Jurnal Sosial Humaniora (JSH)
2020, special edition
ISSN Online: 2443-3527
ISSN Print: 1979-5521

Segmentation Analysis of Students in X Course with RFM Model and Clustering

Yanuar Rafi Rahadian[1], Bambang Syairudin[2]
Faculty of Industrial Technology and System, Institut Teknologi Sepuluh Nopember, Surabaya, 60111
rafirahadian11a@gmail.com, bambangsyairudin@gmail.com

Diterima: 14/04/2020.
Direview: 21/04/2020.
Diterbitkan: 31/08/2020.

Hak Cipta © 2020 oleh Penulis (dkk) dan Jurnal Sosial Humaniora (JSH)
*This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Subject Area: Marketing

Abstract

In the business world, the competition to maintain and obtain more customers has become tougher. The presence of new players entering the market is driven by the developments of internet and advertisement. The X guitar course is an institution engaged in the field of non-formal education services. The customers are the course student that has made the payment transaction. The map of customer segmentation is one of the most important components in finding the main needs of each customer. Know the main needs of each customer is expected to increase the customer's loyalty. Customer segmentation can be done by using the clustering method through a data mining approach in the form of RFM (Recency, Frequency and Monetary) Models. Recency is the data of the last payment transaction date. Frequency shows the number of course payment transactions. Monetary comes from the nominal amount of the transaction. RFM data is combined with the Fuzzy Gustafson-Kessel and K-Means clustering method to produce output in the form of k-clusters of customer. The formed segment is expected to represent the need of customers that vary by using validation process with the Global Silhouette Index. The customer population of the course is 225 students. It has been concluded that the RFM score for each subject by using 3 FGK clusters is the optimum cluster model with the largest Silhouette Index, which is 0.523. This research is expected to provide an in-depth analysis of customer segmentation for X guitar course.

Keywords: *Clustering; customer segmentation; X Course; RFM model; silhouette index.*

Introduction

The competition in the business world mainly focused on maintain and obtain more customers. Every business player is required to be able to observe the customer's changing needs. The tight competition can be seen from the increasing number of new players entering certain Industries, which has been driven by the development of internet technology. The development of the internet in Indonesia encourages a variety of innovations in various lines of social life, including business activities. The internet is a growing technology and is able to develop new business relationships and market opportunities. (Pratminingsih, et al, 2013) states that the internet is a useful tool for gathering information on customers, competitors and potential markets and it can inform about various products and services (S.Pratiminingsih et all, 2013).

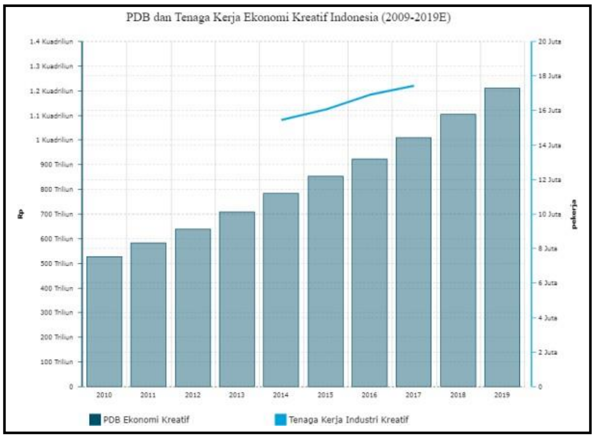


Figure 1. Representation of Increased GDP due to the creative economy

Non-formal education services such as guitar courses are one of the SME (Small Medium Enterprise) business activities which are marketed via the internet. SMEs drive the economies of both developed and developing countries (A.Syukriah and I Hamdani, 2015). The Indonesian Ministry of Cooperatives and Small and Medium Enterprises (*Kementerian Koperasi dan Usaha Kecil Menengah*) and the Indonesian Central Agency on Statistics (*Badan Pusat Statistik*) state that SMEs are businesses with a net worth of at most IDR 200,000,000 excluding land and buildings for business premises (Rahmana A, 2008).

The creative economy has quite a promising potential to support the national economy (B.E Kreatif, 2017). The economic Gross Domestic Product (GDP) that emerged from creative ideas reached IDR 1,009 trillion in 2017, an increase from IDR 922.59 trillion in the previous year. The number of workers involved in the creative economy in 2016 reached 16.91 million workers and increased to 17.43 million workers. Until the end of 2018, the contribution of the creative economy to national GDP is estimated to reach IDR 1,105 trillion and will again increase to IDR 1,211 trillion by 2019. Technology reflects the potential of the digital economy such as e-commerce, online game services, food delivery services and digital video services that can drive the growth of the creative economy. The government targets the value of Indonesia's digital economy in 2020 to reach the USD 130 billion or equivalent to IDR 1,888 trillion. This value is equivalent to 11% of the national GDP. The creative industry in the music sector is one of the fast-growing MSME subsectors. This was reinforced by the increase in the contribution of the music subsector by 7.26% to the GDP of the creative economy in 2015. Non-formal education is an educational activity organized outside the formal education system. The course is an educational activity that takes place in the community carried out deliberately, organized, and systematically to provide one or a series of specific lessons to certain adults or teenagers in a relatively short time (B.E Kreatif, 2017). One form of non-formal education is music courses. Music is a work of sound art in the form of songs or musical compositions, which express the thoughts and feelings of the creator through the elements of music, namely rhythm, melody, harmony, song form/structure, and expression (D.Jamalus, 1988). As part of human life, music is studied in the existing social environment. Humans use and facilitate music as a situational factor for social development (V.J

Koneni, 1982). As part of the culture, the development of music is very dynamic and triggers market demand. A course institution is needed to accommodate market demand. Supporting data needed at the spatial level of the region can determine the sustainability of market demand, especially in a big city such as Surabaya.

Economic growth is one of the macro indicators to see the real economic performance in a region. The rate of economic growth is calculated based on changes in the GDP based on the constant price of the relevant year against the previous year. Economic growth can be seen as an increase in the number of goods and services produced by all business activities of economic activity in an area over a period of a year. From 17 existing economic business sectors, there are 4 samples of business sectors in Surabaya which is shown in Table 1. In Table 1, one of the business sectors experienced positive growth while the rest experienced contraction. The variance in Table 1 measures the range of GDP growth rate is in each business field. The education service sector has a growth variance of 0.05 and is the lowest when compared to sixteen other sectors. This indicates that the education service sector has more stable business sustainability.

