentation and clusters or groups of From the results obtained from the previous stage, of the 104 customers in the Retail & Distribution Services (RDS) sector, 4 segments resulted in 43 customers with Platinum class, 39 customers with gold class, 14 customers with silver class, and 8 customers with platinum level. The most popular services services or product is high speed dedicated internet services, VPN IP package, and service network package as top 3 results. The largest amount of revenue services or product is transponder full time use services, support network and contact center application as top 3 results.

Keywords: Customer Segmentation, Lifetime, Clustering, RFM, K-Means

INTRODUCTION

Competition in the world of the telecommunications industry is getting tougher and the services and offers offered by competitors are increasingly numerous and diverse. Forcing telecommunications industry players to continue to innovate and develop their service products so that service users continue to use the service products provided. The customer is the owner of the position whose most important role is in the development of a business strategy, the customers one of the sources of profit in the company (Ramayu, Susanto, & Mahendra, 2022). Therefore, a good understanding of customers is needed (Wardani, Nugraha, et al., 2022). Where a good understanding of these customers can be used by companies to plant a strategy and target for investment in potential customer assets (Ramayu, 2023). The problem that is often faced in planning a business strategy, of course, is the difficulty in analyzing customer value (Xing & Xin-feng, 2010). Many marketers and businesses have difficulty in identifying the right customers. this has the potential to cause the company to lose potential customers and of course it will be very detrimental to the company (Nugraha & Mahendra, 2020). By segmenting customers based on their behavior, we can better target their actions. An example is a product launching event tailored to marketing targets aimed at meeting customer expectations (Balaji & Srivatsa, 2012). In addition, the data owned by the company is also a very important asset (Adnyana & Sulastra, 2020).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13024 p-ISSN : 2541-044X

The type of characteristics of each customer in this case can be seen by utilizing the RFM method, namely the Recency, Frequency, Monetary method which means Recency (the time of the customer's last transaction), Frequency (the number of customer transactions), and Monetary (the amount of money spent) (Husein, Setiawan, Sumangunsong, Simatupang, & Yasmin, 2022). As well as utilizing the K-Means Algorithm which can be used to carry out the segmentation process or customer groups based on RFM (Budilaksono et al., 2021). RFM itself is an analytical method that is quite well known and often used in marketing strategy through CRM (Customer Relationship Management) approach (Wardani, Arnidya, et al., 2022). CRF can show us the priorities and focus of the campaign. On the other hand, we also have to measure customer value, it will be less than optimal if low value customers are given high priority. We can use Monetary and Frequency values to determine customer value. In this case, we do not know the value limits for Frequency and Monetary, so we decided to use the K-Means Clustering technique. In determining the data group on K-Means, this is done by making a cluster determination that aims to show similarities between data (Putra, Pinem, Silalahi, Gulo, & Liukhoto, 2022). Where each data that has similarities will have a close hierarchical relationship and form a data cluster.

The purpose of this study is to find out how to properly determine the type of promotion that can be offered to potential customers in order to continue using the services provided. However, due to the large number of customers and having a variety of characteristics and different levels of loyalty, of course, the provision of promotions cannot be given equally and evenly. In this case, customer segmentation is very necessary so that the promotions given can be right on target according to the segmentation.

The output built in this study is customer segmentation using an RFM analysis model. The segmentation results were then grouped with the K-Means clustering algorithm. The first output produced is customer segmentation based on frequency during transaction activities, the most frequent transactions that occur in the current time span, and the highest amount of income generated from the entire transaction (Dewi, Gunadi, & Indrawan, 2020). Then the customer data that has been segmented is grouped into 4 clusters. This is the second output that is the purpose of this study. It is hoped that the segmentation and clustering carried out can help ISP companies in designing future business strategies.

LITERATURE REVIEW

Conducting segmentation analysis based on customer characteristics can help companies in determining potential customer segments to be promoted more and allow income to increase due to the segmentation (Taqwim, Setiawan, & Bachtiar, 2019). The method used is RFM in segmentation and combined with the K-Means Algorithm. Where in the study the company wanted to apply different CRM (Customer Relationship Management) in different customer segments so that segmentation was carried out with the RFM Model using the Fuzzy C-Means Clustering algorithm. The results of the clustering are then visualized in a dashboard and usability testing is carried out to calculate the success rate of the research.

