60Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 9 No. 1 2024, pp. 147– 160

Customer Segmentation Using K-Means Clustering with RFM Method (Case Study : PT. Dewangga Travindo Semarang)

Hida Sekar Winaryanti[1], **Heru Pramono Hadi***[2]
Study Program in Information System, Faculty of Computer Science, University of Dian Nuswantoro, Semarang
E-mail : hidasekarwinaryanti@gmail.com[1], heru.pramono.hadi@dsn.dinus.ac.id[2]*
**Corresponding author*

Eko Hari Rachmawanto[3]
Study Program in Informatics Engineering, Faculty of Computer Science, University of Dian Nuswantoro, Semarang
E-mail : eko.hari@dsn.dinus.ac.id[3]

Received 8 December 2015; Revised 10 February 2016; Accepted 2 March 2016

Abstract - PT. Dewangga Travindo is a company that operates in the field of travel services which includes tours, travel, and Hajj and Umrah pilgrimages which is based in the city of Semarang and has received permission from the Ministry of Religion No. D/606 of 2013. Every year there is always an increase in sales of services. Hajj and Umrah. The higher transaction activity every day results in a large buildup of data in the database which will only become data waste. The ability to process data is increasingly sophisticated using data mining, which is an activity of looking for relationships between items to obtain patterns as information to assist in decision making. However, considering the large number of competitors offering the same services, it is necessary to increase competitiveness to overcome market segmentation at PT Dewangga Travindo. For this reason, this research was carried out which aims to overcome customer segmentation using the Clustering method with the K-Means algorithm which produces a visual cluster model with RStudio tools using RFM attributes applied to carry out segmentation. The data used in this research is data on Hajj and Umrah pilgrims in the 2018-2020 period.

Keywords - Data Mining, custpmer segmentation, Clustering, K-Means, RStudio, RFM

1. INTRODUCTION

The high interest of Muslims in Indonesia who are competing to carry out the Hajj and Umrah pilgrimages has resulted in many travel services emerging and providing Hajj and Umrah departure services, such as PT. Dewangga Travindo which provides services for Muslims who want to carry out the Umrah and Hajj pilgrimages. PT. Dewangga Travindo is a company that operates in the field of tourism services with the scope of Tour and Travel, Hajj and Umrah, based in the city of Semarang with its address at Setia Budi No. 91A Srondol Banyumanik Semarang 50263, Central Java. The increasingly rapid development of the travel business sector in Semarang was the beginning of the establishment of PT Dewangga Travindo, which was founded in 1999 and has handled travel for official, private, domestic and foreign government affairs. Since September 13 2013, PT Dewangga Travindo began to transform, focusing on the goal of serving God's guests through Dewangga Lil Hajj Wal Umrah with the permission of the Ministry of Religion number D/606 of 2013 and the permission of the Ministry of Religion No.

756/2016 whose credibility has been tested [1]. PT Dewangga Travindo currently has branches in several cities, namely Pati, Kendal, Solo, Wonosobo and Jogjakarta.

PT Dewangga Travindo often experiences improvements in Hajj and Umrah services every year. In 2018, Dewangga was trusted by 1,500 Umrah pilgrims and in 2019 there were 2,800 pilgrims who used travel services to the holy land. However, considering the high competition between businesses offering similar services, it is necessary to increase competitiveness to overcome customer segmentation at PT Dewangga Travindo. To increase competitiveness, PT. Dewangga Travindo needs to carry out market development and align it with current business developments through implementing business strategies by analyzing congregation data every year. The situation being pursued by PT. Dewangga Travindo in managing customer relationships can completed through a segmentation process, namely by tracing back the history of Hajj and Umrah registrants for a certain period. This customer segmentation process aims to determine customer behavior in order to implement marketing strategies accurately so as to increase profits for the company [2]. Finding customers is the company's most important task, but maintaining customer integrity is considered more important because losing customers means that all business flows with consumers will be cut off [3]. In a dynamic, competitive area, companies need to strive to gain knowledge about the needs and characteristics of consumers that must be maintained. However, in fact, companies constantly have customers who have varying behavior in carrying out transactions [4]. Customer segmentation is a model that groups suitable customers using specific criteria to be used as classification variables. Customers will be in the same environment with the same character if they have certain suitability criteria, while in a different environment there are customers who do not have the same character. After analyzing customer behavior using segmentation, the next step is to plan or prepare future marketing processes according to each customer segment.

Based on the situation described above, the researcher raised the title CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING WITH RFM METHOD (CASE STUDY: PT DEWANGGA TRAVINDO SEMARANG". Congregation data will be classified into several segments which are differentiated based on the characteristics of the congregation which are described through the RFM (Recency, Frequency, Monetary) model. In grouping customers, each customer is assessed for their profitability with the company through transactions that have been carried out using the RFM (Recency, Frequency, and Monetary) method. Grouping based on RFM has been used more than fifty years ago for the purposes of targeting each customer segment, reducing costs per order, and maximizing profitability[5]. Recency is the quantity of purchases since the final transaction was made, Frequency is related to the number of transactions carried out by customers in a certain period, and

Monetary is the size of transactions carried out by customers in a certain period [6]. By segmenting customers, it will make it easier for companies to implement marketing strategies that are appropriate for each type of existing customer, and will also certainly provide benefits for the company in increasing the quality and loyalty of customers towards the company [2]. This customer segmentation uses the Clustering method with the K-Means algorithm. Clustering is a method in data mining by dividing data in a combination into several groups with the similarity of the data in one group being greater than the similarity of the data with the data in other groups [7]. The K-Means algorithm is a clustering algorithm that is able to sort groups of data into several clusters based on the similarity of the data, so that data that has similar characteristics is grouped into one cluster and those that have different characteristics will be included in another cluster group that has similar characteristics [8]. Through a combination of the K-Means algorithm and the Recency Frequency Monetary (RFM) model, customer segmentation results can be used to provide customer scoring (customer scoring) and determine customer profiles more accurately than using the RFM model alone [9].