The X course that was established on January 17th, 2017 is a business engaged in the creative economy in the non-formal education services sector. Business owners see the opportunity for many people in Surabaya who want to learn music, especially guitar, but is constrained by several factors. The main factor is the high price of guitar courses. The X course came to bridge the demand by offering more affordable course fees for Surabaya citizens. Students can take courses in tutoring places or call the teacher to come to their house. In addition, from January 1st, 2019, the X course opened another service division, namely the Research Consultation Service to assist students in conducting research both online and through face-to-face. At present, the X course is centered in Pucang Anom Timur, Gubeng District with two branches which are in Pondok Rosan, Wiyung District and Simorejo Sari, Sukomanunggal District, all in Surabaya city. As of October 31st, 2019, there were 225 students who had taken courses with 52 active students in October 2019. Every day there was a guitar learning class (excluding holidays on Tuesday), in which either the teacher came to the student's house or the students came to the tutoring place, with a turnover of around IDR 12 million per month.

There were business competitors before the X course was established in Surabaya. All competitors certainly have an influence on the development of this business. Until now, the X course has not yet implemented a specific business strategy to retain and satisfy students. Therefore, a retention strategy by maintaining student loyalty, which is to be carried out in this research, is expected to reduce the effects of competitors. The strategy of obtaining and retaining certain customers (students) is considered important to create added value for both the company and the customer (S.I Shim, W.S.Kwon and S.Forsythe, 2013).

Table 1. Comparison of Four business fields on surabaya's growth rate based on 2010 constant price (%)

Business Sector	2014	2015	2016	2017	2018	Variance
Agriculture & Marine	3.54	4.73	4.36	3.35	-1.40	4.92
Mining	3.20	3.98	3.14	2.58	-0.10	1.97
Electricity & Gas Procurement	-1.90	3.12	1.05	1.75	-0.07	2.73
Education Service	5.71	6.31	6.02	5.95	6.24	0.05

Business development will not work without customers. Maintaining customers is as important as finding customers. A decrease in the number of customers will reduce business flow with customers and turnover (M.Mohammadian and I Makhani, 2016). Each customer should be treated individually because each customer has different needs. However, a large number of customers make this approach is not possible. Customer segmentation is an alternative in treating individual customers (K. K. Tsiptsis dan A. Chorianopoulos, 2011).

The X course is one of the creative business fields that need to manage relationships with its customers. The owner cannot directly meet with all customers but can only meet with some customers and all teachers. Therefore, the teacher as a "distributor" plays an important role in helping the achievement of student targets. Not all students are registered students who often make course payments (transactions). The company has never analyzed student behavior in paying for courses, so no strategy has been launched in maintaining student loyalty. The relationship between the owner and the teacher also depends very much on subjective communication. There are adverse effects if business owners erroneously plan and implement student retention strategies, which includes losing students and transferring students to competitors.

The problems experienced by the X course in managing customer relationships can be solved by the segmentation process that is extracting student payment history data at a certain period. Students will be grouped into several segments which are distinguished based on student behavior in making course payments. This student behavior can be described through the RFM (Recency, Frequency, and Monetary) model. The method used in customer segmentation is the K-Means clustering and Fuzzy Gustafson-Kessel (FGK) clustering algorithm. There are three variables for clustering, which are according to R (Recency), F (Frequency), and M (Monetary) scores. Clustering is a method that can be used for grouping based on the similarity of the RFM variables. Customers with the same characteristics will be grouped in one segment whereas customers with the different characteristics will be grouped in the different segments (J. Ong, “Ong, J. O.2013). The formed segment is expected to represent varied consumer needs. The observation object in this research is the X guitar course service in Surabaya. The selection of these services is based on the analysis of Table 1 where the guitar course is one of the educational services.

Previous research regarding the application of customer segmentation, customer satisfaction and guitar course case study has been carried out. Wei, et al. (2016) aims to identify valuable customers and develop marketing strategies for animal hospitals (J. T. Wei, S. Y. Lin, Y. Z. Yang dan H. H. Wu, 2016). The research object customer who has a dog pet. This research uses variables from RFM analysis results and conducts K-Means clustering. Amalia, et al. (2016) conducted an analysis of clusters of banking sub-sector companies based on CAMELS Financial Ratios in 2014 using the Fuzzy K-Means and FGK clustering (N. A. Amalia, D. A. Widodo dan P. P. Oktaviana, 2016). Based on this research, FGK gives the best performance because it gives the smallest icd-rate value. Indah (2018) conducted research related to a music/guitar course in Indonesia (A. R. Indah, 2018). This research related to the description of the implementation and evaluation of the LKP Lily’s Music School Semarang.

This research proposes an analysis of segmentation and satisfaction of students in the X course by using the RFM model and clustering. The customer segmentation carried out in this research aims to be a robust method of representing various customer needs. This research is expected to contribute to the development of a more targeted and optimal CRM (Customer Relationship Management) system in the X course, therefore helping the X course in increasing nominal payment transactions by valuable customers and retaining customers who are making a large profit contribution to the course.

Method

Customer Relationship Management

Based on the SIPOC (supply, input, process, output, and customer) diagram, the customer is the party that uses the output of the process. Each customer has its own characteristics, and strategies are needed in managing relationships with customers. Customer relationship management, often known as CRM, is one of the strategic approaches in handling proper relationships with key customers and other customer groups. CRM provides a better opportunity for business industry players to use available information to find out the type of customer and create added value for each customer. There are three phases in managing customer relationships, which are (R. Kalakota dan M. Robinson, 2001):

1. *Acquire*, which is a phase of the company's strategy to get new customers. Generally, the acceptance of the best services and the company's superior products determines the number and frequency of new customers;
2. *Enhance*, is a phase of the company's strategy to increase profitability or increase profits from existing corporate customers. Business people can make efforts to establish long-term relationships with customers;
3. *Retain*, namely a phase of the company's strategy to retain customers who have high profitability. This phase focus on providing whatever the main customer wants, not based on market demand.

CRM supports a company to provide services to customers in real-time and build relationships with each customer through the use of information about customers (P. Kotler, K. L. Keller, M. Brady, M. Goodman dan T. Hansen, 2012). Companies can also find a picture of the desires and needs of customers so as to adjust the strategy in fulfilling the wants and needs of customers well. The following are four segments in categorizing the characteristics of conducting relationships with customers (D. Peppers dan M. Rogers, 2011):

- Most Valuable Customer (MVC), is the customer who has the highest value in business sustainability. This customer group has the biggest contribution in providing benefits to the company;
- Most Growable Customer (MGC), is a customer who has great potential to become an MVC in the future. Often business people are not aware of customers in this group;
- Below Zero (BZ), is the customer group that provides the least profit compared to other customer groups;
- Migrators, is a group of customers that needs to be further analyzed so that the actual category can be known. This customer group is located between BZ and BWC.

Increasing customer profitability growth is one of the main targets of CRM. This goal can be achieved if the company is able to continue improving the ability to know and understand customer behavior. Therefore, we need a CRM strategy that is able to achieve those goals, one of which is by using customer segmentation. The customer segmentation aims to match the potential of these customers with the undertaken services and marketing strategies so that the provided marketing strategies and services are effective (M. J. Berry dan G. S. Linoff, 2004).