The main objective of applying the K-Means algorithm in order to divide customers into several segments, so that the level of potential of each customer was obtained and known (Widiyanto & Witanti, 2021). The analysis here is carried out in two stages, namely RFM weighting to produce RFM weights where recency is an analysis based on recent transactions, frequency is the number of transaction events that have occurred, and monetary which is the total transaction costs that have occurred or been carried out. After that, it is collected by the K-Means method. The results of the evaluation of the system built found that customer clusters with BZ category values/fields had a percentage value of 54.3%, the MVC category of 21.8% and mgc of 23.9%.

There are many methods that can be used in analyzing Customer Lifetime Value such as using embedding in scientific papers (Chamberlain, Cardoso, Liu, Pagliari, & Deisenroth, 2017) then analysis by utilizing deep neural networks in scientific (Wang, Liu, & Miao, 2019) and scientific works (Pollak, 2021) but these methods are not appropriately used for certain data. On the other hand, the RFM method is often used in cases of Customer Lifetime Value analysis as in previous scientific papers as well as scientific papers (Rumiarti & Budi, 2017), (Alizadeh Zoeram & Karimi Mazidi, 2018), (Juniati, Monalisa, Zafa, & Muslim, 2020), (Christy, Umamakeswari, Priyatharsini, & Neyaa, 2021) and (Department of Informatics Engineering, Widyagama University of Malang, Jalan Borobudur no 35,

*name of corresponding author

Malang, 65125, INDONESIA et al., 2019) and gets optimal results in conducting Customer Lifetime Value analysis by utilizing the RFM method and combined with other algorithmic methods (Brahmana, Mohammed, & Chairuang, 2020).

Each customer has a different strategy in making purchases, promotions, and giving privilege. The first example is for example the existence of customers or customers who deliberately or often make large purchases because they have more funds or want to keep stock of goods. There are also types of customers or customers who make purchases only once in a period of three months directly in large quantities, for example for monthly stock, and some usually make timed purchases, for example in every certain period of time. This is often a scourge of problems for every business actor or company in carrying out customer maintenance and marketing their products. Therefore, knowing the potential and characteristics of each of their customers in purchasing products is important for the company. Customer segmentation itself requires companies to better understand the unique nature and character of their different customers and seek continuity so that they can divide them into specific clusters, so that companies can plan the right approach to each different customer segment (Adnyana & Sulastra, 2020).

The RFM analysis model consisting of Recency, Frequency, Monetary can be interpreted as follows (Christy et al., 2021):

a. Recency: Recency is a variable that is useful in measuring the value of customers based on the range of times of date, month and year of the current transaction (recently) in each customer to date. Keep in mind that the smaller the time range value, the greater the recency value.
b. Frequency: Frequency is the second variable, which is a variable to measure the value of each customer based on the number of transactions that occur when the customer transactions in one period. Thus the greater the number of transactions, the greater the value of f.
c. Monetary: The last variable in the RFM is Monetary, the Monetary variable is a variable that is useful for measuring customer value based on or referring to the amount of money spent by the customer in a certain period. The greater the amount of money spent by the customer, the greater the value of M.

The clustering technique intends to group data into different segments with internal cohesion. Clustering is a process of grouping data into a number of clusters or groups that have similarities or similar characteristics for each data in existing groups (Department of Informatics Engineering, Widyagama University of Malang, Jalan Borobudur no 35, Malang, 65125, INDONESIA et al., 2019). There are many algorithms that can be used to carry out the clustering process, including several: K-Means, LVQ (Learning Vector Quantization), FCM (Fuzzy C-Means), and so on (Karaa & Dey, 2017).

K-Means is an unsupervised machine learning algorithm where this algorithm aims to divide groups of data and find the continuity or continuity or similarity of data and is collected into 1 center, with the number of groups to be represented by variable K which means matching characteristics. The variable K is the number of groups that are cooled. To process K-Means algorithm data, it begins by selecting a random midpoint or centroid as the starting point in each group. Then carry out an iterative calculation process to optimize the position of the centroid until the centroid is in a stable state or the specified number of iterations has been met or achieved. Here are the stages of the K-Means Algorithm:

1) The first stage is to determine the number of clusters by dividing and collecting data into clusters or groups or the like, which will make it easier for researchers to define patterns regarding the segmentation of transactions made by customers.
2) The second stage is to determine the initial centroid point by means of a random retrieval of the data object used.
3) The third stage is to measure the distance for each point of the data object used with the centroid point.
    a) The fourth stage is to obtain the proximity between two objects based on the distance between the two objects. This is also the case with the proximity of a data to a cluster or group which is determined by the distance between the data and the cluster center. At this stage it is also necessary to calculate the distance for each data to each cluster center. And keep in mind that the closest distance between a data and a certain cluster will determine which data is categorized into which