2. RESEARCH METHOD

2.1. State of The Art

The research activities carried out, utilizing previous research as a reference and guide in carrying out procedures in carrying out research. This literature review contains sources taken from scientific journals and articles from at least the last 5 years. In preparing this literature review, at least 2 other similar studies were quoted. In writing a literature review there are four parts, namely the name of the researcher along with the year, problem, method, and research results which are written in the form of a table in the Table 1.

Table 1. State of the art

No.	Researchers (Years)	Problem	Method	Result
1	Anissa Veronika Angelie, 2017	Customer Segmentation at PT. Building Superpower	Clustering K-Means dan model RFM	The visualization is displayed more interactively, web-based and combining several graphs and their contents.
2	Aulia Dewi, Fitria Abdurrachman, Nanan Yudi, 2018	Customer Segmentation at Belle Crown Malang	Metode K-Means Clustering, model RFM	The interactive visualization display combines several slightly different graphs.
3	Fakhri Hadi, dkk. 2017	Mapping and Supporting PT's Customer Management Strategy. Herbal Antidote to Indonesian Alwahidah Pekanbaru	K-Means Clustering berdasarkan RFM Mofek	The results of applying data preprocessing from transaction data are used in applying the RFM model, then the data is processed to determine clusters.
4	Sudriyanto, 2017	Selection of Potential and Loyal Customers at UD. Budi Luhur Probolinggo	Clustering dengan metode RFM dan Fuzzy C-Means	measuring validity by utilizing the Partition Coefficient Index (PCI) with FCM squared and squared produces increasingly large values (closer to 1) which means the quality of the clusters obtained is better [10]
5	Ariesty Rafika, 2015	Customer segmentation for marketing strategy for NANISA Beauty Clinic, Sidoarjo	SOM, Algoritma K-Means dan analisis LRFM	Visualization of neighbor distance in the SOM method produces six customer groups which show that
6	Muhammad Iqbal,2019	The Auliya Tour and Umrah company is having difficulty grouping its congregation.	Klastering dengan metode K-Means	Next, it will be grouped based on LRFM using the K-Means algorithm [11].

2.2. Customer Relationship Management

Customer relationship management is an integrated plan or strategy for identifying, acquiring and retaining customers. And allows organizations to control and coordinate customer relationships through various channels to help companies increase value for each customer. Simply put, customer relationship management is the implementation of business strategies that maximize revenue, profits and most importantly customer satisfaction by coordinating each customer segment, maintaining and increasing treatment that creates satisfaction for customers and implementing customer-centered techniques [13]. Relationships with customers provide a form for creating and maintaining good relationships between the company and each customer. In the field of information technology, customer relationship management is like an integration between business processes and the technology used to fulfill needs. customer. Correct management procedures must understand who the customer is? and what they like and what they don't like. This aims to anticipate customer needs and deal with customers more actively. Customer relationship management can find when customers are not happy, so the company can do things to make customers happy before the customer becomes too dissatisfied and turns to competitors.

Some organizations see customer relationship management as an advantage or even a competency to differentiate the company from competitors. Potential benefits from implementing customer relationship management include optimal marketing costs, more accurate customer targeting, minimizing sales campaign costs, increasing customer loyalty and retention, being able to recognize customer consumption trends and patterns, and easing the flow of information according to organizational needs. There are three phases, namely acquire, enhance, and retain [14] to process customer relationships according to Figure 2:

1. Get the latest customers (acquire), get the latest customers by promoting favorite products using the best services provided by the organization.
2. Growing profit power through existing customers (enhace), creating good relationships between the company and customers by providing the best service through extra comfort at low costs, for example implementing up selling and cross selling.
3. Maintain the integrity of profitable customers (retain), by focusing on services such as providing not what the market needs, but what customers need.

2.3. Clustering

Clustering is the process of grouping objects into segments based on high similarity of characteristics into segments and then each group created is different from other groups. Clustering or segments is an unsupervised data mining method, because there are no attributes used to guide the data in the learning process. Segment analysis in the background of data mining is the stage of placing customers in segments that have similar characteristics. The clustering algorithm forms a model by carrying out a series of iterations and then finishes when the model is centered and the segmentation boundaries are stable. Good cluster results can be seen from the similarity scale and the method applied. Based on the advice of Franley and Raftery, clustering methods are divided into 2 main groups, namely [22]:

1. Hierarchy Method. This method builds clusters by repeating partitions from top to bottom and bottom to top. The result is a dendogram diagram representing object segments and the level of similarity found in the grouping.
2. Partition Method. This method initiates partition 'k' at the beginning. The parameter 'k' describes the number of partitions to build. After that, use an iterative relocation technique, namely repeatedly trying to move objects from each group in order to get the optimal partition. Several partition methods such as K-Means, K-Medoids and CLARANS [13]. The Partition Method is used in this research because the aim is the same, namely to create segments in which each customer only falls into one particular group.