Customer Segmentation

Customer segmentation is the process of distinguishing customer profiles and characteristics. The segment formation by using the help of data mining can enrich the segmentation results (K. K. Tsiptsis dan A. Chorianopoulos, 2011). The purpose of segmentation is to adjust products, services, and marketing messages for each segment (M. J. Berry dan G. S. Linoff, 2004). Customer segmentation is a preparatory step for classifying each customer according to the specified customer group (S. Jansen, 2001). The segmentation process places customers according to the characteristics of similar customer groups. Customer characteristic variables are as follows (M. J. Berry dan G. S. Linoff, 2004)

- Demography, including age, gender, number of family members, the extent of residence, income, profession, last education, homeownership status, social status (position or title), religion andcitizenship;
- Physiography, including personality type, lifestyle, moral values;
- Behavioral, including the principle of the benefit of the product/service sought, the purchase status, the level of use of the product/service, the frequency of purchase;
- Geography, including country, province, city, district, postal code, climate.

Different segmentation schemes can be developed according to the specific business goals of the organization. Segmentation is generally used through market data research to gain insight into customer attitudes, desires, views, preferences, and opinions about the company and competition (K. K. Tsiptsis dan A. Chorianopoulos, 2011). Customer segmentation based on market research and demographics often requires an understanding of the characteristics of all customers to more effectively know which segments are attracting customers. Data mining can be used to develop customer segmentation which also identifies the segmentation of customer behavior (M. J. Berry dan G. S. Linoff, 2004). In addition to external or market research data, transaction and customer payment data can also be used to gain insight into customer behavior. This way, the segmentation will allocate customers to form groups based on the amount of their expenses. This can be used to identify high-value customers and prioritize services (K. K. Tsiptsis dan A. Chorianopoulos, 2011). The company needs to know the customer profile. Customer profiles are very closely related to these customer segments (A. M. Scridon, 2008). Several strategies for analyzing customer profiles are as follows:

- RFM analysis, is one of the most commonly used types of customer profiling. RFM is a method used to segment based on the time of the customer’s last transaction, usually not taking into account the nominal of the transaction made;
- Demographic analysis, is very closely related to the geographical or location of the customer originates. However, in some demographic research, this can also be interpreted to segment according to age, gender, income, and marital status;
- Life stage analysis, is an analysis related to customer behavior. The behavior of each customer is certainly different, therefore it is very interesting to be understood by the business players.

The geographic differences in customer locations have become an important practical component of marketing strategies. This is largely due to organizational expansion goals which force managers to consider the increasingly complex delivery and advertisement system layout of new product launches and management (B. J. Bronnenberg dan P. Albuquerque, 2003). Researchers in the fields of marketing and economics have developed an interest in the spatial aspects of growth and market structure. The resulting research tradition has been called “the new economic geography”.

This flow of research began in 1970 in the field of industrial organizations. The X guitar course needs to treat their customers differently because of different customer locations. Therefore, the map will be formed based on the distribution of students in each segment so that it can be a reference for service providers in treating customers with different geographies.

This flow of research began in 1970 in the field of industrial organizations. The X guitar course needs to treat their customers differently because of different customer locations. Therefore, the map will be formed based on the distribution of students in each segment so that it can be a reference for service providers in treating customers with different geographies.

The process of studying customer behavior is one of the challenges of a company's marketing team (J. Blythe dan P. Megick, 2010). There are many ways to understand customer behavior, one of which is customer behavior in making transactions or the customer's time in conducting transactions. To help with this analysis, this study will display customer transaction habits per week. This will make it easier for the X guitar course in understanding the behavior of customer habit related to certain day/week/month where there are more or fewer customer making transactions. ***RFM (Recency, Frequency & Monetary) Model***

The object of observation in this study is the X guitar course service in Surabaya. The formed segment is expected to represent various consumer needs. The RFM scale attribute can be seen in Table 2.

Table 2. RFM scale attribute

Score	**R-Recency (Days)**	**F-Frequency**	**M-Monetary**
5	Very recent	Very frequent	Very high
4	Recent	Frequent	High
3	Standard	Normal	Normal
2	Not recent	Rare	Low
1	Long ago	Very rare	Very low

The RFM model was originally developed by (A. Hughes, 1994) and (J. R. Bult dan T. Wansbeek, 1995) to differentiate customer profitability based on several attributes, which are the length of time the customer is active, the frequency of customer payments and the nominal amount of money paid by customers. The following explanation of the attributes of the RFM model:

- *Recency of the last payment* (R), is an attribute that states a reviewer or the distance between the customer's last payment date and the current date. If the interval is closer to the current date, the R score gets higher;
- *Frequency of the payment* (F), this attribute shows how often customers make payments. For example, customer A makes payments four times a month more often than customer B who makes payments once a month. F scores will be higher if the frequency of payments is more frequent;
- *Monetary value of the payment* (M), is the nominal amount of money paid by the customer. The greater the nominal paid by the customer, the greater the M score.

There are certain characteristics of the RFM score. The greater the value of R and F indicates the tendency of customers to make repayment transactions (J. Wu dan Z. Lin, 2005). Whereas if M gets higher then there is a tendency for customers to give a good response to the products/services proposed by the company. In determining customer opportunities in responding to offers, RFM score calculation is required. There is a belief in the majority of companies that customers who have become new and most frequent buyers and have made large payments within a certain time frame are most likely to respond positively to the company's offerings in the future (R. J. Baran dan R. J. Galka, 2013). In the X course, the RFM score can be used to determine the suitability of certain customers to get a complete physical catalog offer or just delivery via online messenger's broadcast.

Based on the Great Dictionary of Indonesian Language (KBBI), visualization is the expression of an idea by using pictures, writing (words and numbers), maps, graphics and so on. After obtaining the RFM model, the visualization of the RFM model must be done so that the model can be better understood. The RStudio program is used to visualize the RFM model (W. N. Venables dan D. M. Smith, 2013).

Clustering Analysis

Cluster analysis is a technique for grouping data according to certain characteristics (R. A. Johnson dan D. W. Wichern, 2007). The result must have high homogeneity within a group and have high heterogeneity between groups. Cluster analysis will allocate a group of individuals to independent groups so that the individuals in the group are similar to each other, while the individuals in different groups are dissimilar (S. Sharma, 1996). This grouping is usually called a partition (B. Ruswandi, 2008). The similarity measurement that can be used is the euclidean and mahalanobis distances.

The clustering methods can be grouped based on its distance measurement technique. This distance-based method consists of hierarchical methods (agglomerative), which is including complete linkage, average linkage, and Ward method, and also the non-hierarchical method, which is including K-Means clustering (M. R. Anderberg, 2014). Hierarchical clustering and K-Means clustering only pay attention to the size of the distance between the objects of observation without considering other statistical aspects, such as the distribution of data or the objects on overlapping clusters.

