Data-Driven Strategies for Fuel Distribution in Indonesia: A Case Study of PT Pertamina Patra Niaga

Kania Lovia Tiarazahra[1], Rita Ambarwati[2]\*

Management Study Program, University of Muhammadiyah Sidoarjo, Indonesia [1]

Management Study Program, University of Muhammadiyah Sidoarjo, Indonesia [2]

University of Muhammadiyah Sidoarjo

Sidoarjo, Indonesia

ritaambarwati@umsida.ac.id

***Fuel oil or what is often referred to as BBM is one of the basic needs to drive all community activities. So the government appointed PT Pertamina as a single company which is a state-owned company to facilitate fuel needs for all levels of society. However, with increasing demand, the government formed a new policy to allow private companies to come in to meet all fuel demand. With this, PT Pertamina is no longer the only fuel supplier in Indonesia and must continue to develop mature strategies so that profits do not fade. One way is by examining sales data and predicting customer loyalty. The RFM method followed by the decision tree algorithm and k-means clustering is applied in this research, with the output being able to determine the level of customer loyalty, the level of salesman performance, as well as predicting the potential for customers to churn and its correlation with the salesman's skills. The data used as a reference for the research is sales transaction data obtained from PT Pertamina Patra Niaga Regional Jatimbalinus. And from the research, results showed that the majority of PT Pertamina Patra Niaga Regional Jatimbalinus customers are loyal customers. With a salesman, performance is divided into good performance and less good performance. This grouping is obtained based on the salesman's overall performance track record. As for customer churn predictions, it was found that there was 1 group of customers who were predicted to churn heavily, but this was not influenced by salesman performance, as evidenced by transaction track records in existing data***

***.Keywords—*** ***Churn; Customer Segmentation; Salesman Clustering; Decision Trees; K-means Clustering***

***Abstract*—** **Fuel oil or what is often referred to as BBM is one of the basic needs to drive all community activities. So the government appointed PT Pertamina as a single company which is a state-owned company to facilitate fuel needs for all levels of society. However, with increasing demand, the government formed a new policy to allow private companies to come in to meet all fuel demand. With this, PT Pertamina is no longer the only fuel supplier in Indonesia and must continue to develop mature strategies so that profits do not fade. One way is by examining sales data and predicting customer loyalty. The RFM method followed by the decision tree algorithm and k-means clustering is applied in this research, with the output being able to determine the level of customer loyalty, the level of salesman performance, as well as predicting the potential for customers to churn and its correlation with the salesman's skills. The data used as a reference for the research is sales transaction data obtained from PT Pertamina Patra Niaga Regional Jatimbalinus. And from the research, results showed that the majority of PT Pertamina Patra Niaga Regional Jatimbalinus customers are loyal customers. With a salesman, performance is divided into good performance and less good performance. This grouping is obtained based on the salesman's overall performance track record. As for customer churn predictions, it was found that there was 1 group of customers who were predicted to churn heavily, but this was not influenced by salesman performance, as evidenced by transaction track records in existing data**

***Keyword—*** ***Churn, Segmentasi Pelanggan, Klasterisasi Salesman, Decision Tree K-means Clustering***

# Introduction

Fuel Oil, often called BBM, is one of the needs that underlies all community activities, starting from personal activities to industrial activities. Fuel plays an important role as an element of the economic wheel, both directly and indirectly. Where BBM can be an item that is traded, it can also be a mandatory item so that the operational wheels can turn. PT Pertamina is a state-owned enterprise, one of the directions of the company's movement is to provide fuel, both for the retail sector and the industrial sector. By Law No.8 of 1971, PT Pertamina is authorized by the government to produce and process oil and gas products as well as provide fuel oil and gas needs in Indonesia[1].

As time progresses, the population is increasing, and the need for fuel is also increasing. This is inversely proportional to the availability of fuel in nature which is increasingly depleting. Thus, by Law No.22 of 2001, the government opened the door for private companies and changed the role of PT Pertamina to no longer be the only business actor in the oil and gas sector. This new change in taxation led to an increase in downstream business actors so that consumers have more diverse choices at competitive prices. As was the case in 2006, there was competition between Pertamina, Shell, and Petronas in terms of the distribution of non-subsidized fuel as their prices raced against each other.