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13024 p-ISSN : 2541-044X

cluster. To calculate the distance from an object to a cluster or group using the Euclidean distance space formula as seen in Equation 1 (Ruli & Wuryanto, 2017).

$$D_{L_2}(x_1, x_2) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j-1}^{p}(x_{2j} - x_{1j})^2} \quad (1)$$

Notes:

- X1 = Data Object
- X2 = Centroid
- P = Data Dimension
- D = Distance

b) The fifth stage is to re-determine the new centroid. To reassess the centroid can be done by calculating the median value of the current cluster membership.
c) The sixth stage is to relocate all objects to the centroid as well as the closest group to the new centroid. This allocation is performed on all objects to the cluster or group closest to the new centroid. This process will continue to repeat until no more objects are migrating clusters.
d) At this last stage, if there are no more objects that move clusters or groups then the clustering process can be declared complete.

METHOD

This research was conducted by applying a process that was divided into five stages including: Data Collection, Data Transformation and Data Set Preparation, Segmentation Implementation with RFM analysis, Implementation of the K-Means algorithm to create clusters or groups or the like and evaluation of results. Figure 1 shows the methodology used.

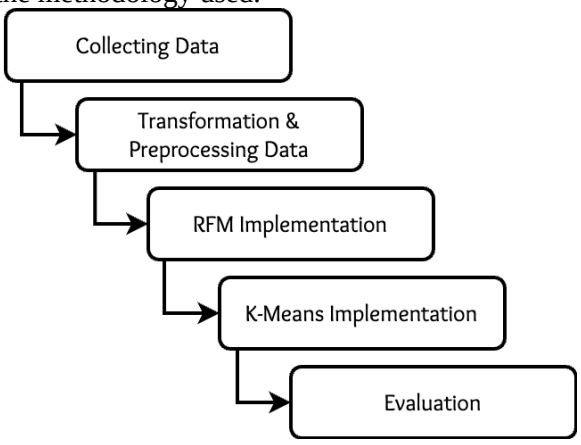

Fig. 1 Research Steps

In this research, the process begins with data collection, followed by integration, cleaning, and feature selection to prepare the dataset. Subsequently, the implementation of techniques like RFM (Recency, Frequency, Monetary) involves calculating scores, defining segments, and interpreting customer behavior. K-Means clustering requires selecting the number of clusters, initializing centroids, assigning data points, and iteratively updating centroids. Lastly, evaluation involves choosing metrics, performing cross-validation, comparing results, and visualizing patterns to determine the most effective approach for achieving research objectives. These steps collectively constitute a comprehensive data mining workflow, allowing researchers to uncover valuable insights and knowledge from complex datasets. Prototyping for the segmentation and clustering process as well as visualization of results is carried out using the Python programming language.

RESULT

Collecting Data Stage

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13024
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The data used in this paper is revenue data for ISP companies in Indonesia in 2021. In the revenue data there are many columns but not all columns are used. The columns used are more like a list of internet subscribers, region, type of service, target market, billing period and billed amount of internet usage. ISP company's revenue dataset, there are about 266655 data lines and 5 market segments.

Transformation & Preprocessing Data Stage

At this stage the revenue data is normalized to return the data to the proper data type, for example the billing data in the dataset used here has a string or object data type which we should adjust to the proper data type, which is float and billing periods as datetime data type. then check the percentage of missing values for each data that will be used and perform cleaning. Figure 2. shows a graph of missing values or empty data for each entity in the dataset. this aims to see whether the entity that will be used for processing with the RFM method has a high missing value or not. In this study we will use several entities including curr (currency), cust_id (customer id), bill_mny (bill amount), bill_prd (billing period), Product_grp (product group), and market_segment (market segment). in Figure 2, we can see that the entity to be used does not have a high percentage of missing value. Furthermore, for entities that are not used and have a high missing value, we can drop them. and have a high missing value, we can drop them.

