ORIGINAL RESEARCH

PATIENT SEGMENTATION BASED ON CUSTOMER LIFETIME VALUE ANALYSIS USING RECENCY, FREQUENCY, MONETARY AND INTERPURCHASE TIME MODELS

Mohamad Shodikin[1] | Chastine Fatichah*[1] | Nungky Taniasari[2]

[1]Informatics Departement, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[2]Faculty of Health Sciences, Universitas Anwar Medika, Sidoarjo, Indonesia

Correspondence

*Chastine Fatichah, Dept of Informatics , Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: chastine@if.its.ac.id

Present Address

Gedung Teknik Informatika, Jl. Teknik Kimia, Kampus ITS Sukolilo, Surabaya 60111, Indonesia

Abstract

Hospitals strategically provide quality health services to the surrounding community. One of the strategic roles is realized in inpatient and outpatient services. Based on the background above, the author examines patient segmentation based on Customer Lifetime Value analysis using the RFMT model. The research stages include calculating the RFMT score, clustering using the K-Means and DBSCAN algorithms, and patient segmentation based on CLV analysis. The dataset in this study was obtained from inpatient and outpatient visits at a hospital from January to December 2022. The research results show four segmentations of outpatients and four inpatients based on CLV values: Champions, Loyal Customers, Potential Loyalists, and Lost Customers. Inpatient segmentation includes the Champion category, 473 (2%) patients belonging to 14 clusters (15.56%). The Loyal Customer Category is 1,727 (7%) patients who are members of 31 clusters (34.44%). The Potential Loyalist category is 3,874 (16%) patients who are members of 31 clusters (34.44%). The Lost Customer Category was 18,516 (75%) patients from 14 clusters (15.56%). Outpatient segmentation includes: Champion category, 3,512 (8%) patients belonging to 10 clusters (10.10%). The Loyal Customer Category is 5,661 (13%) patients who are members of 24 clusters (24.24%). The Potential Loyalist category is 11,070 (25%) patients who are members of 34 clusters (34.34%). In the Lost Customer Category, 23,678 (54%) patients belong to 31 clusters (31.31%).

KEYWORDS:

Fuzzy AHP, CLV, Clusterization, Patient Segmentation, RFMT

1 | INTRODUCTION

A hospital is an organization that has professional and organized medical personnel and permanent medical facilities providing medical services, continuous nursing care, diagnosis, and treatment of diseases suffered by patients[1][2]. The hospital has a mission to provide comprehensive health services (preventive, promotive, curative, and rehabilitative) that are quality and affordable to improve public health. Hospitals have a strategic role in providing quality health services for the surrounding community. One of these strategic roles is realized in inpatient and outpatient services. Understanding the characteristics of inpatients and outpatients based on demographic status, financial status preferences for types of services, and so on is necessary to determine the right marketing strategy to maintain patient visits. The above is known as customer segmentation.

Customer segmentation can be done with demographic, geographic, behavioral, and psychological data on customers[3]. Elrod et al.[4] analyzed several demographic variables (e.g., age, gender, income) to identify important customers. Customer value for a company or organization is determined by the value of Customer Lifetime Value (CLV) during the customer life cycle[5]. LV values help companies and organizations allocate the limited resources available to their customers by categorizing them and assigning a certain weight to each customer[6]. Calculating the appropriate CLV value can help organizations categorize and classify their customers based on CLV ranking[7][8]. CLV ranking is evaluated with one of the popular models, the RFM model, for emphasizing customers who benefit the company[9]. Chen et al.[10] examined customer segmentation based on the RFM model in the online retail industry. From the results of the customer grouping, you can predict which customers have the highest purchasing preferences. Previous studies that focused on the history of inpatient and outpatient medical services for patient segmentation have not been carried out much. So, this research proposes patient segmentation based on Customer Lifetime Value (CLV) analysis using the Recency, Frequency, Monetary, and Interpurchase Time (RFMT) model. This research is expected to be able to fill this gap.

2 | PREVIOUS RESEARCHES

Customer lifetime value (CLV) is the value consumers feel when the costs are paid to obtain a product or service, where, in the end, consumers get an overall evaluation of the product or service's effectiveness[11]. CLV refers to the present value of all benefits created by consumers for the company in the entire process of maintaining a relationship with the company[[12][13]][14]. The future cash flow brought by consumers to the company and the maintenance level of consumer groups can be calculated according to the company's past transactions[9]. In non-contractual relationships, consumers can conduct transactions with several companies simultaneously, and the costs to consumers of switching between several companies are very low. Currently, consumer lifetime value calculations focus on estimating the probability or frequency of a consumer's future repeat purchases based on newness[11][15]. Customer segmentation can be divided into four customer categories based on the CLV value, sequentially the level of customer loyalty based on the CLV index calculation: Champions, Loyal Customers, Potential Loyalists, and Lost Customers[16].

A model called Recency, Frequency Analysis, And Monetary (RFM) was proposed by Hughes[17] to analyze customer behavior. Over the past few decades, the model has become a common paradigm for behavioral segmentation [[16][18]]. The RFM model can characterize customer purchasing behavior by investigating the elapsed time from the last purchase (Recency), the number of purchases (Frequency), and the total expenditure (Monetary) of each customer in a certain period. The customer clustering model with RFM does not focus on attracting new customers but rather on identifying the best customers for marketing performance targets[19]. Recency is the number of days since the last transaction. Frequency is the total number of transactions. Monetary is the total value of transactions during a certain period[19]. To improve accuracy and obtain more information, Zhou et al.[20] In 2021, a new variable purchase time (T) was added to calculate the time between transactions since the patient's first transaction.

Several new approaches have been extended from the RFM model. Yeh et al.[21] extended the RFM model by adding "time since first purchase" (T) and "probability of churn" (C). Yoseph et al.[22] introduced the rate of change in purchases (C) to indicate the quantity and sign of changes in customer purchasing behavior. Interpurchase Time (T), defined as the interval between two consecutive purchases by a customer in the same store or on the same website, has been frequently used to analyze shopping behavior since the mid-1960s[23]. Vakratsas et al.[24] used T to study the relationship between online shopping regularity and propensity. Similarly, Meyer Waarden[25] introduced T to investigate the impact of loyalty programs on customer purchasing

behavior. Guo[26] developed a multi-category inter-purchase timing model to predict purchase likelihood for individual customers. Also, Junpeng Guo et al.[27] utilized a multi-category inter-purchase time model to increase the effectiveness of product recommendations.

A low R-value indicates that consumers have consumed the company's products or services in a shorter time, so consumers are more likely to be touched by the company's marketing information. Thus, such consumers are more likely to make repeat purchases quickly, and companies can obtain greater revenues by investing fewer costs in these consumers[11]. Meanwhile, if the R-value is large, consumers will not consume the company's products or services long.

The F value refers to the frequency of customers purchasing a company's products or services in a statistical period. Generally, the greater the customer's F value, the higher the frequency customers purchase the company's products or services, the higher the customer loyalty, and the greater the customer value[11]. Customer purchase frequency is often considered in combination with customer purchase immediacy. When a customer's purchasing frequency is high, the closer the purchase time is to now, the more likely the customer will continue to consume the company's products. When the time of purchase is far from now, such customers were once valuable customers for some time, but now they are likely to be lost.