So to face business competition, PT Pertamina focuses its business and divides it into several parts according to its business sector, one of which is PT Pertamina Patra Niaga which focuses on PT Pertamina's downstream business, precisely on the distribution and marketing business. PT Pertamina Patra Niaga has full authority in conducting the downstream business of PT Pertamina. This is supported by adequate infrastructure, which can support the distribution and marketing of products produced by Pertamina, such as fuel products, lubricants, LPG, and other products that are ready to meet the demands of consumer needs both in the retail sector and the corporate sector.

Therefore, to continue to maintain and increase company profits, PT Pertamina Patra Niaga needs to look a few steps ahead of the present to see potential profits and losses, one of which is by examining consumers who have the potential to move to competing companies or competitors, hereinafter referred to as churn. Then look for a correlation between their churn potential and the performance of the salesmen that PT Pertamina Patra Niaga has. With the hope that in the future this research can become a new reference in developing sales strategies.

Churn is a term that refers to the meaning of switching consumers to competing companies. As written in research conducted by Radius Poniman, Opim Salim, and Sri Mulyani in an article entitled "Churn Analysis of Cellular Telecommunication Broad Band Services in Sumbagut", the definition of churn is the number of service users who no longer subscribe to service usage. In other studies, the definition of churn is also mentioned. According to Agung Rezkina's view stated in his research entitled "Telkomsel Surakarta's Marketing Public Relations Strategy in Maintaining Customer Loyalty", it is stated that churn is a word used to define customer switching to another operator. So, if churn is placed in this study, the meaning of churn will be adjusted to become the movement of consumers to other competing companies.

This research is focused on the corporate sector where customers are grouped into several groups, namely Industrial Fuel Agents, Bunker Agents, Sea Transportation, BU-PIUNU, Industry, Government Agencies, Public Services, Fisheries, PLN, POLRI, TNI-AD (Army), TNI-AL (Navy), TNI-AU (Airforce), and Land Transportation. Also accompanied by research on salesmen in the corporate sector sales department totaling 8 people. The data used in the research is fuel sales transaction data for a period of 2 years, namely 2021-2022. The method used in finding answers to these objectives uses several methods, namely starting with scoring customer groups through the RFM model (Recency, Frequency, Monetary) to find out which customer groups have the potential to move to competitors or not. The same method is also used in scoring salesmen in interpreting their performance track records which will then be interpreted in Decision Tree C4.5 modeling and K-Means Clustering.

The first method carried out in the research is the analysis method using the RFM model. The RFM model is used to identify the characteristics of each customer group to form a class of each customer group. Analysis with this model is a classic approach that is carried out to determine the interaction patterns and customer patterns in general. After knowing the classes of customers with each different category, then calculated using the C4.5 decision tree algorithm to visualize the results of data processing in the form of a decision tree based on decision-forming criteria.

The same applies to salesman clustering. Starting with giving value to each salesman's track record calculated through the RFM modeling method, which is then interpreted in the k-means clustering algorithm modeling. Kmeans clustering is one of the data mining models for grouping data according to its closest characteristics.[2]. In this study, using the Davies Bouldin method, 2 groups were found with the closest characteristic distance translated as "good performance" and "poor performance" which will then be further investigated regarding its correlation with customer groups that have the potential to churn.

Decision tree modeling and k-means clustering have provided many results in previous studies. For example, in 2018 research was conducted by Ni Wayan Wardani, Gede Rasben Dantes, and Gede Indrawan regarding the prediction of consumer loyalty in a retail company which was also carried out with decision tree modeling, and the results of the test showed that 3 predictions were less precise than the 259 cases studied, which meant that data modeling with this decision tree was quite accurate. Also the same for k-means clustering. The research was carried out in 2021 by Elly Muningsih, Ina Maryani, and Vembria Rose Handayani regarding provincial clustering based on village potential, and the division of clusters in this study is accurate, which is proven by the Davies Bouldin method which shows the results of optimizing the number of clusters is 3 with a davies bouldin index (DBI) value that shows 0.175, which is smaller than the number of DBIs in other clusters.[3].

This research focuses on the attachment of 2 different data processing results, namely between data processing results related to customer groups, and data processing results related to salesmen. Researchers have a simple assumption that the customer group that has the potential to churn has a big influence from salesmen who are less skilled in attracting customers. So from this research, it will be proven, is it true that the group of customers who have the potential to churn is caused by salesmen who are less skillful in attracting consumers. If examined through past research, the shrewdness of a salesman at work, especially in communication skills, is one of the important things that can support marketing success, as conducted by Ahmat Arif Syaifudin and Tutik Al-fiyah in their research which discusses the implementation of integrated marketing communication in supporting marketing success in one of the companies in 2022[4].