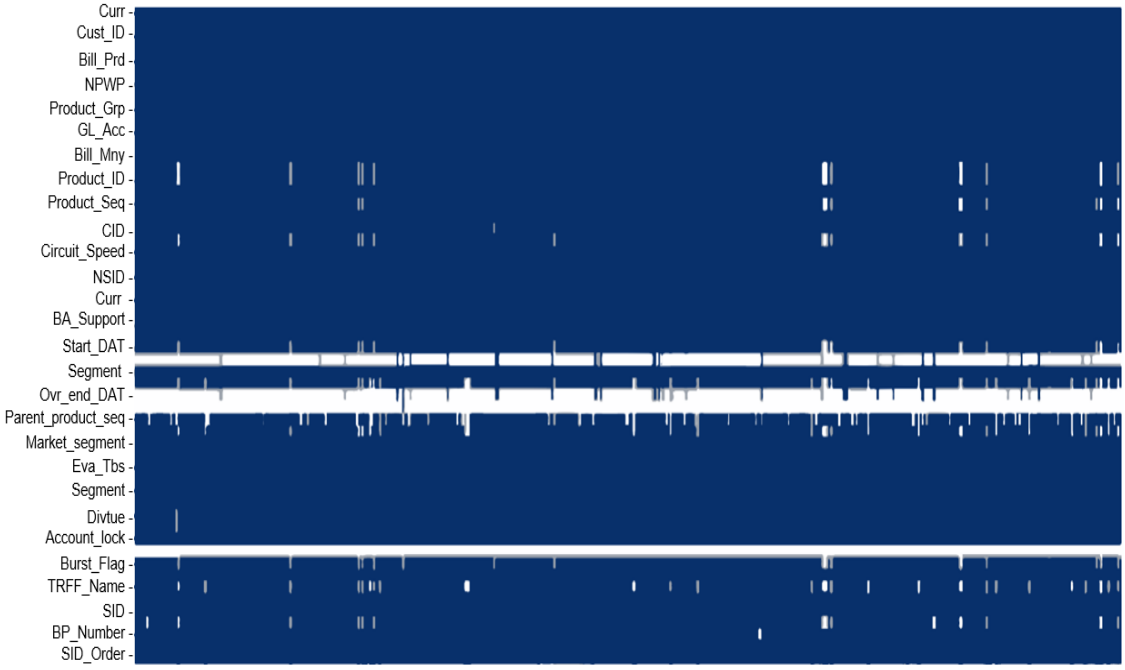

Fig. 2 Missing Value Percentage

Fig. 3 Data after Cleansing.

After cleansing the dataset that has been prepared, the next step is to segment customers, this needs to be done in order to get more specific results, here is customer segmentation based on the business field of the customer. Segmentation is carried out by summing the most transactions carried out in each field. This process is shown in Figure 4.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13024
e-ISSN : 2541-2019
p-ISSN : 2541-044X

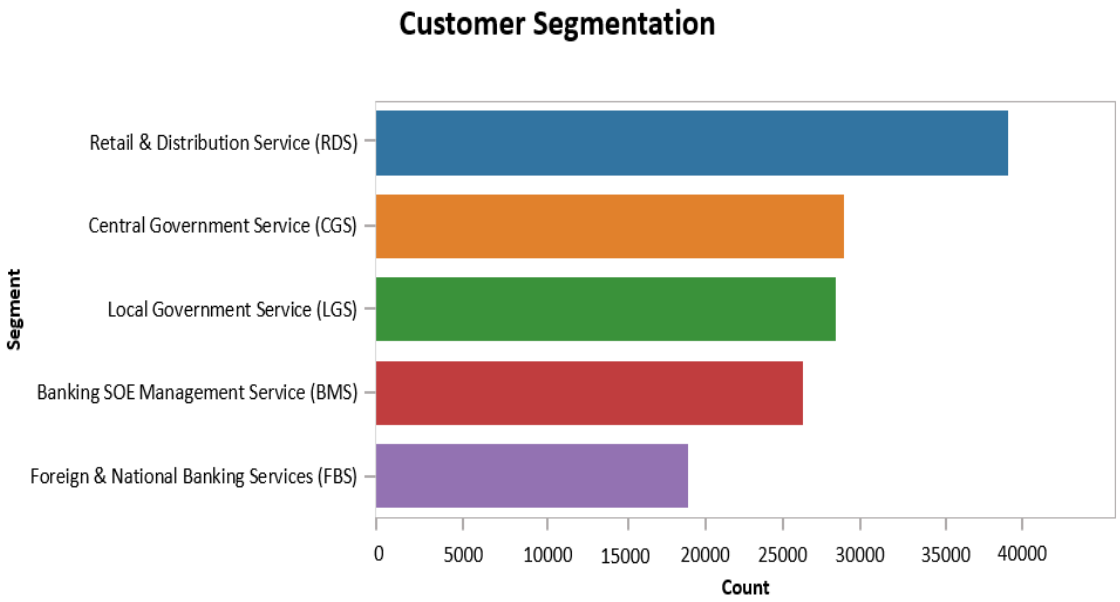

Fig. 4 Customer Segmentation Graph

RFM Implementation Stage

Then in each field of the customer will be processed with the RFM analysis method. In this case, researchers used RFM for the largest transaction data, namely the Retail & Distribution Services (RDS) field with 104 customers. Technically, RFM analysis is done using the Lifetimes library. At this stage Lifetimes will create new data frames automatically based on the data-frames that have been prepared in advance. the data-frames generated with the Lifetimes library will produce output in the frequency, recency, T, and monetary columns. Where T is the age of the customer from the time of the first purchase to the end of the specified time period. Figure 5 is data-frame lifetimes.