2.4. K-Means Clustering

The K-Means algorithm is a clustering algorithm that aims to divide data into various segments. K-Means is widely used in several fields such as data mining, statistical analysis, and other business applications. The clustering process is carried out by the computer by grouping the data used as input without knowing the target class, and getting input in the form of data without class labels. Each cluster has a center point that indicates that cluster. The following algorithm for working on K-Means, clustering [23]:

1. Look for the value of 'k' as a determinant of the number of clusters formed
2. Find the initial value for the segment center point. This stage of finding the initial value is determined randomly.
3. Find the distance between the center point and the point of each object, done using the Euclidean Distance.

4. Classify the data to form clusters with a center point in each cluster with the closest center point. Determining cluster members can be done by calculating the smallest distance to the object.
5. Update the center point value of each cluster.
6. Repeat stage two until the end until the central point value does not change.

2.5. Elbow

The Elbow method is used as a producer of information to determine the best number of clusters by looking at the percentage of comparison results between the number of segments that will form an elbow at a point [24]. The Elbow method provides an idea by selecting segment values and then adding the cluster values as a data model to determine the best cluster [3]. Elbow Method algorithm for determining the k value in K-Means [25]. Each cluster value has a different percentage result and is shown using a graph as a source of information. The graph will show several k values that have experienced the greatest decline and then the results of the k values will decrease slowly until the results of the k values are stable. However, the weakness of this method is that the elbow point cannot always be identified [26].

2.6. Min-Max

The Min-Max method is a simple normalization method by carrying out transformations on the original data. Min-Mix will adjust the specified limits by connecting to the original data [27]. This normalization technique transforms a numerical attribute in a smaller range or scale such as 0.0 to 1.0 [28], with the lowest limit of 0.0 and the highest limit of 1.0. The advantage of using Min-Max is that the comparison value between data before normalization is balanced with the data after normalization and no biased data is produced. Meanwhile, the weakness is that if the data is new, it will be possible to get caught in an "out of bound" error [30].

2.7. RFM Model

RFM analysis is a customer behavior analysis process used in database marketing and direct marketing [33]. RFM aims to determine customer segments based on three variables, namely Recency of the last purchase, Frequency of the purchase, and Monetary value of the purchase [34].

1. Recency, namely the gap between the last time a transaction was made and the current time. The smaller the range, the greater the R value.
2. Frequency, namely the number of times transactions are carried out by customers in a certain period. The more frequencies, the greater the F value.
3. Monetary, namely customer value in the form of the amount of money spent during transactions in a certain period. The more money the customer spends in that period, the greater the M value.

The greater the R and F values, the more likely the customer will make repeat transactions with the company. Apart from that, the greater the M value, the more likely customers are to respond to the company's products and services [35]. Calculating the RFM score for each customer determines the possibility that the customer will respond favorably to the company, for example regarding catalog promotions and offers. Customers who become new buyers and spend the most transactions in a certain period of time are the most likely to respond positively to company offers in the future. Different weightings for each variable can be determined in various ways, namely measurements obtained from experience so that you know the importance of each variable and based on the AHP process. There are two types of analysis:

1. When the RFM variables have the same importance, so the weight of each RFM variable has the same value.

2. When the RFM variable has different levels of importance depending on company characteristics [36].

2.8. Customer Lifetime Value (CLV)

Customer lifetime value or commonly called Customer Lifetime Value (CLV) is an understanding of the present value of all future profits obtained from customers, and must be calculated at the customer segment level which is adjusted to the business process [3]. The application of customer lifetime value in this final project uses the CLV value index approach. The methodology used is weighted RFM based on assessments from the hierarchical analysis process (AHP).

2.9. Research Stages

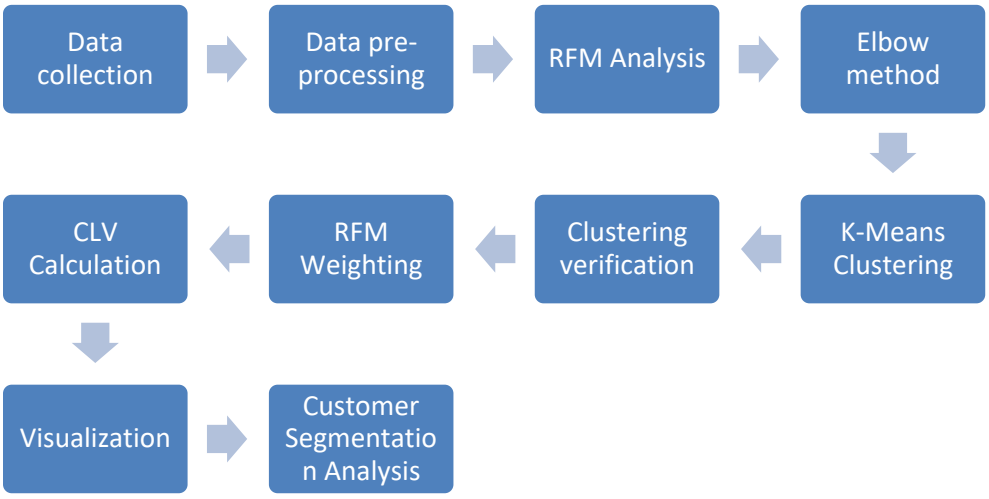

Figure 2. Research Stages

According to Figure 2, the research stages has been describe as follows :