The M value refers to the total customer consumption of a company's products or services in a statistical time. In general, the higher the number of purchases of a company's products or services by customers, the lower the amount of consumption of alternative products or services, the more loyal the customer, and the higher the value[11]. Researchers believe that RFM can well reflect customers' historical consumption behavior, and a CLV value prediction model can be formed based on the three variables R, F, and M[9][28][29].

The RFMT model assessment process is carried out for RFMT analysis using the quantile method. Each customer is given a quantile score with the highest frequency value being 5, while the other quantiles are given scores of 4, 3, 2, and 1 in rank order. This process is also carried out for monetary. The opposite is true for quantiles. The smallest recency and interpurchase time are given a score of 5, while the quantile Recency and other interpurchase times are respectively scored 4, 3, 2, and 1[19].

After RFMT score calculation, algorithm clustering can group RFMT data into certain clusters[30][31]. Clustering is the process of grouping a set of data objects with similar values in data attributes into one group or cluster, and each cluster formed does not have similar values to other clusters[32]. The purpose of clustering is to organize data into classes so that there is high and low intra-class similarity[33][12]. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm can find high-density core samples and expand clusters from these samples. Two main parameters of the algorithm are determinedimum number of samples and epsilon. The first parameter determines the minimum number of points classified as a core sample. This parameter defines the noise tolerance level of the algorithm[34]. The DBSCAN algorithm can find any cluster of any shape and effectively identify existing noise points[35]. Meanwhile, K-Means is an algorithm that is easy to use to cluster large amounts of data and outlier data very quickly based on the level of similarity and dissimilarity of objects[36][37][38]. K-Means Clustering can progressively improve the quality of clustering[10].

From an optimization perspective, the main goal of clustering is to maximize homogeneity (internal) within a cluster[1]and heterogeneity (external) between different clusters 1. In research by Rimelda et al.[39] comparing the DBSCAN and K-Means algorithms for grouping countries in the world regarding recovery and recovery rates in handling the Covid-19 pandemic, the conclusion is that the K-Means algorithm is superior to DBSCAN with the best Silhouette Index value of 0.6902 with a value of k = 8. Research conducted by Mustakim et al.[40] regarding the application of the DBSCAN algorithm in clustering Twitter text quotes for the Pekanbaru City Election, a combination of Eps = 0.1 and MinPts = 10 with the best Silhouette Index value of 0.413.

3 | METHOD

The process diagram used to segment patients can be seen in Figure 1 .

Based on Figure 1 , it shows that the research started from collecting data on outpatient and inpatient visits. The RFMT model assessment process is carried out for RFMT analysis using the quantile method. After calculating the RFMT score, the clustering algorithm can group RFMT data with the DBSCAN and K-Means algorithms. Next, the DBSCAN and K-Means algorithms
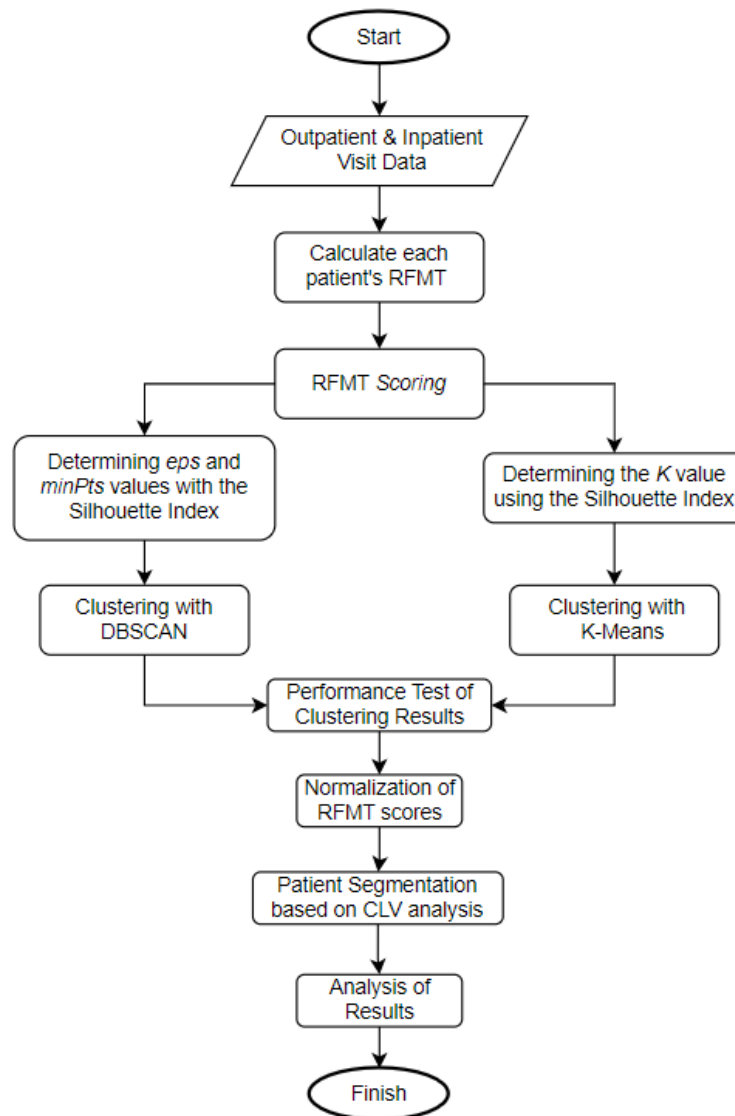
FIGURE 1 Research method

compare their quality using the Silhouette Index (SI) method to get the best clustering algorithm. After forming clusters on the inpatient and outpatient datasets, a data normalization process was carried out on the Recency (R), Frequency (F), Monetary (M), and Interpurchase Time (T) attributes using the Min-Max method for the CLV analysis process. Determination of CLV in this study uses normalized RFMT values using the min-max and weighted RFMT methods.

3.1 | Datasets

This study uses two historical datasets of patient visits: inpatient visits and outpatient visits. The dataset was obtained from SIMRS Private Hospital X from January to December 2022.

Based on Table 1 , it can be seen that the outpatient dataset consists of 10 features, and there are 185,196 visit data. The outpatient dataset shows every patient who visited the hospital to access outpatient clinic services. These are identified based on patient ID, visit ID, visit date, guarantor, billing, gender, patient age, and medical diagnosis. Age category data based on the Indonesian Ministry of Health (2009) is divided into six groups, namely toddlers (0 - 5 years), children (6 - 11 years), teenagers (12 - 25 years), adults (26 - 45 years), elderly (46 - 65 years), and seniors (66 - above). The patient's medical diagnosis data uses International Classification of Diseases (ICD) codes.

TABLE 1 The features on outpatient services datasets.