	CUSTOMER_REF	frequency	recency	T	monetary_value
0	2000104	10.0	329.0	329.0	48000000.0
1	2000105	0.0	0.0	30.0	0.0
2	2000168	10.0	329.0	329.0	18000000.0
3	2000174	10.0	329.0	329.0	33600000.0
4	2000197	2.0	61.0	61.0	203061866.5
...	...	...	...	...	...

Fig. 5 Data-Frame Lifetimes

After segmentation is carried out, it is continued by displaying the visualization of segmentation in the form of a graph in accordance with the provisions of the RFM analysis. The first graph is with the Recency approach visualized in Figure 6 (a). In Figure 6 (a), there are at least 45 customers who had the last transaction more than 300 days ago. The segmentation chart in terms of frequency approach visualized in Figure 6 (b). In figure 6 (b), there are at least 27 customers who have about 10 routine transactions every month (repeat orders). The segmentation chart is in a monetary approach visualized in Figure 6 (c). In figure 6 (c), there are more than 80 customers who have transactions under Rp. 10 million, and less than 10 customers who have transactions over Rp. 60 million.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13024
e-ISSN : 2541-2019
p-ISSN : 2541-044X

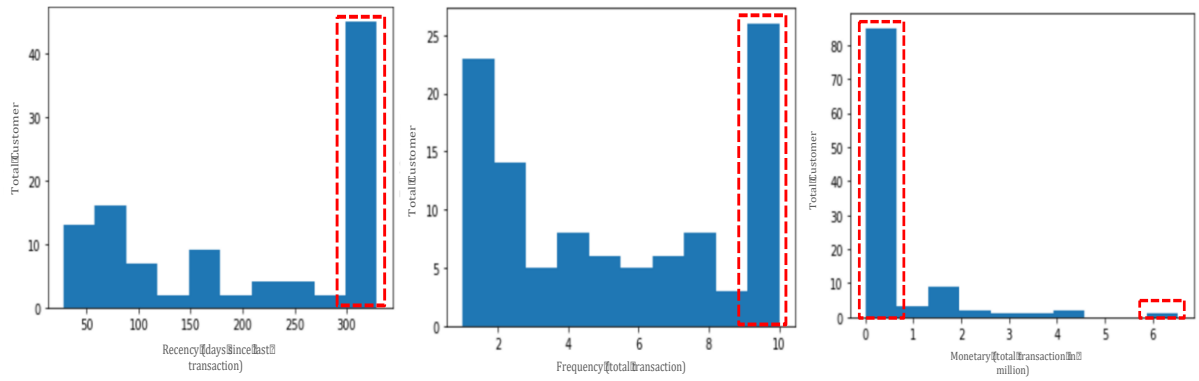

Fig. 6 Recency, Frequency and Monetary Segmentation Graph

K-Means Implementation Stage

After the segmentation data is obtained and can be visualized, the next stage is to group the data into clusters using the K-Means clustering algorithm. Algorithm cluster graph demotion in Python can be seen in Figure 7.

First, find the number of clusters by determining how many centroids are used as a reference for grouping. Cluster selection is also done with the K-Means Elbow algorithm which is generated automatically by the model machine. The results of the elbow graph in Figure 8, then the number of clusters to be used is 4 clusters. This is based on the sharpest graphics drop.

```
for k in range(1, 15):
    kmeans = KMeans(n_clusters=k)
    member = kmeans.fit_predict(np.array(rfm['frequency'])
             .reshape(-1, 1))
    score.append(kmeans.inertia_)
```