1. At the data collection stage, the required data is collected as the main support in working on the final assignment. The data taken is adjusted to the topic and problem limits in the final project. In this stage, data was obtained from PT. Dewangga Travindo Semarang branch includes customer data for Hajj and Umrah candidates for the period 2018-2022. Apart from data collection, interviews were conducted with sources regarding current conditions related to the pandemic and related to segmentation as supporting information.
2. Data preprocessing. At this stage, raw data is processed to suit needs. Data preprocessing will carry out attribute selection, clean data rows that have empty values, combine raw data and process RFM analysis. The output of this process is data that is ready to enter the clustering stage. For incomplete data or in other words eliminating invalid data, cleaning is carried out to remove redundant data. Cleaning rows of data using features in Excel, namely remove duplicates, and removing unnecessary attributes, so that from 2400 data to 2144 data and leaving only the attributes used to carry out this final assignment, namely passport, cost and date.

Table 2. Sample dataset

Name	Sex	Star	Birth Of		Age	Date Of Paspor				Document			
			Place	Date		No	Issue	Expiry	Issuing Office	Book	Photo	KTP	KK
Hijrah Nurdin M Siga	F	*5	Palu	1979-11-18	40	B4039058	2016-05-04	2021-05-04	Semarang	✓	✓	✓	✓

Bandiyah Wasino Rana Karsya	F	*5	Wonosobo	1974-08-08	45	C5654691	2019-11-28	2024-11-28	Semarang	✓	✓	✓	✓
Imam Baedowi Purwadi	M	*5	Demak	1978-11-07	41	C3615832	2019-04-10	2024-04-10	Semarang	✓	✓	✓	✓
Eko Suwarni Sudirman	F	*5	Cilacap	1960-02-04	59	C4243043	2019-08-13	2024-08-13	Semarang	✓	✓	✓	✓

3. RFM Analysis. When carrying out a transformation, there are several stages, including changing the value into RFM form. Search for recency, frequency, and monetary attributes by aggregation using queries in the SQL Serven Management Studio application tool. The results are in the form of a file format and normalizes the data resulting from searching for RFM values.
4. Elbow Method (determining the k value). The Elbow method in this first stage aims to help determine values as input in implementing K-Means. The method used is the Elbow method. The initial process is to initiate the range of k values that will be processed in this method. The output obtained is in the form of a k value selected from the graphic results showing elbow points.
5. Clustering with K-Means. The Clustering process is carried out using RStudio tools. The results of the normalized RFM analysis will be processed using the Elbow method to obtain the k value, then clustering will be carried out using K-Means. The Cluster output that is formed will be given a name to make it easier to remember the characteristics of its customers. This process is carried out to find the customer segmentation owned by PT> Dewangga Travindo. The input to the K-Means process is the result of RFM analysis which has been normalized and the k value obtained using the Elbow method. Determining the centroid value (midpoint) and the distance of each object to the centroid. So the output is in the form of a cluster with a fixed centroid.
6. Verify Clustering Results. The process is carried out to ensure that each prospective congregation is appropriately grouped into that segment. This stage is carried out by calculating the distance of the prospective jamah to the center point of the group using the Euclidean Distance algorithm. At this stage, the cluster results will measure the level of performance of the cluster model that was formed in the previous stage. At this stage it will be known how well the clusters are separated and how closely related objects are in one cluster. Performance testing is carried out internally using SEE, connectivity and Dunn index calculations.
7. RFM weighting. The third weighting of this research is the AHP process. Previously, input from AHP was the result of a survey in the form of a questionnaire filled out by the company. AHP results will give different weights to each RFM variable (Recency, Frequency, Monetary). The weighting results will be carried out by a consistency ratio test to determine the consistency of the RFM weights that have been calculated.
8. CLV calculation. Before calculating the CLV index, first look for the normalized average of recency, frequency and monetary in each cluster. The value obtained is multiplied by the weight resulting from the AHP method. The CLV index is obtained from the sum of the RFM variable values in each cluster. The size of the large CLV index determines the level of customer loyalty
9. Creating Visualizations. This stage includes creating a prototype-based interface with the RShiny application. This visualization aims to make it easier to read clustering results so that you can plan customer relationship management more quickly. The visualization output is presented with diagrams and information in the form of index ranges and characteristics in each group formed.
10. Analysis of Customer Segmentation Results. Each cluster formed will be analyzed to determine customer characteristics. The results of the analysis will explain the

characteristics of each segment in the form of similarities in PT customer behavior. Dewangga Travindo and will indirectly explain the differences with other segments. Apart from that, analysis is carried out to visualize the graphs and images created.