# Methods

The data used as research material in this journal is secondary data originating from related companies. The data used is data that describes purchases made by consumers which are grouped into several groups based on the type of consumer itself, which will then be segmented again based on the consumption level of each consumer. The period taken is for the past 2 years, to be precise in 2021-2022.

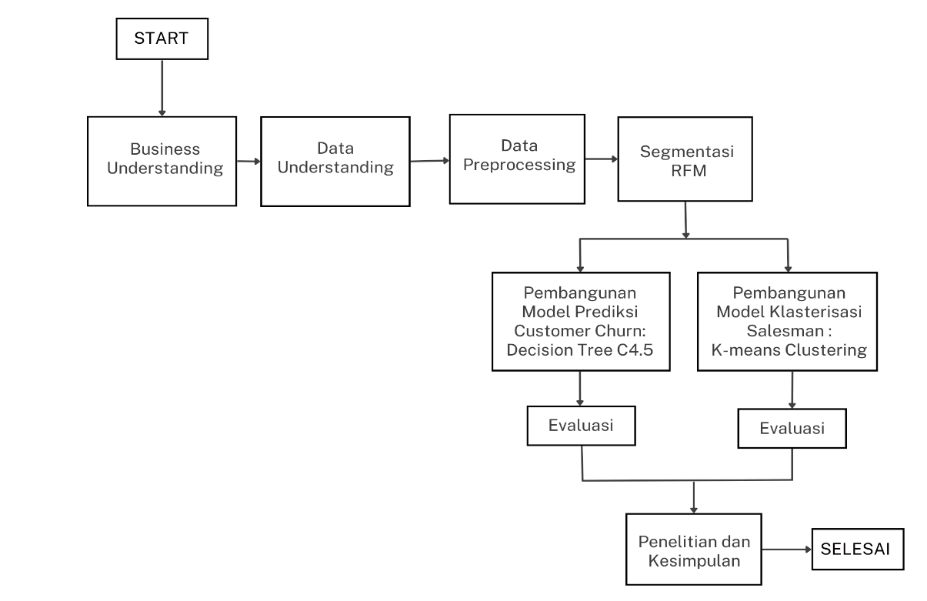


Fig. 1. Diagram of Research Highlights.

In analyzing the data, several analytical techniques are used in sequence to interpret the results of data processing that is easy to understand. The sequence of analysis techniques is shown in Figure 1.

## Business Understanding

Business understanding is a stage of understanding related to the business to be studied. Understanding covers business objectives, the actual situation in the field, and a general design related to the purpose of conducting research using existing data.

## Data Understanding

Data Understanding is a stage where an understanding of the data that has been obtained is carried out. At this stage, the researcher gets help from the data owner source to understand the data provided. The data obtained is daily consumer transaction data with various types of materials, located in the East Java, Bali, and Nusa regions for 2 years. 113,270 data records for 2021 and 110,364 data records for 2022

## Data Preprocessing

This is the stage where the data is sorted and prepared for the data mining process. All unused attributes will be cleaned up to leave the attributes that are important for the data processing stage. In RFM modeling, both in the customer segment and the salesman segment, the attributes used are the customer group, salesman, calendar per day, amount of income, and description of materials purchased

## RFM Segmentation

RFM is a method of grouping customer classes based on transaction loyalty. In this segmentation process, a score of 1 to 3 is given to each recency, frequency, Image 2. Research Overview Diagramand monetary domain in each customer group. Score 3 for the highest score and score 1 for the lowest score where the final score is determined by a combination of scores in each domain.

The recency score is taken from the distance between the last transaction and the current date. The customer group that has the closest transaction date to the present has the highest score, which is a score of 3. Also related the frequency value is taken from the frequent number of transactions from each customer group. The customer group that has the most frequent number of transactions is given the highest score, which is score 3. As for the monetary value, it is taken from the total number of transactions. Related to the score range of each domain is obtained from processing raw data in Excel with the help of the pivot data feature. And obtained the score range as shown in Table 1.

TABLE I. RFM SCORE AND CLASS RANGE FOR CUSTOMER GROUP SEGMENT

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | **Value** | **Score** | **Meaning** |
| *Recency* | o - ≤44658 | 1 | Longest |
| ≤44659 - ≤44922 | 2 | Somewhat Longer |
| ≤44923 - 44927 | 3 | New |
| *Frequency* | o - ≤3550 | 1 | Rare |
| ≤3551 - ≤14109 | 2 | Medium |
| ≤14110 - 84794 | 3 | Often |
| *monetary* | 0 - 186520927985 | 1 | low |
| ≤186520927986 - ≤1463821002601 | 2 | Medium |
| ≤1463821002602 - 15370672220547 | 3 | high |

As for the scores for the salesman segment, there are only differences in the frequency and monetary domains, not for recency, because, in the raw data, all salesmen have the last transaction data recorded on the same day. So the table for the salesman segment is as shown in table 2 below.