Feature	Description
Patient ID	A unique number that each patient has
Visit ID	Number of patient visits to the hospital
Date of visit	Date of patient visit to hospital
Destined clinic	The destination clinic service location
Service guarantor	How to pay or guarantee patient services
Service bill	Total service charge bill
Gender	Patient gender
Age group	Patient age group categories (toddlers, children, teenagers, adults, elderly, seniors)
Primary medical diagnosis	ICD10 patient's primary medical diagnosis
Secondary medical diagnosis	ICD10 secondary medical diagnosis of the patient

TABLE 2 Features on outpatient services datasets

Feature	Description
Patient ID	A unique number that each patient has
Visit ID	Number of patient visits to the hospital
Date of visit	Date of patient visit to hospital
Service guarantor	How to pay or guarantee patient services
PatientClassantor class	The class of service rights that the patient has at the time of servClass
Service class	Class of service received by the patient during service
Service bill	Total service charge bill
Gender	Patient gender
Age group	Patient age group categories (toddlers, children, teenagers, adults, elderly, seniors)
Primary medical diagnosis	ICD10 patient's primary medical diagnosis
Secondary medical diagnosis	ICD10 secondary medical diagnosis of the patient
Length of hospitalization	The patient's length of stay in hospital

Based on Table 2 , it can be seen that the inpatient dataset consists of 12 features, and there are 29,687 visit data. The inpatient dataset shows every patient who visits the hospital to access inpatient services identified based on patient ID, visit ID, visit date, guarantor, bill, guarantor class, service class, gender, patient age, length of stay,y, and medical diagnosis. Age category data based on the Indonesian Ministry of Health (2009) is divided into six groups, namely toddlers (0 - 5 years), children (6 - 11 years), teenagers (12 - 25 years), adults (26 - 45 years), elderly (46 - 65 years), and seniors (66 - above). The patient's medical diagnosis data uses International Classification of Diseases (ICD) codes.

The outpatient research dataset consists of 10 features, with 185,196 visit data. Outpatients who had visited more than once were 25,101 patients or 57% of the total outpatients. The inpatient dataset consists of 11 features and 29,687 visit data. Inpatients who had visited more than once were 3,287 patients or 13% of the total inpatients. Age category data based on the Indonesian Ministry of Health (2009) is divided into six groups, namely toddlers (0 - 5 years), children (6 - 11 years), teenagers (12 - 25 years), adults (26 - 45 years), elderly (46 - 65 years), and seniors (66 - above). The patient's medical diagnosis data uses International Classification of Diseases (ICD) codes.

3.2 | RFMT Analysis

The patient medical record number attribute in the outpatient and inpatient datasets determines the RFMT value. The time interval from the last patient visit to the data collection date is calculated as the Recency (R) value. The number of patient visits is calculated as the Frequency (F) value. All service bills from the same patient are accumulated as a Monetary value (M). The fourth variable, Inter purchase Time (T), measures the average distance between consecutive visit patterns. Suppose the first customer and last purchase dates are noted as T1 and Tn , respectively. In that case, the customer's holistic shopping cycle (L) can be approximated by the months between T1 and Tn, and thus, the value of T (in months) can be calculated using equation 1.

$$T = L/(F - 1) = (Tn - T1)/(F - 1) \tag{1}$$

TABLE 3 The RFMT value for outpatient.

Value	Mean	Minimum	Maximum
R	137.00	1.00	365.00
F	4.54	1.00	97.00
M	1234810.90.00	7000.00	27636522.00
T	12.12	0.00	177.00

TABLE 4 The RFMT value for inpatient.

Value	Mean	Minimum	Maximum
R	168.48	1.00	365.00
F	1.20	1.00	21.00
M	6931461.00	544502.00	207421522.00
T	6.09	0.00	179.00

TABLE 5 Inpatient RFMT scoring rules

Category	Score	R (Day)	F	M (million)	T (Day)
Very low	1	> 210	1	< 0.6	0 or > 210
Low	2	150 – 210	2	0.6 – 3	150 – 210
Medium	3	90 – 150	3	3 – 6	90 – 150
High	4	60 – 90	4	6 – 9	60 – 90
Very High	5	< 60	> 4	> 9	< 60

Only patients who made at least two purchases (i.e., $F >= 2$) in a given period were considered to calculate T. After preprocessing the RFMT data; the dataset included 185,196 outpatient visits involving 43,921 patients and 29,529 inpatient visits involving 24,590 patients. The table below shows the RFMT values for outpatients and inpatients.

Based on Table 3 , it can be seen that the number of outpatient visits per year of 43,921 patients shows an average Recency (R) value for outpatients of 137 days, where the minimum value is 1 day, and the maximum value is 365 days. Meanwhile, the Frequency (F) value averaged 5 times in the research period. The minimum value is 1 time, and the maximum value is 97 visits a year. Based on the transaction value (M), the average outpatient costs IDR 1,234,810 per year. The minimum fee incurred is IDR 7,000, and the maximum transaction fee is IDR 27,636,522. The average distance between visits was 12 days. The largest interval between visits was 177 days.

Based on Table 4 , it can be seen that the number of visits per year for inpatients is 24,590 patients, where the average Recency (R) value for inpatients is 168 days, with a minimum value of 1 day and a maximum value of 365 days. Meanwhile, the Frequency (F) value averages one hospitalization per year. The minimum value is 1 time,e, and the maximum value is 21 visits a year. Based on the transaction value (M), the average cost for inpatients is IDR 6,931,461 per year. The minimum fee incurred is IDR 544,502, and the maximum transaction fee is IDR 207,421,522. The average distance between visits was 6 days. The largest interval between visits was 179 days.

To analyze clusters, we need to categorize the parameters R, F, M, and T into five categories (Very Low, Low, Medium, High, and Very High) with quintile scores of 1 to 5. Based on the Focus Group Discussion (FGD) results with Hospital Management, we agreed to categorize RFMT parameters, as shown in the following table.

Based on Table 5 , it shows that the patient category is very low if the patient's last visit was more than 210 days, the frequency of visits is only 1 time in one period, where the patient's bill value is less than IDR 600,000.00. The distance between visits is zero or greater than 210 days. The patient category is low if the patient's last visit was between 150 - 210 days, the frequency of visits is 2 times in one period, where the value of the patient's bill is between IDR 600,000.00 - IDR 3,000,000.00 and the distance between visits is 150 - 210 days. The patient category is medium if the patient's last visit was between 90 - 150 days, the frequency of visits is 3 times in one period, where the patient's bill value is between IDR 3,000,000.00 - IDR 6,000,000.00 and the distance between visits is 90 - 150 days. The patient category is high if the patient's last visit was between 60 - 90 days, the frequency of visits is 4 times in one period, where the patient's bill value is between IDR 6,000,000.00 - IDR 9,000,000.00 and the distance between visits is 60 - 90 days. The patient category is very high if the patient's last visit was less than 60 days,

TABLE 6 Outpatient RFMT scoring rules

Category	Score	R (Day)	F	M (million)	T (Day)
Very low	1	> 180	1	<0.007	0 or > 170
Low	2	90 – 180	2-5	0.007-0.1	90-170
Medium	3	30-90	5-10	0.1-0.4	930-90
High	4	15-30	10-15	0.4-1.2	15.30
Very High	5	< 15	> 15	> 1.2	< 15

the frequency of visits is more than 4 times in one period, where the value of the patient's bill is more than IDR 9,000,000.00, and the distance between visits is less than 60 days.

Based on Table 6 , it shows that the patient category is very low if the patient's last visit was more than 180 days, the frequency of visits is only 1 time in one period, where the patient's bill value is less than IDR 7,000.0,0. The distance between visits is zero or greater than 170 days. The patient category is low if the patient's last visit was between 90 - 180 days, the frequency of visits is 2 - 5 times in one period, where the value of the patient's bill is between IDR 7,000.00 - IDR 100,000.00, and the distance between visits is 90 - 170 days. The patient category is medium if the patient's last visit was between 30 - 90 days, the frequency of visits is 5 - 10 times in one period, where the patient's bill value is between IDR 100,000.00 - IDR 400,000.00, and the distance between visits is 30 - 90 days. The patient category is high if the patient's last visit was between 15 - 30 days, the frequency of visits is 10 - 15 in one period, where the value of the patient's bill is between IDR 400,000.00 - IDR 1,200,000.00 and the distance between visits is 15 - 30 days. The patient category is very high if the patient's last visit was less than 15 days, the frequency of visits is more than 15 times in one period, where the value of the patient's bill is more than IDR 1,200,000.0,0, and the distance between visits is less than 15 days.