The K-Means clustering is a very popular and common method. This method groups objects into *k* clusters and the division of clusters is based on differences in the average value of an object to the center of the cluster. The methodology for clustering by using K-Means algorithm is described in (R. A. Johnson dan D. W. Wichern, 2007). The K-Means algorithm has some weaknesses regarding the complexity of this algorithm in detecting 'natural' clusters, in which the clusters have different sizes, thicknesses or shapes that are not oval-shaped.

Fuzzy Gustafson-Kessel (FGK) clustering is the development of K-Means clustering and Fuzzy C-Means (B. Feil, B. Balasko dan J. Abonyi, 2007). The matrix-forming value in this grouping method is called the adaptive distance norm which is updated in each iteration. Therefore, this method is able to further adjust the geometric shapes of membership functions that are appropriate for the data set. Fuzzy cluster analysis considers the level of membership that includes the fuzzy set as a weighting basis. The main difference between FGK clustering and the hierarchical and non- hierarchical cluster methods is its ability to handle uncertainty. Mauliyadi, et al. (2003) stated that the accuracy of FGK clustering is higher than the K-Means clustering [34] and Amaliya, et al. (2013) obtained lower icd- rate from FGK clustering than Fuzzy C-Means clustering which means that the FGK clustering gives better performance (N. A. Amalia, D. A. Widodo dan P. P. Oktaviana,, 2016).

The algorithm of FGK clustering is as follows (B. Feil, B. Balasko dan J. Abonyi, 2007):

1. Enter N number of data that will be grouped.
2. Determine the number of clusters (L) where $1 < L < N$, the weighting exponent (m) where $m > 1$, the maximum number of iteration $(maksIter)$, and the target error (ε) where $\varepsilon > 0$.
3. Set the initial objective function as 0 and the iteration as 1.
4. Form the matrix U as the initial partition matrix.

$$U = \begin{bmatrix} u_{11}(f_1) & u_{c2}(f_2) & \cdots & u_{1N}(f_N) \\ u_{21}(f_1) & u_{c2}(f_2) & \cdots & u_{2N}(f_N) \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1}(f_1) & u_{c2}(f_2) & \cdots & u_{cN}(f_N) \end{bmatrix} \quad (1)$$

5. Calculate the center of each cluster k (v_k).

$$v_k = \frac{\sum_{j=1}^{N} (\mu_{ij})^m f_j}{\sum_{j=1}^{N} (\mu_{ij})^m}, k = 1,2,..., L \quad (2)$$

6. Calculate the covariance matrix of each cluster (F_k)

$$F_k = \frac{\sum_{j=1}^{N} (\mu_{ij})^m (f_j - v_k)^T A((f_j - v_k)}{\sum_{j=1}^{N} (\mu_{ij})^m} \quad (3)$$

7. Calculate the distance D_{ij} by using Eq. (4) where

$$A = [\det(F_k)^{\frac{1}{N}} F_k{}^{-1}].$$

$$D_{ij} = \left\| x_j - v_k \right\| = (f_j - v_k)^T A(f_j - v_k) \quad (4)$$

8. Calculate the objective function in Eq. (5) for iteration *t*.

$$J_{FGK}(X, U, V) = \sum_{k=1}^{L} \sum_{j=1}^{N} (\mu_{ij})^m D_{ij} \quad (5)$$

9. Calculate the new membership function U_{t+1} by using Eq. (6).

$$\mu_{ij} = \left[\sum_{k=1}^{L} \left(\frac{D(f_j, v_k)}{D(f_j, v_j)} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (6)$$

10. Compare the membership function in matrix U until convergence where $t > maks\ iter$ or $\|U_{t+1} - U_t\| < \varepsilon$. If $\|U_{t+1} - U_t\| \geq \varepsilon$ then repeat calculating the center of each cluster k (v_k).

Performance measurement of clustering results is a method to determine the validity of the clustering. One of the evaluation methods for measuring the clustering performance is the global silhouette, which formula is shown in Eq. (7) [36]. This method used to evaluate the quality of clusters produced from the clustering process. The validity of clustering can be seen by the level of optimization of a cluster and homogeneity among cluster members. The value of the silhouette ranges between -1 $\leq S \leq$ 1, where the results of clustering are good if the silhouette value is positive (0-1) [35]. This indicates that the data is in the right group.:

$$S = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{b_i - a_i}{\max\{a_i, b_i\}} \right) \quad (7)$$

Research Methodology

There are several steps of analysis used in this research, which is shown in Figure 2. In general, the flowchart diagram of this research is based on KDD (Knowledge Discovery on Database) or data mining. There are 3 steps in KDD, namely the preprocessing, the data mining and the postprocessing [10]. Data mining is discovering new information by looking for certain patterns or rules from a very large amount of data [19]. Data mining plays a role in the process of finding interesting and hidden patterns of a large data set stored in a database, data warehouse, or other data storage [19].

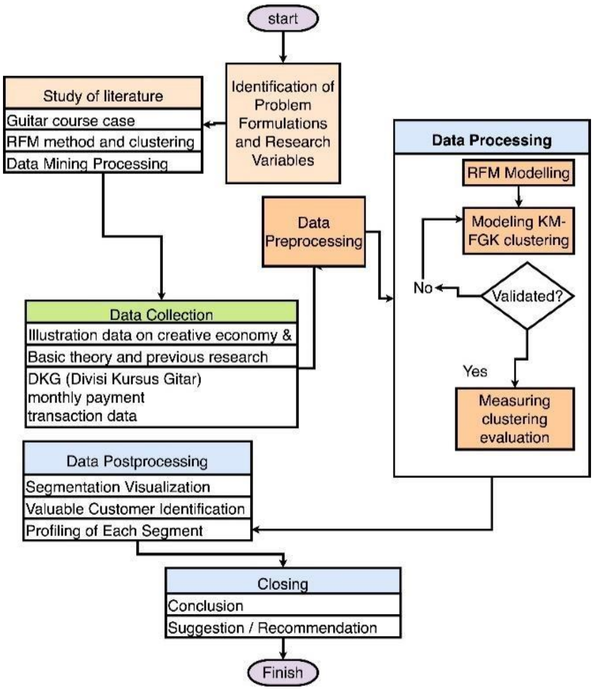

start
Study of literature
Guitar course case
RFM method and clustering
Data Mining Processing
Identification of Problem Formulations and Research Variables
Data Processing
RFM Modelling
Modeling KM-FGK clustering
Data Preprocessing
No
Validated?
Yes
Measuring clustering evaluation
Data Collection
Illustration data on creative economy &
Basic theory and previous research
DKG (Divisi Kursus Gitar) monthly payment transaction data
Data Postprocessing
Segmentation Visualization
Valuable Customer Identification
Profiling of Each Segment
Closing
Conclusion
Suggestion / Recommendation
Finish

Figure 2. Flowchart Diagram of the Research

In this research, we collected secondary data from the X guitar course, a service provider for guitar learning services that was operating since January 7th, 2017 in the Surabaya city. The data includes historical data on student payment transactions for courses from February 9th, 2018 until October 31st, 2019 with a total of 634 payment transactions from 216 unique users registered as course students in that period. The data is presented in an Excel file with the attributes of the transaction date, student's home district, student's name and nominal of payment.