TABLE II. RFM SCORE AND CLASS RANGE FOR THE SALESMAN SEGMENT

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | **Value** | **Score** | **Meaning** |
| *Frequency* | 0 - 18124 | 1 | Rare |
| ≤18125 - ≤37254 | 2 | Medium |
| ≤37255 - 47947 | 3 | Often |
| *monetary* | 0 - 2515383469894 | 1 | low |
| ≤2515383469895 - ≤3958875694211 | 2 | Medium |
| ≤3958875694212 -15472713667590 | 3 | high |

## Construction of Customer Churn Prediction Model with Decision Tree C4.5

Data mining for research on customer group segments using C4.5 decision tree modeling. The data processed in this modeling uses data that is categorical and not numerical [10]. After determining the score for each domain, the next step is determining the customer label. Customer labels in this research are determined by several categories. The first is the Superstar category. This category is aimed at the customer class that has the highest score. Next is the Golden category. The Golden category is aimed at the class that has the highest monetary score after superstar, a high frequency score, and an average recency score. The next category is the Typical category. This category is intended for customer classes that have an average score on each domain. Next is the Occational category. This category is intended for the customer class that has the second lowest monetary score after dormant, but has a high score for frequency and a low score for recency. Next is the Everyday category. This category is intended for the class of customers who have an average recency score. The frequency score is low, and the monetary score is medium to low. And the last one is the Dormant category. This category is intended for the customer class that has the lowest combined score.

Based on the division using the RFM method, there are 27 customer classes that have various combinations of scores, with the highest score being 333 and the lowest score being 111. This score table with category labels will determine the average class of customers owned by the company, whether loyal or not. and can be a determining factor in formulating the next strategy. The score table with category labels is shown in table 3 as follows.

TABLE III. SCORE TABLE WITH CATEGORY LABELS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **SCORE** | | | **Label Cust Group** |
| **R** | **F** | **M** |
| K1 | New | Rare | low | everyday a |
| K2 | New | Rare | Medium | typical b |
| K3 | New | Rare | high | golden c |
| K4 | New | Rare | low | everyday a |
| K5 | New | Some  what often | Medium | superstar c |
| K6 | New | Some  what often | high | superstar b |
| K7 | New | often | low | typical d |
| K8 | New | often | Medium | golden a |
| K9 | New | Rare | high | golden c |
| K10 | New | Rare | low | everyday a |
| K11 | New | Rare | Medium | typical b |
| K12 | New | Rare | high | golden c |
| K13 | New | Some  what often | low | typical c |
| K14 | New | Some  what often | Medium | superstar c |
| K15 | New | often | high | a superstar |
| K16 | New | often | low | typical d |
| K17 | New | Rare | Medium | typical b |
| K18 | New | Rare | high | golden c |
| K19 | New | Rare | low | everyday a |
| K20 | Some  what Longer | Rare | Medium | everyday c |
| K21 | Some  what Longer | Some  what often | high | golden b |
| K22 | Some  what Longer | Some  what often | low | everyday d |
| K23 | Some  what Longer | often | Medium | everyday e |
| K24 | Longest | often | high | golden d |
| K25 | Longest | Rare | low | a dormant |
| K26 | Longest | Rare | Medium | dormant b |
| K27 | Longest | Rare | high | golden e |

Furthermore, this table will be processed into the RapidMiner application for C4.5 decision tree modeling, whose detailed process will be shown in the results and conclusions chapter.

## Construction of Salesman Clustering Model with K-means Clustering

Modeling using the k-means clustering algorithm for the salesman segment is not much different from the customer group segment. Before finally entering k-means clustering modeling, the data must first be processed with RFM modeling to be able to highlight the characteristics required by each salesman. So from RFM modeling, it is known that the value possessed by each salesman from each domain is as follows as shown in table 4.