3.3 | Clustering

Next, the data clustering process was carried out using the DBSCAN and KMeans algorithms on a dataset with RFMT quintile scores. In the clustering experiment, we used different epsilon (Eps) and minimum number of samples (MinPts) parameter values in DBSCAN and K parameter values in K-Means. In this study, the Silhouette Index (SI) was used to evaluate the quality of the clustering results.

3.3.1 | Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

The algorithm can find high-density core samples and expand clusters from them. Two main parameters of the algorithm determine clusters: MinPts and Eps. The MinPts parameter determines the minimum number of points that can be classified as a core sample, and the Eps parameter defines the algorithm's noise tolerance level (Pietrzykowski, 2020).

1. Determine the values of the MinPts and Eps parameters.

2. Randomly determine the p-value or starting point.

3. Calculate Eps or all distances from points whose density is reachable to p using the Euclidean distance formula using equation 2.

$$d_{ij} = \sqrt{\sum (i\alpha - \times j\alpha)\rho 2\alpha} \tag{2}$$

4. Where $\times i\alpha$ is the variable of α from the object i (i =1, …, n; a=1, …, p) and dij are euclidean distance values .

5. A cluster is formed when the points that meet Eps are more than MinPts, and point p is the core point.

6. Repeat steps 3 – 4 until the process is carried out at all points. If p is a border point and no points whose density is reachable concerning p, then the process continues to another point.

3.3.2 | **K-Means**

K-Means is a data grouping method that can group existing data into two or more groups. This method groups data so that data with similar characteristics will be included in the same group, and data with different characteristics will be grouped into different groups (Wang et al., 2019). According to Tan (2005), the stages of how the algorithm works K-Means are as follows:

1. Determine the k value to determine the number of clusters.

2. Determine the initial value of the centroid or cluster center point randomly, and the next stage uses equation 3.

$$V_{ij} = \frac{1}{n} \sum_{k=0}^{N_i} X_{kj} \tag{3}$$

With:
V_{ij} = cluster centroid of i for the variable j
N_i = the amount of data in the i cluster
i, k = index of the cluster
j = index of the variable
$X_{k,j}$ = the data value of k in the cluster for the variable j

3. Calculating the distance between the centroid point and the point on each object can be done using Euclidean Distance with equation 4.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_t - t_t)^2} \tag{4}$$

With:
D_e = Exclusion Distance
$_i = amount of data$
$(x, y) = data coordinate$
$(s, t) = centroid coordinates$

4. Group the data to form clusters with the centroid point of each cluster being the closest centroid point, taking into account the minimum distance to the object.

5. Update the centroid value on each cluster.

6. Repeat steps 2 to 5 until the centroid point value does not change.

3.3.3 | **Silhoutte Index (SI)**

One method for measuring the optimality of clustering results is the Silhouette Index method. The Silhouette Index method is a measure of validation based on internal criteria (Nahdliyah et al., 2019). This cluster validity method is based on internal criteria, where evaluation of object placement in each cluster is carried out by comparing the average distance of objects both in the same cluster and different clusters (Wang et al., 2019). To calculate the Silhouette Index value, you can use equation 5.

$$S = \frac{1}{n} \sum_{k=0}^{N_i} \left(\frac{b(i) - a(i)}{\max(a(i), b(i))} \right) \tag{5}$$

With:
a(i) = the average distance of sample i to other samples in the cluster.
b(i) = represents the minimum sample distance from sample i to another cluster.

TABLE 7 The RFMT weight using Fuzzy AHP technique,

Service	WR	WF	WM	WT
Outpatient	0.28	0.33	0.13	0.26
Inpatient	0.15	0.32	0.39	0.14

3.3.4 | Min-Max Normalization

After forming clusters on the inpatient and outpatient datasets, the data normalization process is carried out on the Recency (R), Frequency (F), Monetary (M), and Interpurchase attributes. Time (T) uses the Min-Max method for the CLV value calculation process. Min-Max Normalization is a data processing technique used to change the value of a variable within a certain range. Min-Max will adjust to the specified limits by correlating with the original data (Patro et al., 2015). This normalization technique transforms numerical attributes on a smaller scale, such as 0 to 1 (Junaedi et al., 2011). This scale value shows that the lowest value limit is zero, and the highest is 1. Min-Max normalization can be calculated using equation 6. (Khajvand et al., 2011).

$$X' = \frac{x - \min_a}{\max_a - \min_a} ((\text{newmax} - \text{newmin}) + \text{newmin}) \tag{6}$$

With:

X' = Normalized value

x = Original value that will be processed for normalization

$\min_a$ = Lowest value for each variable

$\max_a$ = Highest value for each variable

newmin = Maximum range of X which has a value of 0

newmax = Maximum range of X which has a value of 1

3.4 | Segmentation Based on CLV Analysis

The RFMT-based CLV method provides a reliable basis for measuring customer lifetime value and understanding market segmentation with various recency, frequency, and monetary values. There are many assessment methods for these three variables. In the first study, Arthur Hughes (the founder of the RFMT model) considered the RFMT scoring method, which was used to separate customers into five equal groups. According to him, these three variables are given the same weight to calculate the composite score (Hughes, 1994). Meanwhile, Stone (1994) states that different businesses have different characteristics and assign different variable weights to the RFMT measure according to the nature of the business. For example, he proposed a sequence of F, R, and M to analyze the value offered by customers who have used credit cards because frequency of use is more important in such business cases. The research application of CLV uses RFM T values normalized using the min-max method. RFM T is weighted based on relevant company-level assessments through a hierarchical analysis process (AHP) using the Fuzzy AHP technique. Fuzzy AHP is an analysis method developed from traditional AHP. Although AHP is commonly used to handle qualitative and quantitative criteria in MCDM, fuzzy AHP is considered better at describing vague decisions than traditional AHP (Kong, F, 2015). The research results for calculating RFMT weights using the Fuzzy AHP technique are shown in Table 7 .

Table 7 shows the Fuzzy-AHP calculation results showing different RFMT weights for outpatients and inpatients. Outpatients have the highest weights in the F criteria (0.33), R criteria (0.28), T criteria (0.26), and M criteria (0.13). Meanwhile, inpatients showed the highest weight in criteria M (0.39), criteria F (0.32), criteria R (0.15), and criteria T (0.14). The weighted RFMT method is used to calculate the CLV for each cluster. The CLV value of each cluster can be calculated using equation 7.

$$\text{CLV}_{c_i} = \text{NR}_{c_i} \times \text{WR}_{c_i} + \text{NF}_{c_i} \times \text{WF}_{c_i} + \text{NM}_{c_i} \times \text{WM}_{c_i} + \text{NT}_{c_i} \times \text{WT}_{c_i} \tag{7}$$

With =

NR_{c_i} = Normalized Recency Value of cluster c_i

WR_{c_i} = Weighted Recency

NF_{c_i} = Normalized Frequency Value of cluster c_i

TABLE 8 The patient segmentation on CLV value.