There are four research variables that are used in research, which are shown in Table 3. The payment date variable is ordinal scale due to the order of differences between dates. The district and name of the student variable have a nominal scale because there are no degrees of difference. Because there is an absolute value, the nominal of Payment has a ratio scale. The first variable has a format explaining the year, month and date adjusting of the transaction. The second and third variables have general format types (no special treatment). The fourth variable has a number format because it contains the amount of money in IDR.

The first step is the preprocessing of secondary data. This step focuses on data cleaning, which includes adjusting the input format of the Excel file to the RStudio program. Each research variable derived from secondary data must be in accordance with the input format. The first predictor variable (X_1) is the date of the occurred transaction date which is initially irregular with the format of dd-mm-yy (date, month, year) that must be changed to the yyyy-mm-dd format. The second predictor variable (X_2) is the district of students. The third predictor variable (X_3) is the student's name which is initially recorded very irregularly (informal recording) so a more formal format by using an initial capital letter is needed. The fourth predictor variable is the nominal of payment, which was originally in the form of currency IDR XXX,XXX was changed to the XXX (in 1,000 IDR) number format.

The next step is the secondary data processing stage (data mining). The processing is divided into 2 phases, namely the RFM modeling which followed by the clustering process. All phases are carried out by using RStudio programming. RFM modeling phase begins with the input process by reading data from CSV and then proceed by showing data with raw_data initiation. Determination of the current day is initiated, with the current day the same as November 5th, 2019. Afterward, the nominal of payment transaction is calculated, the last payment date is searched, and the number of recency (days), as well as the number of orders/payments, is calculated for each transaction data. Those four attributes were combined with the preprocessed secondary data to form the RFM table.

Table 3.
Research variable

Variable	Variabel Name	Scale	Format
X_1	Transaction Date	Ordinal	yyyy-mm-dd
X_2	Student's District	Nominal	General
X_3	Student's Name	Nominal	General
X_4	Nominal of Payment	Ratio	Number

The RFM table contains columns for students' names, number of recency (in days), number of transactions, nominal of payment for each transaction, recency score, frequency score, and monetary score. After the RFM scores have been obtained, the next phase is the clustering phase to segment the customer based on its similarities. This phase aims to group 216 students into several segments based on the respective R, F, and M scores. This phase uses FGK clustering and K-Means clustering for comparison.

Result and Discussions

Collection of Secondary Data

The collection of customer's data is an important thing in the business industry. In this research, the main objectives for collecting data for business purposes are as follows: 1) increase the nominal payment transactions by valuable customers; and 2) retain customers who make a large profit contribution to the course. Those two business objectives need to be linked to the customer relationship management strategy in the customer segmentation process and embodied in the business objectives in the data mining process. This customer segmentation aims to build consumer profiles related to customer payment patterns and transaction

Table 4.
Input data of payment transactions in the X course

No	Transaction Date X_1	Origin of Entry	Mode	Student's Name X_2 & X_3	Amount X_4
1	2018-02-09	Studio	Cash	Fandi Sukolilo	IDR 140,000
2	2018-02-11	Studio	Cash	Lintang Sawahan Kupang	IDR 250,000
3	2018-02-13	Studio	Cash	Didin Wonokromo	IDR 70,000
⋮	⋮	⋮	⋮	⋮	⋮
642	2019-10-31	House	BCA Transfer	Angga Gedangan	IDR 300,000

Table 5.

Second input of payment transactions after data cleaning

No	Transaction Date X_1	Student's District X_2	Student 's Name X_3	Amount X_4
1	2018-08-23	Sukolilo	Abas	360
2	2019-01-27	Gubeng	Abbygail	280
3	2019-03-03	Gubeng	Abbygail	280
⋮	⋮	⋮	⋮	⋮
634	2019-07-31	Tambaksari	Zidan	500

re expected to b e distinguished in the profitabl

history so that customers a e (valuable) and unprofitable segments. The example from 642 rows of initial input data for student payment transactions obtained from the X course can be seen in Table 4.

Preprocessing

This process is the attribute selection and cleaning data stage. From a total of 5 attributes contained in the raw data, 4 attributes will be selected. Attributes that will be used for the next process are transaction date, student's home district, student’s name, and amount of payment. In the data cleaning stage, there were 8 transaction lines that are not entered because the location of the student districts was outside the target marketing areas such as Blitar, Lamongan, Pamekasan and Makassar. The example of 4 attributes and 634 rows of RFM input data in the form of student payment transactions obtained from the X course can be seen in Table 5.

Data Mining

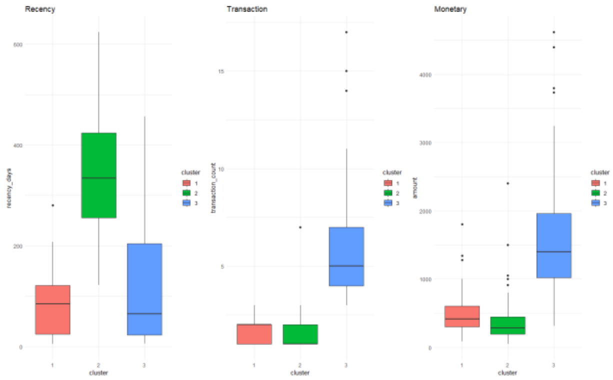
After the second input of the payment transaction is completed, it will be carried out to get a score each stage of R, F, and M by using RStudio software. The R (recency) value is the difference between the current time (November 5th, 2019) with the last time each student made a payment transaction. The R value obtained by using LUBRIDATE function. The F (frequency) value is a value that illustrates the number of payment transactions made by students. The F value is obtained from calculating the number of transaction dates with the COUNT and GROUP BY functions. The M (monetary) value is the total cost paid by students. The M value is obtained by adding up the cost with the SUM function. After the script above is run on RStudio, it will produce an output in the form of RFM scores for every 216 students of the X course which can be seen in Table 6.