Fig. 7 Algorithm Cluster Graph Demotion using Python

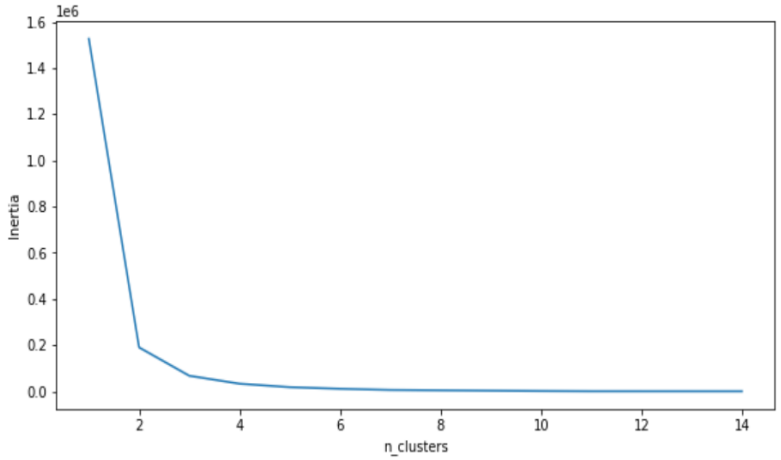

Fig. 8 Cluster Graph Demotion

Then proceed with assigning clusters with the K-Means algorithm. Clusters are assigned to each segment approach on the RFM as seen in Figure 9. Each individual cluster score from each data to get an RFM score. The results of this process can be seen in Figure 10. The last step is to label each RFM score that has been obtained previously, then the consumer list table is obtained. The label given is adjusted to the customer level that has been used previously by the company, namely Silver, Gold, Platinum, and Diamond. Label of RFM Score can be seen in Figure 11. Then the cluster results are displayed into the form of a graph, thus producing a visual graph as shown in Figure 13.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13024
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Customer_ID	Frequency	Recency	T	Monetary_value	Recency_Cluster	Frequency_Cluster	Monetary_Cluster
2000104	10.0	329.0	329.0	48000000.0	2	3	1
2000168	10.0	329.0	329.0	18000000.0	2	3	1
2000174	10.0	329.0	329.0	33600000.0	2	3	1
2000197	2.0	61.0	61.0	203061866.5	1	0	0
2000284	10.0	329.0	329.0	4300000.0	2	3	1

Fig. 9 Cluster Chosen

Frequency	Recency	T	Monetary_value	Recency_Cluster	Frequency_Cluster	Monetary_Cluster	RFM_Score
10.0	329.0	329.0	48000000.0	2	3	1	6
10.0	329.0	329.0	18000000.0	2	3	1	6
10.0	329.0	329.0	33600000.0	2	3	1	6
2.0	61.0	61.0	203061866.5	1	0	0	1
10.0	329.0	329.0	4300000.0	2	3	1	6

Fig. 10 RFM Score

Customer_ID	Frequency	Recency	T	Monetary_value	Recency_Cluster	Frequency_Cluster	Monetary_Cluster	RFM_Score	Label
2000104	10.0	329.0	329.0	48000000.0	2	3	1	6	Platinum
2000168	10.0	329.0	329.0	18000000.0	2	3	1	6	Platinum
2000174	10.0	329.0	329.0	33600000.0	2	3	1	6	Platinum
2000197	2.0	61.0	61.0	203061866.5	1	0	0	1	Bronze
2000284	10.0	329.0	329.0	4300000.0	2	3	1	6	Platinum

Fig. 11 Label of RFM Score

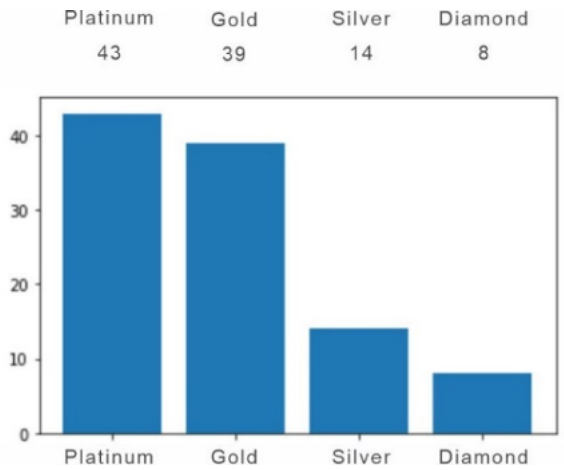

Fig. 13 Cluster Visualization

Evaluation Stage

In table 1 are the services or products that are most in demand, and the largest amount of revenue.