TABLE IV. RFM VALUE FOR EACH SALESMAN

|  |  |  |  |
| --- | --- | --- | --- |
| **SBM region** | **R** | **F** | **M** |
| SBM Industry I | 0.75890411 | 41503 | 8460422329347 |
| SBM Industry II | 0.75890411 | 20346 | 3958875694211 |
| SBM Industry III | 0.75890411 | 47947 | 2515383469894 |
| SBM Industry IV | 0.75890411 | 18124 | 1305280048298 |
| SBM Industry V | 0.75890411 | 37254 | 2620658189947 |
| SBM Industry VI | 0.75890411 | 33424 | 15472713667590 |
| SBM Industry VII | 0.75890411 | 12190 | 2411370220183 |
| SBM Industry VIII | 0.75890411 | 12846 | 2642144188420 |

From this table, it can be seen that the *recency* domain has the same value, because from the raw data itself, each salesman has the same track record for dates, and there is no accompanying hour information. After obtaining the value of each domain for each salesman, the next is data processing in RapidMiner with the *k-means clustering* algorithm to determine the cluster of each data.

To determine the number of clusters for determining the centroid point, researchers use the Davies-Bouldin method which looks at the proximity distance between the data and the center point of each cluster from the predetermined center. This method has an internal scheme test, where the suitability of the number of clusters is seen by comparing the Davies Bouldin value between the number of clusters. This method was first disseminated by David L. Davies and Donald W. Bouldin in 1979.[3]. As for the salesman data in this study, the Davies Bouldin method is explained in the following table.

TABLE V. DAVIES BOULDIN VALUE FOR EACH CLUSTER

|  |  |
| --- | --- |
| **Davies Bouldin** | |
| **k** | **DB** |
| 2 | 1.549 |
| 3 | 2.900 |
| 4 | 3.592 |
| 5 | 3.829 |

It can be seen from the table above, that the smallest closeness value is owned by cluster 2 which is 0.959, which means that the right cluster division for salesman data is 2 clusters. Henceforth, k-means clustering processing on RapidMiner is divided into 2 clusters and will be explained in more detail in the results and conclusions chapter.

Clustering salesman data is done to distinguish between groups of salesmen who have good performance and groups of salesmen who have poor performance. If it has been grouped, then it will be easier for the author to decide the answer to the research question "Is the group of customers who have the potential to churn caused by the lack of salesman skills?".

# Results and Discussion

* 1. *Test Results on Customer Group Data*

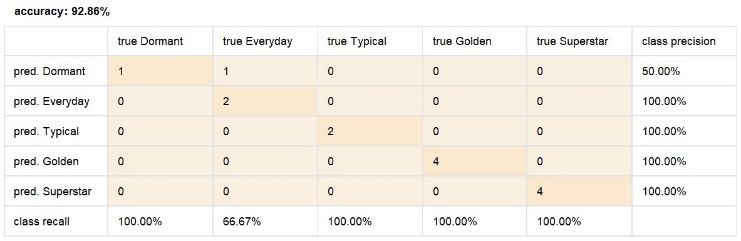
Testing on customer group data is done in 2 ways, namely with RFM modeling and classification using the C4.5 decision tree algorithm. RFM modeling, in addition to being used for data preprocessing before modeling using a decision tree, RFM modeling that includes scoring can also be used to designate 1 customer group that has the highest percentage of churn. Modeling through the C4.5 decision tree algorithm can be used to identify the character of the BBM industry customer group at PT Pertamina Patra Niaga Regional Jatimbalinus as a whole, to get a conclusion on the majority of customers owned by the company and whether they have a high level of loyalty or not, as well as one of the determinants of whether the current strategy is appropriate.

Fig. 2. PerformanceVector Classification Table for Customer Group Data.

As stated in Chapter II, RFM modeling is used them to prepare data which is then processed in decision tree modeling. RFM modeling can also be used to determine which customer group has the highest percentage of churn.

TABLE VI. RFM LABEL TABLE FOR CUSTOMER GROUP DATA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Customer group** | **Max of Calendar Day** | **Frequency** | **Monetary** | **Label** |
| PLN | 44926 | 84794 | 15,370,672,220,547 | Superstar |
| Industry | 44926 | 23419 | 11,069,681,813,245 | Superstar |
| Industrial Fuel Agent | 44926 | 37068 | 4,167,413,285,932 | Superstar |
| Sea Transportation | 44926 | 22285 | 3,973,001,549,093 | Superstar |
| Navy | 44926 | 12003 | 1,463,821,002,601 | Golden |
| Agent Bunker | 44923 | 2833 | 1,086,402,885,241 | Typical |
| POLRI | 44926 | 13242 | 678,250,613,276 | Golden |
| TNI - AD (Army) | 44922 | 8532 | 579,156,467,053 | Golden |
| Land Transportation | 44926 | 14109 | 556,924,549,043 | Golden |
| TNI - AU (Airforce) | 44923 | 3550 | 186,520,927,985 | Typical |
| Government Agencies | 44926 | 1169 | 141,688,074,946 | Everyday |
| BU-PIUNU | 44658 | 451 | 87,070,804,854 | Everyday |
| Fisheries | 44848 | 97 | 13,986,458,975 | Everyday |
| General Services | 44463 | 82 | 12,257,155,099 | Dormant |