Table 6.
RFM output from each student in the X course

Customer ID	Recency	Frequency	Monetary	RFM
Abas	Sukolilo	Abas	360	112
Abbygail	Gubeng	Abbygail	280	334
Abhi_Nina	Gubeng	Abbygail	280	534
⋮	⋮	⋮	⋮	⋮
Zidan	Tambaksari	Zidan	500	424

Table 7.
Example of the FGK Cluster member

Segment 1	Segment 2	Segment 3
Abhi Nina	Abas	Abbygail
Achmad	Abiantoro	Adit
Adin	Adam	Agnes
⋮	⋮	⋮
Zidan	Yudha	Zehra



RFM score becomes the input at the next stage, which is the clustering stage. The clustering stage aims to group 216 students into several segments based on the respective R, F, and M scores. The number of segments to be formed for comparison is 3, 4 and 5 clusters or segments. The clustering methods are of Fuzzy Gustafson-Kessel Clustering and K-Means Clustering which performance will be compared by using global silhouette. The highest or optimum silhouette value is 0.523 which is obtained by using the FGK method for clustering into 3 segments. Table 7 shows the example of cluster members obtained by using the FGK method for clustering into 3 segments. There are 70 students as the members of segment 1, 83 students as the members of segment 2, and 63 students as the members of segment 3.

Postprocessing

Visualization aims to help decision-makers (the X course) in analyzing customers. There are several visualizations used in this research, which are:

1. Boxplot, which describes the form of the population distribution of each segment in the form of skewness, the size of the central tendency and the size of the distribution of observational data;
2. Monetary segment size, to ensure that the logic used for transaction customer classification is sound and practical;
3. RFM score distribution, which aims to see the tendency of certain RFM scores to become members of certain segmental;
4. Calendar heatmap, useful to see the behavior of customer's payment time;

5. The 2D plot, to show the closeness between the object of observation;
6. Hotspot map, which is a geographical visualization of students as customers.

Figure 3 is a combined boxplot for recency, frequency and monetary. The leftmost boxplot shows recency where segment 1 has the narrowest whisker compared to other segments and has slices with segment 3 which has the longest whisker. There is a similarity between the newness of students in segment 1 and segment 3. The middlebox plot shows the frequency where segment 3 has the widest and highest whisker. This emphasizes the striking difference between segment 3 and other segments. The rightmost plot shows monetary where segment 3 has the widest and tallest whisker. The middle value in segment 2 is lower than segment 1. This indicates that the money paid by students in segment 2 tends to be lower than segment 1. The biggest median is in segment 3 with transactions amounting to IDR 1,400,000 then followed by IDR 412,500 in segment 2 and IDR 280,000 in segment 1.

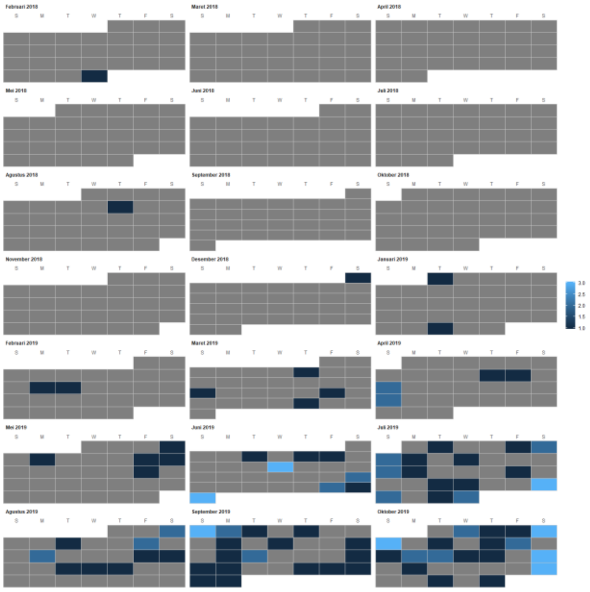


Figure 4. RFM Calender Heatmap Segment 1

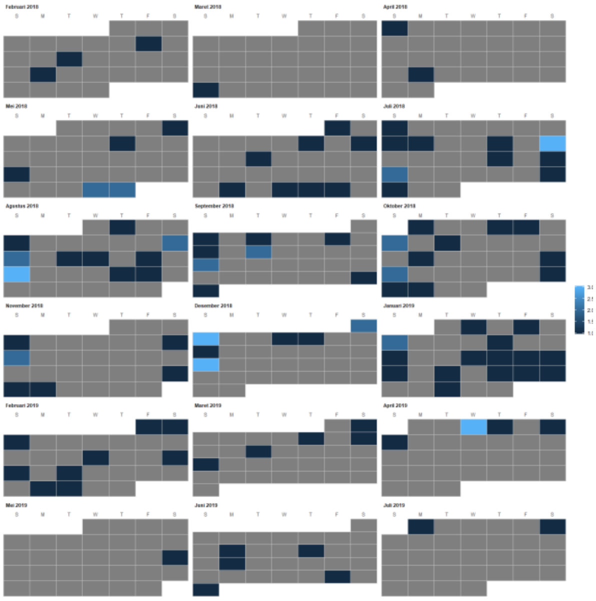


Figure 5. RFM Calender Heatmap Segment 2

Based on Figure 4, there are a number of habits practiced by students who are members of segment 1. Customers in segment 1 rarely make transactions during 2018, in fact only recorded 3 transactions, namely in February, August and December 2018. Segment 1 customers most often make transactions on days week as many as 12 days where the 5 days of the week are in September 2019. Based on the frequency of the number of transactions, segment 1 customers make the most transactions in a day are as many as 3 units of transactions and spread in September-October 2019.

Based on Figure 5, there are a number of habits performed by students who are members of segment 2. Segment 1 customers rarely make transactions in February 2018, March 2018, August 2019 and October 2019. Segment 2 customers most often make transactions on Sundays for 31 days where 4 days of which week is in July 2018. Based on the frequency of the number of transactions, segment 1 customers make the most transactions in a day is 3 units of transactions and spread in July 2018, August 2018, December 2018 (2 days) and April 2019.

Based on Figure 6, there are a number of habits practiced by students who are members of segment 3. Segment customers always make payments every month. The least number of customer days in paying for course fees occurred in March 2018 (2 transaction days) and most transactions occurred in October 2019 (21 transaction days).