3. RESULTS AND DISCUSSION

This process consists of 2 stages, namely finding the k value using the elbow method, then working on the clustering stage using the K-Means method. These two stages use the RStudio application tool. Before carrying out this process, first input a numeric data set and then initialize the name of the data set to make the calculation process easier. Like the following query (the results are data containing the Recency value which has been reversed, with other data needed). The data set that has been entered into data preprocessing is ready to be used. Next, look for the k value using the elbow method. The method for selecting the number of clusters is to observe a significant decrease in SSE and at that point the SSE value begins to stabilize, then that point becomes an elbow point on the graph. This stage uses a trial number of segments 1-10. The final stage is to look at the grouping results to see the elbow points that are formed.

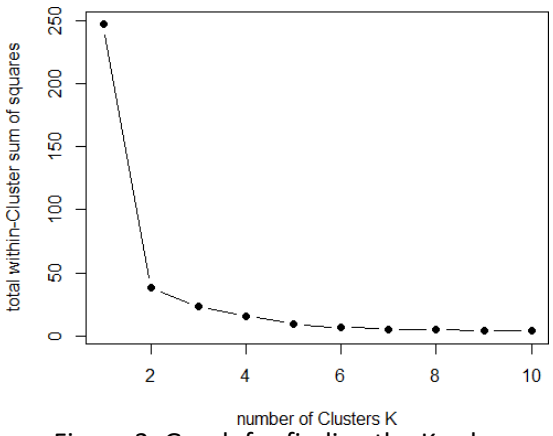

Figure 3. Graph for finding the K value

```
K-means clustering with 2 clusters of sizes 925, 1187

Cluster means:
  balik.rnew balik.fnorm balik.mnorm
1  0.2013081 0.001081081  0.03736216
2  0.8346841 0.005897220  0.02660489

Clustering vector:
   [1] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  [34] 1 1 1 1 1 2 2 2 2 1 2 1 1 2 1 2 1 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 2
  [67] 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 1 2 1 2 2 2 2 2 1 1
 [100] 2 1 2 2 2 2 2 1 2 1 2 1 1 1 1 1 1 1 1 1 2 2 2 1 2 1 1 1 2 2 2 2 2 1
 [133] 2 2 2 2 2 1 1 2 2 2 1 1 1 1 2 2 2 1 1 1 1 1 1 1 2 2 1 2 2 2 2 2 2 2
 [166] 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 1 1 2
 [199] 2 1 1 2 2 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 1 2 1
 [232] 2 1 1 1 1 1 2 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2
 [265] 2 2 2 1 1 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2
 [298] 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 2 1 1 1 2 2 2 1 2 1 2
 [331] 2 2 2 2 2 2 1 2 1 1 1 1 2 2 1 1 2 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1
 [364] 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1
 [397] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [430] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [463] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
 [496] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [529] 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [562] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
 [595] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [628] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [661] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [694] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [727] 2 1 1 1 1 1 1 2 2 1 2 2 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [760] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
 [793] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 2 2
 [826] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
 [859] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [892] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1
 [925] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
 [958] 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [991] 1 1 1 1 1 1 1 1 1 1
 [ reached getOption("max.print") -- omitted 1112 entries ]

Within cluster sum of squares by cluster:
[1] 19.97560 18.11929
 (between_SS / total_SS =  84.6 %)
```

Figure 4. Cluster Vector of K-Means over 2 cluster

The plot resulting from the Elbow method is depicted as shown in the image. The resulting points that form an elbow are between points 1 - 3, after there is no drastic decrease, so the conclusion is that the number of clusters according to the Elbow method is 2 clusters. After getting the k value, the clustering calculation stage can be carried out. Starting from creating an initiation for the name which contains the function to activate K-Means with the R application then inputting 2 k values (from the results of the elbow calculation). K-Means can appear with the initiation call. Next, combine 'back' and 'cluster' using the "data.frame" function and then save it using CSV format using the "write.csv" function in R.Figure 4. Output Centroid of 2 Cluster.

This process can determine whether each customer is exactly in the cluster created, therefore you must verify the data using Euclidean Distance. The distance to the center point is calculated for each customer. This distance calculation can see how close the customer is to the center point. The segment center point table is split, then a new column is given for the results of the congregation's distance to the center point. The stage for searching min-values in columns and rows with the correct number of center points, which contains the results of calculating the distance and appearance of the column with the minimum value in this study is used (2-3). Clustering data that is divided into 2 clusters needs to be tested for performance to determine the optimal level, using other methods. In this test, the Connectivity, SSE and Dunn Index methods are used. The SSE method requires normalized data from each RFM attribute, the same as the Dunn Index and Connectivity methods. The SSE algorithm calculates the square value of the distance of each data to the center point, then calculates the total. The results are displayed with a diagram plot and SSE values per segment. The SSE method in R uses a query as in Figure 5.

```
> View(balik)
> datass <- data.frame(balik$rnew, balik$fnorm, balik$mnorm)
> wss = kmeans(datass, centers =1)$tot.withinss
> for (i in 2:10) wss[i] = kmeans(datass, centers=i)$tot.withinss
> plot(wss, xlab='Number of Clusters', ylab='Within-Clusters sum of squares')
> write.csv(wss, 'sse.csv')
```