Segment	Patient Segmentation	CLV value
1	Champions	>= 0.80
2	Loyal Customers	0.60 – 0.79
3	Potential Loyalists	0.40 – 0.59
4	Lost Customers	< 0.40

WF_{c_i} = Weighted Frequency NM_{c_i} = Normalized Monetary Value of cluster c_i
WM_{c_i} = Weighted Monetary
NT_{c_i} = Normalized Interpurchase Time value from cluster c_i
WT_{c_i} = Weighted Interpurchase Time

Sequentially, the level of customer segmentation based on the CLV index calculation is Champions, Loyal Customers, Potential Loyalists, and Lost Customers[7]. Based on the results of the Focus Group Discussion (FGD) with Hospital Management, it was agreed that the segmentation level for the CLV value was as shown in the following Table 8 .

Based on Table 8 , it shows that the patient segmentation level is included in the Champions category if the CLV value is >= 0.80. The Loyal Customers category includes the patient segmentation level if the CLV value is between 0.60 – 0.79. The Potential Loyalist category includes the patient segmentation level if the CLV value is between 0.40 – 0.59. The patient segmentation level is included in the Lost Customers category if the CLV value is less than 0.40.

4 | RESULTS AND DISCUSSION

Understanding the characteristics of each cluster group based on clustering can help hospital management implement more specific marketing strategies in the outpatient and inpatient segments to achieve better patient retention goals. The distribution of RFMT on a discrete scale can be seen in Figure 2 and Figure 3 .

Figure 2 shows the distribution of outpatients with the highest R score of 15,000 patients who fall into category 1 (very low). The highest F score of 25,000 patients was in category 1 (very low). The highest M score was 18,000 patients who fell into category 3 (medium). The highest T scores of 19,000 patients were in category 1 (very low). The RFMT variable shows an uneven distribution. Patients with the same scores had similar outpatient or inpatient visit behavior.

4.1 | Clustering Inpatient RFMT Dataset

Based on the results of the Elbow method for the inpatient dataset in Figure 4 , it shows that the best eps parameter value for the inpatient dataset is 0.9.

Based on the results of the Elbow method for the inpatient dataset in Figure 5 , it can be seen that the best k parameter value for the inpatient dataset is 10. Results Testing the quality of the RFMT dataset clustering results for inpatients using the DBSCAN and K-Means algorithms using Silhouette Index (SI) values has been carried out. Clustering uses the K-Means algorithm, which has the best SI value of 0.58 and an optimal k value of 10. Clustering uses the DBSCAN algorithm, which has the best SI value of 0.99, a value of 0.8, and a MinPts value of 8.

4.2 | Clustering the Outpatient RFMT Dataset

Based on the results of the Elbow method for the outpatient dataset in Figure 6 , it shows that the best eps parameter value for the outpatient dataset is 0.9.

Based on the results of the Elbow method for the outpatient dataset in Figure 7, it can be seen that the best k-parameter value for the outpatient dataset is 7. The results of testing the quality of the outpatient RFMT clustering dataset using the DBSCAN and K-Means algorithms using Silhouette Index (SI) values have been carried out. Clustering uses the K-Means algorithm, which has the best SI value of 0.47 and an optimal k value of 7. Clustering uses the DBSCAN algorithm, which has the best SI value of 0.91 with an Eps value of 0.9 and a MinPts value of 75.
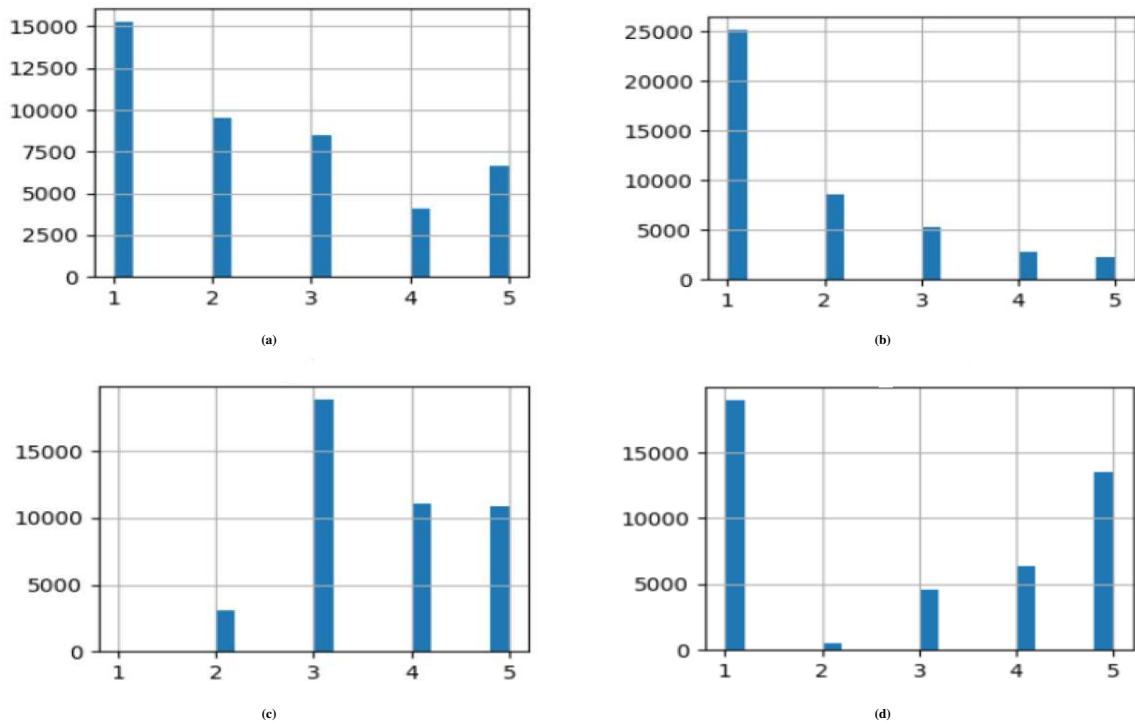
FIGURE 2 The distribution of outpatient RFMT: (a) r_{rank}, (b) f_{rank}, (c) m_{rank} and, (d) t_{rank}.

TABLE 9 The RFMT value based on segmentation inpatients.

Segmentation	RFMT value			
	Value	Mean	Min	Max
Champions	R	0.67	0.00	1.00
	F	0.78	0.50	1.00
	M	1.00	1.00	1.00
	T	0.94	0.75	1.00
Loyal Customers	R	0.62	0.00	1.00
	F	0.37	0.25	0.75
	M	0.91	0.75	1.00
	T	0.81	0.25	1.00
Potential Loyalists	R	0.58	0.00	1.00
	F	0.20	0.00	0.25
	M	0.68	0.50	1.00
	T	0.58	0.00	1.00
Lost Customers	R	0.41	0.00	1.00
	F	0.00	0.00	0.00
	M	0.50	0.25	1.00
	T	0.00	0.00	0.00