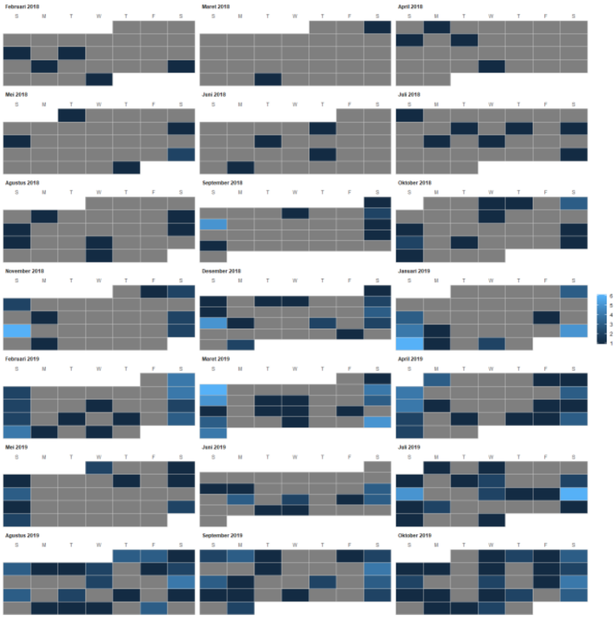


Figure 6. RFM Calender Heatmap Segment 3

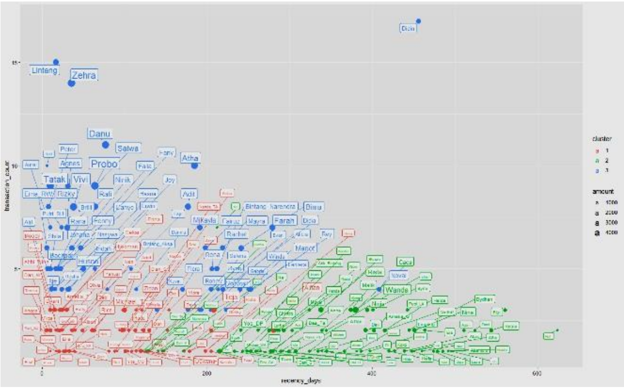


Figure 7. RFM 2D Plot Visualization

Based on Figure 7, there are differences between segments 1, 2 and 3. The larger dot circle indicates that the amount paid by the customer will be even greater and vice versa. The red color represents the members in segment 1, then the green color represents the members in segment 2 and the red color represents the members in segment 3. It appears that segment 1 has a low recency-frequency tendency, segment 2 has a high recency tendency but a low frequency is the opposite recency-frequency segment 3 highest.

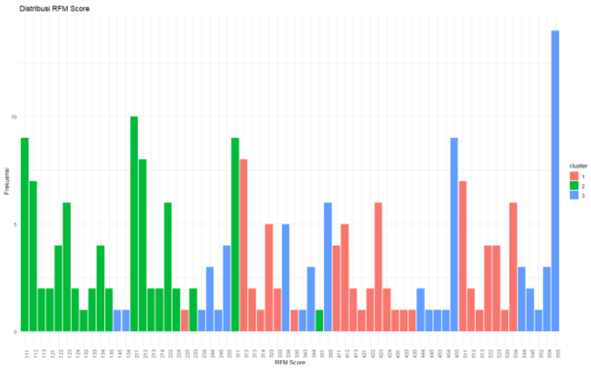


Figure 8. RFM Score Distribution Visualization

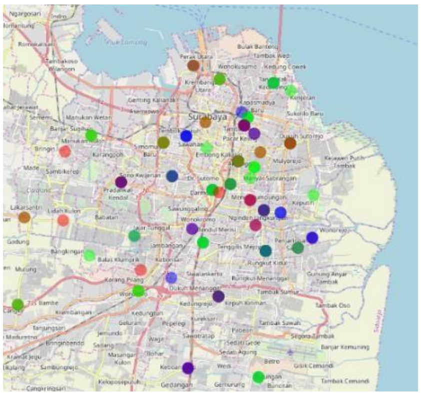


Figure 9. RFM Hotspot Map Visualization

Figure 8 explains the distribution of RFM scores into each segment. There is a tendency that the lowest RFM score of 18 categories of RFM scores has been entered into segment 2 (111, 112, 113, 121, 122, 123, 132, 133, 134, 211, 212, 213, 214, 222, 224, 233, 311). While there is a tendency that the highest RFM score has entered segment 3 even though there are 2 students who come from 2 categories of low RFM scores (145 and 154) so that the initial conclusion can be drawn that segment 2 tends to be opposite to segment 1 and segment 3.

Based on Figure 9, it can be seen that students' geography has spread. According to interviews with The X course, this geographic analysis can be utilized in determining promotion and retention strategies if there are specific patterns in certain segments. The dot that is colored red, green and blue contains a mixture of more than 1 segment. The red dot is not clearly visible, while the blue dot is clearly visible in the district of sawahan, which indicates that almost all students living in the district of sawahan are members of segment 3. The green dot looks spread without any specific pattern, so there is no need for a special promotion strategy. The following is a summary explanation of the descriptive statistical characteristics of each segment to identify the most valuable customers (segments).

Based on Table 8, there are differences in valuable segments when viewed from each indicator. Recency is better if the average, median and payment time span is shorter so that students who are listed in segment 1 are the most valuable customers. While the frequency and monetary are getting better if the average, median and range are getting bigger so that students who are enrolled in segment 3 are the most valuable customers.

Table 8.
Descriptive statistics of the X Course's student profile

Recency	Segmen 1	Segmen 2	Segmen 3
Minimum	5	122	6
Median	84,50	334	65
Mean	82,94	335,90	113,50
Range	275	502	450
Maximum	280	624	456
Frequency			
Minimum	1	1	3
Median	2	1	5
Mean	1,72	1,60	6,03
Range	2	6	14
Maximum	3	7	17
Monetary			
Minimum	Rp87.500	Rp45.000	Rp315.000
Median	Rp412.500	Rp280.000	Rp1.400.000
Mean	Rp497.100	Rp399.300	Rp1.620.000
Range	Rp1.712.500	Rp2.355.000	Rp4.305.000
Maximum	Rp1.800.000	Rp2.400.000	Rp4.620.000

Conclusions

Based on the research that has been done, it can be concluded that the RFM method followed by the Fuzzy Gustafson-Kessel (FGK) clustering can be used as the main choice in segmenting. The optimum segment is three. The results of the FGK application were evaluated with a silhouette index of 0.523. The most valuable segment is segment 3. The profile of each segment has been described in 6 types of visualization which can produce conclusions:

1. There are variations in the results of the boxplot & segment size analysis in observing central tendency. Discount policy is suitable for segment 2 so that recency can be shortened in the future. The Monetary Segment size reinforces the result that segment 3 is the most customer who makes monthly course payments of IDR 1,400,000. The X course can provide attractive promos so that students in segment 1 and segment 3 are interested in increasing their hours of study (the frequency of courses which are initially once a week becomes twice a week) so that the monetary can be pushed to close to IDR 1,400,000.
2. RFM score distribution shows the similarity of customers in segment 1 and segment 3. Because of these similarities, The X course can consider implementing the same retention strategy.
3. Calendar heatmap is useful to see the time behavior of customers' payment habits where segment 2 needs to be encouraged to make transactions immediately because recency scores tend to be low.
4. 2D plot to show the closeness between the objects of observation where if The X course wants to focus on retaining students, segment 3 can be given attractive promos. The X course also needs to increase the frequency of student transactions in segment 1 because the volume of transactions tends to be low.
5. Hotspot map to show the closeness between the objects of observation where if The X course wants to focus on retaining students, segment 3 can be given attractive promos. The X course also needs to increase the frequency of student transactions in segment 1 because the volume of transactions tends to be low.