Figure 5. Query for searching the SSE Value

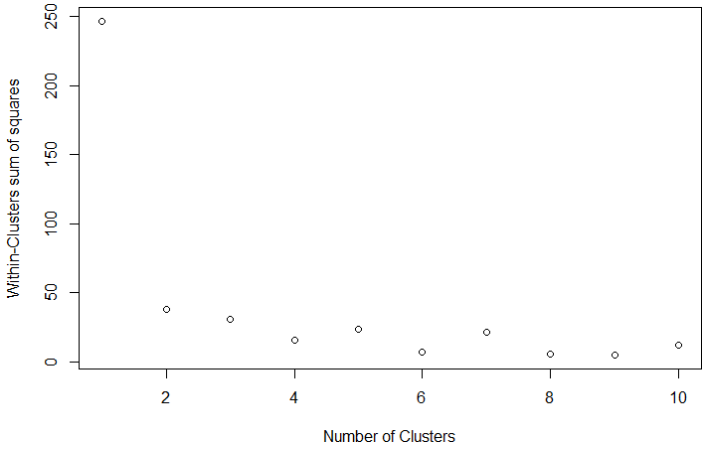

Figure 6. SSE Graphic

In the Connectivity and Dunn Index methods, all that is needed is the normalized value of each RFM attribute and a column containing the name of each row. Dunn Index can select the number of segments with the highest value. Meanwhile, connectivity selects many segments with the lowest value. Connectivity and Dunn index in R, using Query : don't forget to install package (clValid). In this test, it has been obtained:

- In the SSE method. The graphic display is almost the same as the Elbow method, and the SSE / WSS values are the same as the Elbow method.
- On the Dunn Index and Connectivity methods. The connectivity method chooses 3 segments, because it has the lowest value of 6.5206. Meanwhile, the Dunn Index method selects 3 segments because the lowest value is 0.4488. Not suitable for selecting the number of clusters using the Elbow method. Then do the clustering stage again using 3 k values.

The next stage of RFM analysis is related to the initial stage of CLV analysis or weighting of RFM with AHP. AHP data was obtained from filling out a questionnaire, there were 3 respondents who filled out the questionnaire, namely General Manager, Finance and Marketing. The questionnaire given contains several respondent data such as the respondent's name, the respondent's position, and how long the respondent has served. The guidelines for filling in are similar to the explanation of the RFM and preference scale (comparison table of criteria and scores) in chapter 2. The results of making a questionnaire using Google Form are presented in the attachment. The weighting process uses data obtained after filling out the questionnaire, which is then averaged by looking at the comparison of each criterion.

Table 3. Normal Comparisson Matrix

Comparisons Matriks				Bobot
Keriteria	Recency(R)	Frequency(F)	Monetary(M)	
Recency(R)	0,30	0,29	0,31	0,30
Frequency(F)	0,66	0,66	0,65	0,66
Monetary(M)	0,04	0,04	0,04	0,04

The weights obtained need to be tested for consistency, so that they know whether the respondents' filling in the questionnaire is consistent or not. Each element in the pairwise comparison table is multiplied by the weight of the criteria in the comparison matrix. The result is an overall score obtained by adding up each row as shown in Table 4. The x value can be found

using the overall score results as input to calculate the CI, by dividing each overall score row by the weight. Next, the results are averaged as shown in Table 5.

Table 4. Overall Score

Keriteria	Recency(R)	Frequency(F)	Monetary(M)	Overal Score
Recency(R)	0,30	0,29	0,31	0,902
Frequency(F)	0,68	0,66	0,64	1,973
Monetary(M)	0,04	0,04	0,04	0,127

Table 5. Result of x value

Recency(R)	= 0,902 / 0,30 = 3,0007
Frequency(F)	= 1,973 / 0, 66 =3,0016
Monetary(M)	= 0,127 / 0,04 = 3, 0001
Hasil X	$\frac{(3,0007 + 3,0016 + 3,0001)}{3} = 3,0008$

Based on the AHP calculation process, it can be seen that Frequency (F) is an important criterion used in assessing the best customers. After Frequency, the Recency value is prioritized compared to Monetary. The final choice criteria is Monetary. The CI value obtained is ≠ 0 because the weight obtained is based on inconsistent questionnaire filling, but it can be accepted because the CI value obtained is smaller than 0.1, namely 0.0006. Data from cluster results and cluster results that have been tested for performance are needed in later CLV calculations. It is necessary to search for the average index for each variable in the CLV calculation. The average index can be obtained from calculating the average RFM in each cluster (midpoint value per cluster) that has been obtained.

Table 6. Average Value of RFM 2 Cluster

Cluster	SR	SF	SM
1	0.2013081	0.001081081	0.03736216
2	0.8346841	0.00589722	0.02660489

Table 7. Average Value of RFM 3 Cluster

Cluster	SR	SF	SM
1	0.9106159	0.009319287	0.03042139
2	0.7524912	0.002192982	0.02247368
3	0.2013081	0.001081081	0.03736216

Then the weights from the AHP calculations are used for CLV calculations, by adding up the product of the average RFM index (SR, SF, SM) with the weight RFM (BR, BF, BM). In Table 9, the ranking of each cluster for 2 clusters. Cluster 2 received rank 1 because its CLV received a higher total recency, frequency, monetary value compared to the recency, frequency, monetary value in cluster 2. Judging from the graph formed from the visualization, the following characteristics are obtained as shown in Table 10.