4.3 | Patient Segmentation

After forming clusters on the inpatient and outpatient datasets, the data normalization process was carried out on the Recency (R), Frequency (F), Monetary (M), and Interpurchase Time (T) attributes using the theMin-Maxx method and the CLV value was calculated based on The weighted RFMT method uses the Fuzzy AHP technique. Patients are divided into four categories for patient segmentation based on each CLV value cluster: 1. Champions, 2. Loyal Customers, 3. Potential Loyalists, and 4. Lost Customers. RFMT values based on inpatient segmentation are presented in Table 9 and Table 10 below.
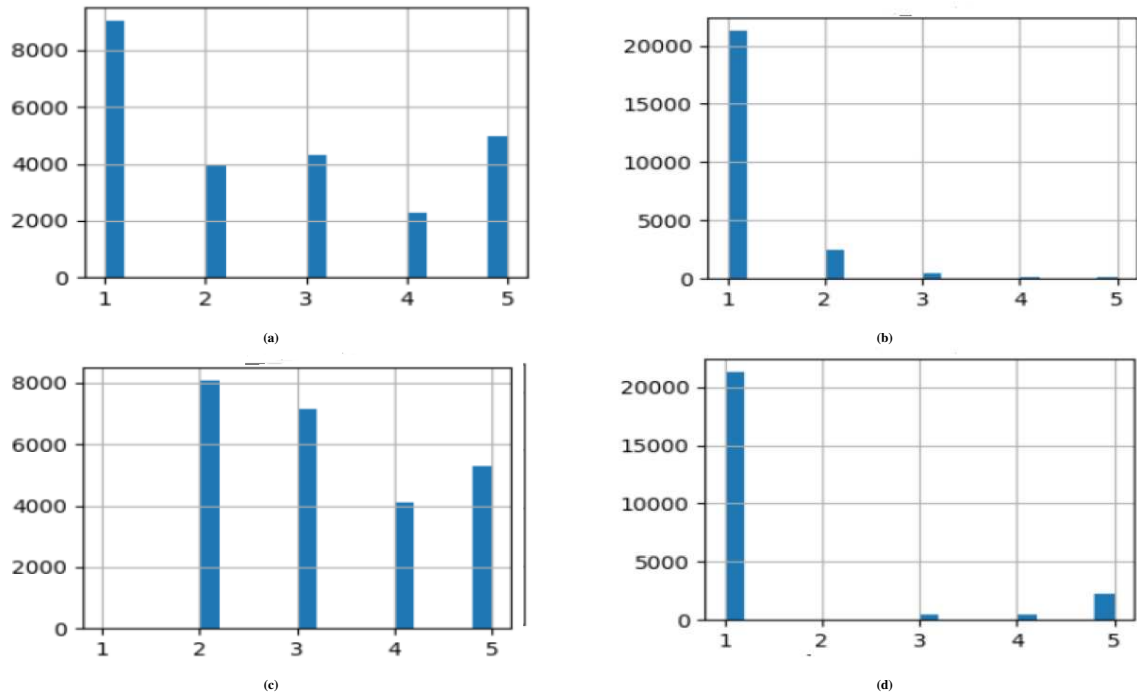
Based on Table 9 above, it can be seen that:

FIGURE 3 The distribution of inpatient RFMT (a) r_{rank}, (b) f_{rank}, (c) m_{rank} and, (d) t_{rank}.
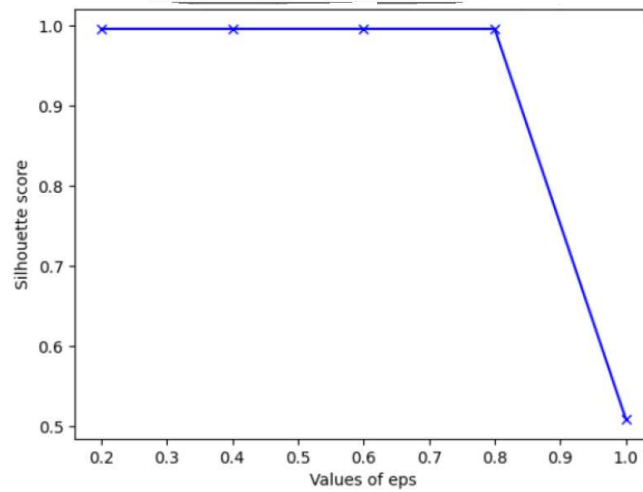

FIGURE 4 Silhoutte Index for finding optimal eps and minPts values for the DBSCAN algorithm

1. In the champions segment, the average transaction value for inpatients is more than IDR 9,000,000, where the last visit to the hospital was 90 days ago, the frequency of visits is 4 times in one period, and the distance between visits is 60-90 days.
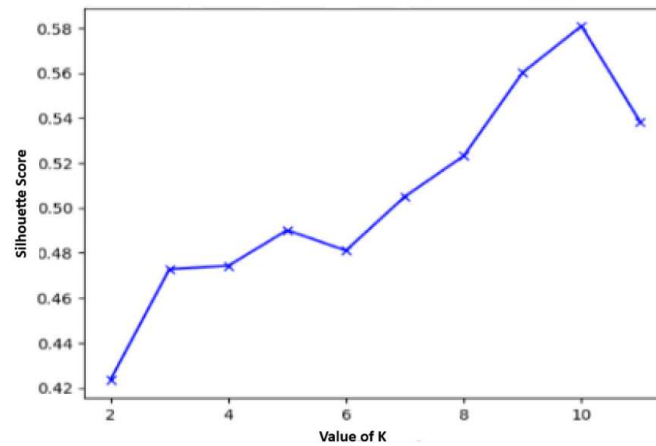
2. In the loyal customer segment, the average transaction value for inpatients is more than IDR 6,000,000, where the last visit to the hospital was 90 days ago, the frequency of visits is 3 times in one period, and the distance between visits is 60-90 days.

3. In the potential loyalist segment, the average transaction value for inpatients is IDR 3,000,000 – IDR 6,000,000, where the last visit to the hospital was 90 days ago, the frequency of visits is 1 time in one period and the distance between visits is 90- 150 days.

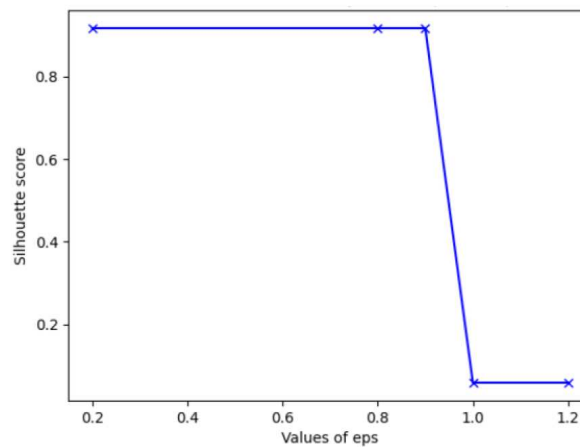**FIGURE 5** Silhoutte Index for finding the most optimal k value for the K-Means algorithm

FIGURE 6 Silhoutte Index for finding optimal eps and minPts values for the DBSCAN algorithm

4. In the lost customers, the average transaction value of inpatients is more than IDR 3,000,000 – IDR 6,000,000, where the last visit to the hospital was 150-210 days ago, the frequency of visits is 1 time in one period or the distance between visits are more than 210 days.

Based on Table 10   above, it can be seen that:

1. the champions segment, the average outpatient transaction value is IDR 1,000,000, where the last visit to the hospital was 15-30 days ago, the frequency of visits is 10-15 times in one period, and the distance between visits is 15-30 days.

2. In the loyal customer segment, the average outpatient transaction value is IDR 1,000,000, where the last visit to the hospital was 30-90 days ago, the frequency of visits is 5-10 times in one period,d and the distance between visits is 15- 30 days.

3. In the potential loyalist segment, the average outpatient transaction value is IDR 400,000-IDR 1,000,000, where the last visit to the hospital was 90-180 days ago, the frequency of visits is 2-5 times in one period, and the distance between visits is 15-30 days.