The following are suggestions that can be considered for future works:

1. The attribute that can detect the length of a customer's membership is needed. This attribute can facilitate the X course in knowing the loyalty of students following the course.
2. Monetary attributes can be changed to profit. The largest monetary contribution can still be defeated with the largest profit contribution. In the future, The X course needs to add expenditure aspects so that profits can be analyzed.
3. Dynamic visualization is needed so that the data can be changed immediately which will automatically change the output as well.

References

S. Pratminingsih, C. Lipuringtyas dan T. Rimenta, (2013). Factors Influencing Customer Loyalty Toward Online Shopping," *International Journal of Trade, Economics and Finance,* vol. 4, no. 3.

A. Syukriah dan I. Hamdani. (2013). Peningkatan eksistensi UMKM melalui Comparative Advantage dalam rangka menghadapi MEA 2015 di Temanggung. *Economics Development Analysis Journal*, vol. 2, no. 2.

A. Rahmana. (2008). *Keragaman Definisi UKM di Indonesia*.

B. E. Kreatif. . (2008). Data Statistik dan Hasil Survei Ekonomi Kreatif, Jakarta: Badan Ekonomi Kreatif.

I. Abdulhak dan U. Suprayogi. (2012). Penelitian Tindakan dalam Pendidikan Nonformal, Jakarta: PT. Raja Grafindo Persada.

D. Jamalus. . (1988). Pengajaran Musik Melalui Pengalaman Musik. Departemen Pendidikan dan Kebudayaan, Jakarta.

V. J. Konečni. (1982). Social Interaction and Musical Preference. In Psychology of Music. dalam *Psychology of Music*, Academic Press. pp. 497-516.

S. I. Shim, W. S. Kwon dan S. Forsythe. . (2013). Enhancing Brand Loyalty Through Brand Experience: Application of Online Flow Theory," New Orleans.

M. Mohammadian dan I. Makhani. (2016). RFM-Based Customer Segmentation as an Elaborative Analytical Tool for Enriching the Creation of Sales and Trade Marketing Strategies," *International Academic Journal of Accounting and Financial Management,* vol. 3, no. 6, pp. 21-35.

K. K. Tsiptsis dan A. Chorianopoulos. (2011). Data Mining Techniques in CRM: Inside Customer Segmentation, John Wiley & Sons.

J. Ong, "Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri,* vol. 12, no. 1, pp. 10-20.

J. T. Wei, S. Y. Lin, Y. Z. Yang dan H. H. Wu. (2016). Applying Data Mining and RFM Model to Analyze Customers' Values of a Veterinary Hospital," dalam *Wei, J. T., Lin, S. Y., Yang, Y. Z., & WInternational Symposium on Computer, Consumer and Control (IS3C)*.

N. A. Amalia, D. A. Widodo dan P. P. Oktaviana. (2016). Analisis Clustering Perusahaan Sub Sektor Perbankan berdasarkan Rasio Keuangan CAMELS Tahun 2014 menggunakan Metode Fuzzy C-Means dan Fuzzy Gustafson Kessel," *Jurnal Sains dan Seni ITS,* vol. V, no. 2.

S. R. Indah. (2018). Penyelenggaraan Program Kursus Musik (Studi Pada Lembaga Lily's Music School Semarang)," *Jurnal Eksistensi Pendidikan Luar Sekolah (E-Plus),* vol. III, no. 1.

S. Payne dan P. Frow (2005). A Strategic Framework for Customer Relationship Management," *Journal of marketing,* vol. 69, no. 4, pp. 167-176.

R. Kalakota dan M. Robinson. (2001). E-business 2.0: Roadmap for Success, 2nd penyunt., Boston: Addison-Wesley Longman Publishing.

P. Kotler, K. L. Keller, M. Brady, M. Goodman dan T. Hansen. (2012). Marketing Management, England: Pearson Education Limited.

D. Peppers dan M. Rogers. (2011). Managing Customer Relationship, 2nd penyunt., New Jersey: John Wiley & Sons Inc.

M. J. Berry dan G. S. Linoff. (2004). Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management, 2nd penyunt., Indianapolis: John Wiley & Sons.

S. Jansen, "A Vodafone Case Study. (2007). dalam *Customer segmentation and customer profiling for a mobile telecommunications company based on usage behavior*, Maastricht, 2007, p. 66.

A. M. Scridon. (2008). Understanding Customers – Profiling Segmentation," Babes Bolyai University, Cluj.

B. J. Bronnenberg dan P. Albuquerque. (2003). Geography and Marketing Strategy in Consumer Packaged Goods. In Geography and Strategy," *Emerald Group Publishing Limited,* pp. 215-237.

J. Blythe dan P. Megicks. (2010). Marketing Planning: Strategy, Environment and Context, Pearson Education Canada.

A. Hughes. (1994). Strategic Database marketing, Probus.

J. R. Bult dan T. Wansbeek. (1995). Optimal Selection for Direct Mail," *Marketing Science*, vol. 14, no. 4,pp. 378-394.

J. Wu dan Z. Lin. (2005). Research on Customer Segmentation Model by Clustering," *7th International Conference on Electronic Commerce,* pp. 316-318.

R. J. Baran dan R. J. Galka. (2013). CRM: The Foundation of Contemporary Marketing Strategy, Routledge.

W. N. Venables dan D. M. Smith. (2013).*An Introduction to R.*

R. A. Johnson dan D. W. Wichern, (2007). Applied Multivariate Statistical Analysis, 6th penyunt., New Jersey: Pearson Prentice Hall.

S. Sharma. (1996). *Applied Multivariate Techniques,* New York: Wiley.

B. Ruswandi, (2008). *Diktat Perkuliahan Praktikum Statistika Multivariat*, Studi Matematika Fakultas Sains dan Teknologi UIN Jakarta.

M. R. Anderberg. (2014). Cluster Analysis for Applications: Probability and Mathematical Statistics: a Series of Monographs and Textbooks, 19th penyunt., Academic Press.

B. Feil, B. Balasko dan J. Abonyi. (2007). Visualization of Fuzzy Clusters by Fuzzy Sammon Mapping Projection: Application to The Analysis of Phase Space Trajectories, Springer.

A. Mauliyadi, M. Sofyan dan M. Subianto. (2003). Perbandingan Metode Fuzzy C-Means (FCM) dan Fuzzy Gustafson-Kessel (FGK) Menggunakan Data Citra Satelit Quickbird," *Jurnal Transenden*.

P. J. Rousseeuw .(1987). "Silhouettes: a Graphical Aid to The Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65.

X. Wang dan Y. Xu. (2019). An Improved Index for Clustering Validation based on Silhouette Index and Calinski-Harabasz Index,.