Table 8. Rating of CLV 2 Cluster

Cluster	(SR * BR)	(SF*BF)	(SM*BM)	Total	Peringkat
1	0,06039	0,00071	0,00149	0,06260	2
2	0,25041	0,00389	0,00106	0,25536	1

Table 9. Rating of CLV 3 Cluster

Cluster	(SR * BR)	(SF*BF)	(SM*BM)	Total	Peringkat
1	0,27318	0,00615	0,00122	0,28055	1
2	0,22575	0,00145	0,00090	0,22809	2
3	0,06039	0,00071	0,00149	0,06260	3

Table 10. Characteristic of 2 Clusters

Cluster 1	Cluster 2
Peringkat = 2	Peringkat= 1
Anggota = 925	Anggota=1187
Karakteristik = Recency: 122-349 hari Frequency:1-3 kali Monetary: 23000000-74200000	Karakteristik = Recency: (-139)- 15 hari Frequency:1-5 kali Monetary: 22000000-141300000

Table 11. Characteristic of 3 Clusters

Cluster 1	Cluster 2	Cluster 3
Peringkat = 1	Peringkat = 2	Peringkat = 3
Anggota= 617	Anggota = 570	Anggota= 925
Karakteristik = Recency: (-139)- (-70) hari Frequency:1-5 kali Monetary: (Rp) 22000000-141300000	Karakteristik = Recency: (-46)- 15 hari Frequency: 1-2kali Monetary:(Rp) 22500000-52900000	Karakteristik = Recency: 122-349 hari Frequency: 1-3kali Monetary: (Rp) 23000000-74200000

4. CONCLUSION

Based on the research carried out, it was concluded that the K-Means method could be an option for solving problems in customer segmentation. In applying the K-Means method, the results are tested first to be more precise in determining clusters, namely testing using the connectivity and dunn index methods to test the results of the elbow method calculation which in this study produces 2 segments. After testing, it turned out that the correct segment was 3 segments. Meanwhile, the RFM method used functions as a description of the character of each cluster that has been formed because it can determine the transaction behavior of each congregation at PT. Dewangga Travindo Semarang. Result:

• In 2 segments. Segment 1 is ranked 2nd, with 925 members, with criteria R= 122- 349 days, F= 1-3 times, M = IDR 23,000,000-IDR 74,200,000. • Segment 2 is ranked 1st, with 1187 members, with criteria R= (-139)-15 days, F= 1-5 times, M=Rp. 22,000,000-Rp. 141,300,000,

• In 3 segments. Segment 1 is ranked 1st, with 617 members, with the criteria R=(- 139)-(-70) days, F=1-5 times, M= IDR 22,000,000-IDR 141,300,000. Segment 2 is ranked 2nd, with 570 members, with criteria R= (- 46)-15 days, F=1-2 times, M= IDR 22,500,000-IDR 52,900,000. Segment 3 is ranked 3rd, with 925 members, with criteria R=122-349 days, F=1-3 times, M=Rp23,000,000-74,200,000.

In the entire segment, the best customer obtained was Anetonia, and the less profitable customer was Sri. There is a relationship between the RFM model and the K-Means method to be able to form the right segments according to the conditions of the congregation because each congregation definitely has a different RFM value which is then processed using the K-Means method, then the appropriate segment is obtained by considering the distance to the closest point. RFM of each congregation, as well as knowing which congregations are profitable. In this prototype-shaped visualization, it can appear interactive by combining various graphs and box plots made according to existing segments. For further research, it would be better to add time spans for transaction data and variables so that the analysis carried out is sharper and broader, such as adding a location map to find out in which areas profitable customers are spread.

REFERENCES

[1] karir.com, “PT Dewangga Tour & Travel.” https://www.karir.com/companies/25471.

[2] F. Hadi, M. Mustakim, D. O. Rahmadia, F. H. Nugraha, N. P. Bulan, and S. Monalisa, “Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru),” *J. Sains dan Teknol. Ind.*, vol. 15, no. 1, pp. 69–76, 2017.

[3] F. T. Informasi, “Segmentasi Pelanggan Menggunakan Clustering K-Means Dan ModelRfm (Studi Kasus : Pt . Bina Adidaya Surabaya) Customer Segmentation Using K- Means