Table. 1 The Most Popular Services and The Largest Amount of Revenue

No of Order	Services or Product of The Most Popular Services
1	High speed dedicated internet service
2	VPN IP Package
3	Service network package
4	Standard Internet Services
5	Device Services
6	Manage Service Device
7	Manage Service MRC
8	VPN IP Node
9	Wifi Internet Service Standard
10	Mobile Connectivity Service

No of Order	Services or Product of The Largest Amount of Revenue
1	Transponder Full Time Use Services
2	Support Network
3	Contact Center Application
4	Mobile Device Bundling & MDM
5	Ticketing Services
6	VPN IP
7	PayTV Advertising - Monthly Recurring Charges
8	Device Services
9	Satellite based internet service
10	Data Center / Cloud

*name of corresponding author

DISCUSSIONS

From the results obtained from the previous stage, of the 104 customers in the Retail & Distribution Services (RDS) sector, 4 segments resulted in 43 customers with Platinum class, 39 customers with gold class, 14 customers with silver class, and 8 customers with platinum level. with the following distribution characteristics. For the Diamond cluster, customers tend to have higher RFM values for the frequency attribute, in the gold cluster they are more likely to be on monetary attributes, in the silver cluster they are more likely to be on the recency attribute and in the bronze cluster they tend to be on the recency and frequency attributes with balanced results but with a lower monetary value. From the results of RFM analysis on retail data, it can be seen that the product 'High speed broadband internet' is most in demand by most customers, but not the largest amount of revenue. the largest amount of revenue comes from the fewest and rarest transactions, namely Transponder Full Time Use Services or satellite rental services.

CONCLUSION

From the results of the discussion, it can be concluded that the Retail & Distribution Services (RDS) segment produces 2 outputs. The first output that is generated is customer segmentation based on frequency, it is known that the product service 'High speed broadband internet' is the service that is most in demand and has the highest transaction frequency of all existing services. while in terms of the highest amount of revenue generated from all transactions comes from Transponder Full Time Use Services or satellite rental for several years. Then the second output that is generated is 4 clusters obtained from the results of customer segmentation as the figure 14. In this output it is known that cluster gold is the cluster with the highest number of subscribers. Members of the gold cluster tend to have monetary attributes, meaning customers in the gold cluster have bills with large values. In our opinion, if it is linked back to the case of ISP users, especially in the Retail & Distribution Services sector, customers in this cluster are willing to pay more to get better service.

REFERENCES

Adnyana, I. G., & Sulastra, I. M. D. J. (2020). Implementation of Data Backup and Synchronization Based on Identity Column Real Time Data Warehouse. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *11*(1), 9. https://doi.org/10.24843/LKJITI.2020.v11.i01.p02

Alizadeh Zoeram, A., & Karimi Mazidi, A. (2018). A new approach for customer clustering by integrating the LRFM model and fuzzy inference system. *Iranian Journal of Management Studies*, (Online First). https://doi.org/10.22059/ijms.2018.242528.672839

Balaji, S., & Srivatsa, D. S. K. (2012). Customer Segmentation for Decision Support using Clustering and Association Rule based approaches. *Engineering Technology*, *3*(11), 525–529.

Brahmana, R. W. S., Mohammed, F. A., & Chairuang, K. (2020). Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *11*(1), 32. https://doi.org/10.24843/LKJITI.2020.v11.i01.p04

Budilaksono, S., Jupriyanto, Suwarno, M. A. S., Suwartane, I. G. A., Azhari, L., Fauzi, A., … Effendi, M. S. (2021). Customer Profilling for Precision Marketing using RFM Method, K-MEANS algorithm and Decision Tree. *SinkrOn*, *6*(1), 191–200. https://doi.org/10.33395/sinkron.v6i1.11225

Chamberlain, B. P., Cardoso, Â., Liu, C. H. B., Pagliari, R., & Deisenroth, M. P. (2017). Customer Lifetime Value Prediction Using Embeddings. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1753–1762. Halifax NS Canada: ACM. https://doi.org/10.1145/3097983.3098123

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, *33*(10), 1251–1257. https://doi.org/10.1016/j.jksuci.2018.09.004

*name of corresponding author

Department of Informatics Engineering, Widyagama University of Malang, Jalan Borobudur no 35, Malang, 65125, INDONESIA, Marisa, F., Syed Ahmad, S. S., Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Jalan Hang Tuah jaya Durian Tunggal, Melaka, 76100, MALAYSIA, Mohd Yusof, Z. I., Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Jalan Hang Tuah jaya Durian Tunggal, Melaka, 76100, MALAYSIA, … Department of Informatics Engineering, Pradnya Paramita School of Informatics Management and Computer, Jalan LA Sucipto no 249-A, Malang, INDONESIA. (2019). Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. *International Journal of Integrated Engineering*, *11*(3). https://doi.org/10.30880/ijie.2019.11.03.018

Dewi, N. W. E. R., Gunadi, I. G. A., & Indrawan, G. (2020). Detection of Class Regularity with Support Vector Machine methods. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *11*(1), 20. https://doi.org/10.24843/LKJITI.2020.v11.i01.p03