4. In the lost customer's segment, the average outpatient transaction value is IDR 400,000-IDR 1,000,000, where the last visit to the hospital was 90-180 days ago, the frequency of visits is 2-5 times in one period, or the distance between visits is more than 15-30 days.
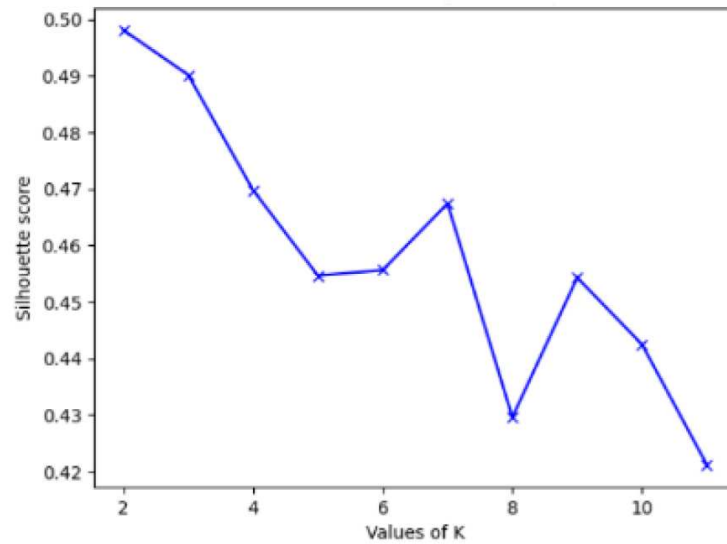
FIGURE 7 Silhoutte Index for finding the most optimal k value for the K-Means algorithm

TABLE 10 The RFMT value based on segmentation outpatient patients.

Segmentation	RFMT value			
	Value	Mean	Min	Max
Champions	R	0.87	0.50	1
	F	0.82	0.50	1.00
	M	0.97	0.75	1.00
	T	0.92	0.75	1.00
Loyal Customers	R	0.63	0.00	1
	F	0.50	0.00	1.00
	M	0.93	0.50	1.00
	T	0.87	0.50	1.00
Potential Loyalists	R	0.37	0.00	1.00
	F	0.27	0.00	0.50
	M	0.77	0.50	1.00
	T	0.78	0.50	1.00
Lost Customers	R	0.37	0.00	1
	F	0.27	0.00	0.50
	M	0.77	0.50	1.00
	T	0.78	0.50	1.00

TABLE 11 Segmentation of each inpatient cluster

Segmentation	CLV value			
	Count	Mean	Min	Max
Champions	14	0.87	0.80	1.00
Loyal Customers	31	0.68	0.61	0.77
Potential Loyalists	31	0.50	0.41	0.59
Lost Customers	14	0.26	0.1	0.39

The research results show that there are four segments of inpatients, namely:

1. Champions category, there were 473 (2%) patients who were members of 14 clusters (15.56%)

2. Loyal Customers category is 1,727 (7%) patients who are members of 31 clusters (34.44%)

3. Potential Loyalist category is 3,874 (16%) patients who are members of 31 clusters (34.44%)

4. Category Lost Customers were 18,516 (75%) patients who were members of 14 clusters (15.56%)

TABLE 12 The segmentation of each outpatient cluster.

Segmentation	CLV value		
	Count	Mean	Min	Max
Champions	10	0.88	0.80	1
Loyal Customers	24	0.69	0.60	0.79
Potential Loyalists	34	0.50	0.40	0.59
Lost Customers	31	0.26	0.03	0.39

The research results show that there are four segments of outpatients, namely:

1. Champions category there were 3,512 (8%) patients who were members of 10 clusters (10.10%)

2. Loyal Customers category is 5,661 (13%) patients who are members of 24 clusters (24.24%)

3. Potential Loyalist category is 11,070 (25%) patients who are members of 34 clusters (34.34%)

4. Lost Customers category is 23,678 (54%) patients who are members of 31 clusters (31.31%)

5 | CONCLUSION

This study used RFMT values, the min-max, and weighted RFMT methods to determine CLV. The results showed that outpatient and inpatient segmentation were categorized into patient segment segments on similar behavior according to the RFMeach patient's RFMT value model, which can characterize patient visit behavior by investigating the elapsed time. The RFMT model does not focus on attracting new patients but on identifying the best patients to improve marketing performance.

This research shows four segmentations of outpatients and four inpatients based on CLV values: Champions, Loyal Customers, Potential Loyalists, and Lost Customers. Inpatient segmentation includes: Champions Category, 473 (2%) patients belonging to 14 clusters (15.56%). In the Loyal Customers category, 1,727 (7%) patients were members of 31 clusters (34.44%). In the Potential Loyalist category, there were 3,874 (16%) patients in 31 clusters (34.44%). In the Lost Customers category, 18,516 (75%) patients were members of 14 clusters (15.56%). Outpatient segmentation includes: Champions Category, 3,512 (8%) patients belonging to 10 clusters (10.10%). In the Loyal Customers category, 5,661 (13%) patients are members of 24 clusters (24.24%). In the Potential Loyalist category, 11,070 (25%) patients were members of 34 clusters (34.34%). In the Lost Customers category, 23,678 (54%) patients were in 31 clusters (31.31%).

Segmentation of inpatients and outpatients is important in hospital management and mapping marketing strategies. Based on CLV analysis using the RFMT model developed in this research, management can use the segmentation of outpatients and inpatients to manage better patient services. Future research could develop a prediction model for patient visits using machine learning methods.

CREDIT

References

1. Benabdellah AC, Benghabrit A, Bouhaddou I. A survey of clustering algorithms for an industrial context. Procedia Computer Science 2019;148:291–302. http://dx.doi.org/10.1016/j.procs.2019.01.022.

2. Aditama TY. Manajemen Administrasi Rumash Sakit (Hospital Administration Management). 2 ed. Jakarta, Indonesia: UI Press; 2006.

3. Utama A G dST S Nair R K. Influences of Advertisement on Buying Behaviour. Finance in Asia Pasific 2021;10:100–110. https://api.semanticscholar.org/CorpusID:169600687.

4. Elrod C, Stanley S, Cudney E, Fisher C. Empirical Study Utilizing QFD to Develop an International Marketing Strategy. Sustainability 2015;7(8):10756–10769. http://dx.doi.org/10.3390/su70810756.

5. Patro SGK, Sahu KK. Normalization: A Preprocessing Stage. International Advanced Research Journal in Science Engineering and Technology 2015;2(3):20. http://www.iarjset.com/upload/2015/march-15/IARJSET%205.pdf.

6. Greenberg P. CRM at the speed of light: Capturing and keeping customers in Internet real time. Journal of Service Research 2001;9(3):139–150.

7. Hiziroglu A, Sengul S. Investigating Two Customer Lifetime Value Models from Segmentation Perspective. Procedia Social and Behavioral Sciences 2012 10;62:766–774. http://dx.doi.org/10.1016/j.sbspro.2012.09.129.

8. Junaedi H, Budianto H, Maryati I, Melani Y. Data Transformation Pada Data Mining. Ideatech 2011;p. 93–99.

9. Hu YH, Yeh TW. Discovering valuable frequent patterns based on RFM analysis without customer identification information. Knowledge Based Systems 2014 5;61:76–88. http://dx.doi.org/10.1016/j.knosys.2014.02.009.