Clustering and Rfm Model (Case Study : Pt . Bina Adidaya Surabaya SegmentasiPelanggan Menggunakan Clusteri," 2017.
[4] Aviliani, U. Sumarwan, I. Sugema, and A. Saefuddin, "Segmentasi nasabah tabungan mikro berdasarkan recency, frequency, dan monetary : kasus bank bri," *Financ. Bank. J.*, vol. 13, no. 1, pp. 95–109, 2011.
[5] R. Kohavi and R. Parekh, "Visualizing RFM segmentation," *SIAM Proc. Ser.*, no. April,pp. 391–399, 2004, doi: 10.1137/1.9781611972740.36.
[6] M. I. Istiana, "Segmentasi Pelanggan menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran pada LAROIBA Seluler," vol. 1, pp. 3–4, 2013.
[7] D. Zheng, "Application of silence customer segmentation in securities industry based on fuzzy cluster algorithm," *J. Inf. Comput. Sci.*, vol. 10, no. 13, pp. 4337–4347, 2013, doi: 10.12733/jics20102432.
[8] G. F. Wulandari, "Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) Pada Hijab Miulan," *Ind. Mark. Manag.*,vol. I, no. segmentasi pelanggan, p. 7, 2014.
[9] Y. Nugraheni, "Data Mining Using Fuzzy Theory for Customer Relationship Management," *Lontar Komput.*, vol. 4, no. 1, pp. 188–200, 2013.
[10] Sudriyanto, "Clustering Loyalitas Pelanggan Dengan Metode RFM (Recenty, Frequency, Monetary) dan Fuzzy C-Means," *Pros. SNATIF Ke-4*, pp. 815–822,2017.
[11] A. Rafika, *Segmentasi Pelanggan Menggunakan Som, Algoritma K-Means Dan AnalisisLrfm Untuk Penyusunan Rekomendasi Strategi Pemasaran Pada Klinik Kecantikan Nanisa, Sidoarjo*. 2015.
[12] M. Iqbal, "Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 97–104, 2019.
[13] F. Buttle, *Customer Relationship Management-Concepts and Technologies (Second Edition)*, Second., vol. 53, no. 9. Hungary: Elsevier, 2013.
[14] R. Kalakota and M. Robinson, *E-Business 2.0 - Roadmap for Success Second Edition*. Canada: Wiley Publishing, Inc, 2000.
[15] D. P. Hidayatullah, R. I. Rokhmawati, and A. R. Perdanakusuma, "Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2406–2415, 2018.
[16] M. J. A. Berry dan G. S. Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management*. Canada: John Wiley & Sons, Inc, 2000.
[17] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
[18] R. D. Syah, "Metode Decision Tree Untuk Klasifikasi Hasil Seleksi Kompetensi DasarPada Cpns 2019 Di Arsip Nasional Republik Indonesia," *J. Ilm. Inform. Komput.*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.
[19] D. D. Efraim Turban, Ramesh Sharda, *Decision Support And Business Intelligence Systems (9th Edition)*. New Jersey: Pearson, 2011.
[20] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
[21] Kenneth Jensen, "Cross-Industry Standard Process for Data Mining (CRISP-DM)."
[22] C. Fraley and A. E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis," *Comput. J.*, vol. 41, no. 8, pp. 586–588, 1998, doi: 10.1093/comjnl/41.8.578.
[23] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm," *J. Transform.*, vol. 8, no. 1, p. 33, 2010, doi: 10.26623/transformatika.v8i1.45.

[24] E. Muningsih and A. B. S. I. Yogyakarta, “Optimasi jumlah cluster k-means dengan metode elbow untuk pemetaan pelanggan,” *Pros. Semin. Nas. ELINVO*, no. September,pp. 105–114, 2017.
[25] B. Purnima and K. Arvind, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014,[Online]. Available: https://www.ijcaonline.org/archives/volume105/number9/18405-9674.
[26] T. M. Kodinariya and P. R. Makwana, “Review on determining number of Cluster in K-Means Clustering,” *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 2321–7782, 2013.
[27] S. G. K. Patro and K. K. sahu, “Normalization: A Preprocessing Stage,” *Iarjset*, pp. 20–22, 2015, doi: 10.17148/iarjset.2015.2305.
[28] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, “Data Transformation pada Data Mining,” *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.
[29] Y. C. Sitanggang, C. Dewi, and R. C. Wihandika, “Pemilihan Rute Optimal Penjemputan Penumpang Travel Menggunakan Ant Colony Optimization Pada Multiple Travelling Salesman Problem (M-TSP),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 3138–3145, 2018.
[30] E. Martiana, *Data Preprocessing : Data Transformation*. Surabaya: EEPIS-ITS, 2013.
[31] A. Aigustin, “Identifikasi Tanda Tangan Menggunakan Manhattan Distance dan Sum Square Error dengan Ekstraksi Ciri Dimensi Fraktal,” 2014.
[32] A. R. H. Sisca Indah Pratiwi, Tatik Widiharih, “ANALISIS KLASTER METODE WARD DAN AVERAGE LINKAGE DENGAN VALIDASI DUNN INDEX DAN KOEFISIEN KORELASI COPHENETIC (Studi Kasus: Kecelakaan Lalu Lintas Berdasarkan Jenis Kendaraan Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018),” *J. gaussian*, vol. 8, pp. 486–495, 2019, [Online]. Available: http://ejournal3.undip.ac.id/index.php/gaussian.
[33] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.
[34] Arthur M. Hughes, *Strategic Database Marketing*. Chicago: Probus Publishing, 1994.
[35] Z. Li, “Research on customer segmentation in retailing based on clustering model,” *2011Int. Conf. Comput. Sci. Serv. Syst. CSSS 2011 - Proc.*, pp. 316–318, 2005, doi: 10.1109/CSSS.2011.5974496.
[36] C. H. Cheng and Y. S. Chen, “Classifying the segmentation of customer value via RFMmodel and RS theory,” *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 4176–4184, 2009, doi: 10.1016/j.eswa.2008.04.003.
[37] B. W. Taylor, *introduction to Management science*, 11th ed., vol. 83, no. 3. United Statesof America: Prentice Hall, 2013.