Husein, A. M., Setiawan, D., Sumangunsong, A. R. K., Simatupang, A., & Yasmin, S. A. (2022). Combination Grouping Techniques and Association Rules For Marketing Analysis based Customer Segmentation. *SinkrOn*, *7*(3), 1998–2007. https://doi.org/10.33395/sinkron.v7i3.11571

Juniati, F., Monalisa, S., Zafa, R., & Muslim, M. (2020). Klasterisasi Customer Lifetime Value Dengan Model LRFM Menggunakan Algoritma Fuzzy C-Means. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, *6*(1), 38. https://doi.org/10.24014/rmsi.v6i1.8635

Karaa, W. B. A., & Dey, N. (2017). *Mining Multimedia Documents*. CRC Press. Retrieved from https://books.google.co.id/books?id=-EQPEAAAQBAJ

Nugraha, P. G. S. C., & Mahendra, G. S. (2020). Explorasi Algoritma C4.5 dan Forward Feature Selection untuk Menentukan Debitur Baik dan Debitur Bermasalah pada Produk Kredit Tanpa Agunan (KTA). *JST (Jurnal Sains dan Teknologi)*, *9*(1), 39–46. https://doi.org/10.23887/jst-undiksha.v9i1.24627

Pollak, Z. (2021). *Predicting Customer Lifetime Values – E-Commerce Use Case*.

Putra, A. Z., Pinem, R. W., Silalahi, S., Gulo, F., & Liukhoto, J. A. A. (2022). Classification of Covid-19 Patient Spread Rate By Age and Region With K-Means Algorithm. *SinkrOn*, *7*(3), 1085–1989. https://doi.org/10.33395/sinkron.v7i3.11603

Ramayu, I. M. S. (2023). Rancangan Sistem Informasi Persediaan Stok Obat Dengan Metode System Development Life Cycle di Apotek Nusa Farma Nusa Penida. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, *4*(1), 110–120. https://doi.org/10.35870/jimik.v4i1.130

Ramayu, I. M. S., Susanto, F., & Mahendra, G. S. (2022). Penerapan Data Mining dengan Algoritma C4.5 dalam Pemesanan Obat Guna Meningkatkan Keuntungan Apotek. *Prosiding Seminar Nasional Manajemen, Desain & Aplikasi Bisnis Teknologi (SENADA)*, *5*, 237–245. Denpasar: Institut Desain Bali.

Ruli, R. D. F., & Wuryanto, E. (2017). *Penerapan Clustering K-Means pada Customer Segmentation Berbasis Recency Frequency Monetary (RFM) (Studi Kasus: PT. Sinar Kencana Intermoda Surabaya)*. 418–427. Surabaya: Universitas Airlangga.

Rumiarti, C. D., & Budi, I. (2017). Segmentasi Pelanggan pada Customer Relationship Management di Perusahaan Ritel: Studi Kasus PT Gramedia Asri Media. *urnal Sistem Informasi (Journal of Information System)*, *13*(1), 1–10.

Taqwim, W. A., Setiawan, N. Y., & Bachtiar, F. A. (2019). Analisis Segmentasi Pelanggan Dengan RFM Model pada PT. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *3*(2), 1986–1993.

*name of corresponding author

Wang, X., Liu, T., & Miao, J. (2019, December 16). *A Deep Probabilistic Model for Customer Lifetime Value Prediction*. arXiv. Retrieved from http://arxiv.org/abs/1912.07753

Wardani, N. W., Arnidya, D. J., Putra, I. N. A. S., Desmayani, N. M. M. R., Nugraha, P. G. S. C., Hartono, E., & Mahendra, G. S. (2022). Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM. *Jurnal Manajemen dan Teknologi Informasi*, *12*(2), 113–124. https://doi.org/10.5281/zenodo.7178249

Wardani, N. W., Nugraha, P. G. S. C., Hartono, E., Suryawan, I. W. D., Dirgayusari, A. M., Darmadi, I. W., & Mahendra, G. S. (2022). Penerapan Data Mining Untuk Klasifikasi Penjualan Barang Terlaris Menggunakan Metode Decision Tree C4.5. *Jurnal Teknologi Informasi dan Komputer*, *8*(3), 268–279.

Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *1*(1), 204–215. https://doi.org/10.24002/konstelasi.v1i1.4293

Xing, B., & Xin-feng, W. (2010). The Evaluation of Customer Potential Value Based on Prediction and Cluster Analysis. *2010 International Conference on Management Science & Engineering 17th Annual Conference Proceedings*, 613–618. Melbourne, Australia: IEEE. https://doi.org/10.1109/ICMSE.2010.5719866

*name of corresponding author