10. Karima FA, Shiddiqi AM. Optimization of Access Point Positioning on Wi-Fi Networks Using the K-Means Clustering Method. IPTEK The Journal for Technology and Science 2022 5;33(1):13. http://dx.doi.org/10.12962/j20882033.v33i1.12402.

11. Fader PS, Hardie BGS, Lee KL. RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research 2005 11;42(4):415–430. http://dx.doi.org/10.1509/jmkr.2005.42.4.415.

12. Malav A, Kadam K, Kamat P. Prediction of Heart Disease Using K-Means and Artificial Neural Network as Hybrid Approach to Improve Accuracy. International Journal of Engineering and Technology 2017 8;9(4):3081–3085. http://dx.doi.org/10.21817/ijet/2017/v9i4/170904101.

13. Anitha P, Patil MM. RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University Computer and Information Sciences 2022;34(5):1785–1792. http://dx.doi.org/10.1016/j.jksuci.2019.12.011.

14. Tukel OI, Dixit A. Application of customer lifetime value model in make-to-order manufacturing. Journal of Business amp and Industrial Marketing 2013 7;28(6):468–474. http://dx.doi.org/10.1108/jbim-04-2013-0099.

15. Gupta S, Hanssens D, Hardie B, Kahn W, Kumar V, Lin N, et al. Modeling Customer Lifetime Value. Journal of Service Research 2006 11;9(2):139–155. http://dx.doi.org/10.1177/1094670506293810.

16. Chang HC, Tsai HP. Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Systems with Applications 2011;38(12):14570–14580.

17. Chang HC, Tsai HP. Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Systems with Applications 2011;38(12):14499–14513. http://dx.doi.org/10.1016/j.eswa.2011.05.034.

18. Rungruang C, Riyapan P, Intarasit A, Chuarkham K, Muangprathub J. RFM model customer segmentation based on hierarchical approach using FCA. Expert Systems with Applications 2024 3;237:121449. http://dx.doi.org/10.1016/j.eswa.2023.121449.

19. Christy J, Umamakeswari A, Priyatharsini L, Neyaa A. RFM ranking – An effective approach to customer segmentation. Journal of King Saud University Computer and Information Sciences 2021;33(10):1241–1251.

20. Zhou J, Wei J, Xu B. Customer segmentation by web content mining. Journal of Retailing and Consumer Services 2021;61:102588. http://dx.doi.org/10.1016/j.jretconser.2021.102588.

21. Yeh IC, Yang KJ, Ting TM. Knowledge discovery on RFM model using Bernoulli sequence. Expert Systems with Applications 2009;36(3):5866–5871. http://dx.doi.org/10.1016/j.eswa.2008.07.018.

22. Yoseph F, Heikkila M. Segmenting Retail Customers With An Enhanced RFM And A Hybrid Regression/Clustering Method. 2018 International Conference on Machine Learning and Data Engineering (iCMLDE) 2018;p. 108–116. http://dx.doi.org/10.1109/icmlde.2018.00029.

23. Morrison DG. Interpurchase Time and Brand Loyalty. Journal of Marketing Research 1966 8;3(3):289–291. http://dx.doi.org/10.1177/002224376600300309.

24. Vakratsas D, Bass FM. The relationship between purchase regularity and propensity to accelerate. Journal of Retailing 2002 6;78(2):119–129. http://dx.doi.org/10.1016/s0022-4359(02)00068-4.

25. Meyer-Waarden L. The influence of loyalty program membership on customer purchase behavior. European Journal of Marketing 2008;42(1/2):87–114.

26. Guo RS. A multi-category inter purchase time model based on hierarchical Bayesian theory. Expert Systems with Applications 2009 4;36(3):6301–6308. http://dx.doi.org/10.1016/j.eswa.2008.08.059.

27. Guo J, Gao Z, Liu N, Wu Y. Recommend products with consideration of multi-category inter-purchase time and price. Future Generation Computer Systems 2018 1;78:451–461. http://dx.doi.org/10.1016/j.future.2017.02.031.

28. Hood N, Clarke G, Clarke M. Segmenting the growing UK convenience store market for retail location planning. The International Review of Retail Distribution and Consumer Research 2015 11;26(2):113–136. http://dx.doi.org/10.1080/09593969.2015.1086403.

29. Khajvand M, Tarokh MJ. Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. Procedia Computer Science 2011;3:1327–1332. http://dx.doi.org/10.1016/j.procs.2011.01.011.

30. Kong F, Liu H. Applying fuzzy analytic hierarchy process to evaluate success factors of E-commerce. International Journal of Information and Systems Sciences 2005 4;1(3-4):406—412.

31. Maryani I, Riana D. Clustering and profiling of customers using RFM for customer relationship management recommendations. In: The 5th International Conference on Cyber and IT Service Management (CITSM) Denpasar, Indonesia: IEEE; 2017. p. 1–6. http://dx.doi.org/10.1109/citsm.2017.8089258.

32. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Data Mining 2006;p. 83–124. http://dx.doi.org/10.1016/b978-0-12-381479-1.00003-4.

33. Jing W, Zhao C, Jiang C. An improvement method of DBSCAN algorithm on cloud computing. Procedia Computer Science 2019;147:596–604. http://dx.doi.org/10.1016/j.procs.2019.01.208.

34. Pietrzykowski M. Comparison of mini-models based on various clustering algorithms. Procedia Computer Science 2020;176:3563–3570. http://dx.doi.org/10.1016/j.procs.2020.09.030.

35. Wang X, Xu Y. An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. IOP Conference Series and Materials Science and Engineering 2019;569(5):52024. http://dx.doi.org/10.1088/1757-899x/569/5/052024.

36. Nahdliyah MA, Widiharih T, Prahutama A. Metode K-Medoids Clustering Dengan Validasi Silhouette Index dan C-Index: Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018) (K-Medoids Clustering Method Using Silhouette Index and C-Index Validation: Case Study of The Regency/City Crimes Rate in Central Java in 2018). Jurnal Gaussian 2019 5;8(2):161–170. https://ejournal3.undip.ac.id/index.php/gaussian/article/view/26640.

37. Metisen BM, Sari HL. Analisis Clustering Menggunakan Metode K Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila (Clustering Analysis Using the K Means Method for Grouping Product Sales in Fadhila Supermarket). Jurnal Media Infotama 2015;11(2):110–118. https://jurnal.unived.ac.id/index.php/jmi/article/view/258.

38. Miglautsch JR. Thoughts on RFM scoring. Journal of Database Marketing amp and Customer Strategy Management 2000 8;8(1):67–72. http://dx.doi.org/10.1057/palgrave.jdm.3240019.

39. Adha R, Nurhaliza N, Sholeha U, Mustakim M. Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia (Comparison of DBSCAN and K-Means Clustering Algorithms for Grouping Covid-19 Cases in the World). SITEKIN: Jurnal Sains and Teknologi & Industri 2021;18(2):206–211. https://ejournal. uin-suska.ac.id/index.php/sitekin/article/view/12469.

40. Mustakim, Indah RNG, Novita R, Kharisma OB, Vebrianto R, Sanjaya S, et al. DBSCAN algorithm: Twitter text clustering of trend topic Pilkada Pekanbaru. Journal of Physics and Conference Series 2019 11;1363(1):12001. http://dx.doi.org/10. 1088/1742-6596/1363/1/012001.