The table 3 shows the RFM value for each customer group, as well as the label determined based on the value range of each customer group. So it can be seen, that the customer group that has great potential in churning is the customer group in the "general service" segment, where the RFM value of this customer group is the lowest compared to other customer groups.

As for modeling with the decision tree algorithm, data processing begins with RFM modeling which then produces 5 class categories, namely dormant, everyday, typical, golden, and superstar. This is then processed using the C4.5 decision tree algorithm in RapidMiner, and the following results are obtained.

The table is a table generated from the *performance classification* operator*.* This operator is useful as a validity tester of the test data. The results obtained from this table where the total accuracylevel is 92.116%. This means that this test has a high level of validity.

From data processing using a *decision tree, the* results of the decision tree are as follows

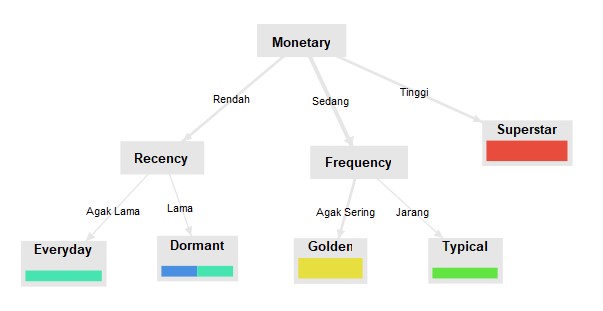


Fig. 3. Class Category Decision Tree of Customer Group Data.

It can be seen from the decision tree that the factor that most influences the category grouping is the factor of monetary or in other words the total transaction amount factor, which is then followed by the frequency and recency factors. From this decision tree, it can be concluded that the majority of PT Pertamina Patra Niaga regional Jatimbalinus customer groups in the Fuel Industry segment fall into the superstar category, which means that there is little chance for these customer groups to churn. In other words, it can be said that when viewed through this research method, the strategy implemented in the last 2 years has been included as good, judging by the majority of customers who have high loyalty.

* 1. *Test Results on Salesman Data*

Salesman data testing is tested through 2 methods, RFM modeling and *k-means clustering* algorithm modeling*.* Through the RFM method, transaction data that has been carried out by salesmen is processed in such a way that it is then modeled in RapidMiner to extract the essence of the data processing results. In this study, researchers used *kmeans clustering* to find out the lowest salesman performance among the salesmen that the company has, which will then be studied about the purchasing power of customer groups in the same segment.

Tests carried out on salesman data using the RFM method, produce the following output

TABLE VII. TEST RESULTS USING RFM MODELING ON SALESMAN DATA

|  |  |  |  |
| --- | --- | --- | --- |
| **SBM region** | **R** | **F** | **M** |
| SBM Industry I | 0.75890411 | 41503 | 8460422329347 |
| SBM Industry II | 0.75890411 | 20346 | 3958875694211 |
| SBM Industry III | 0.75890411 | 47947 | 2515383469894 |
| SBM Industry IV | 0.75890411 | 18124 | 1305280048298 |
| SBM Industry V | 0.75890411 | 37254 | 2620658189947 |
| SBM Industry VI | 0.75890411 | 33424 | 15472713667590 |
| SBM Industry VII | 0.75890411 | 12190 | 2411370220183 |
| SBM Industry VIII | 0.75890411 | 12846 | 2642144188420 |

The test results of hundreds of thousands of salesman transaction data summarized according to each salesman are shown in the table. It is found from the test results, that there are 8 salesmen who each have RFM scores according to the transaction traces they have made in the last 2 years. Furthermore, the results of this test will be processed in RapidMiner to be able to retrieve test results according to the objectives of this study.

The results of the salesman data test with RFM modeling are processed in RapidMiner using the *k-means clustering* algorithm*.* In this algorithm, each salesman will be grouped based on the similarity of the characteristics of each data with a predetermined centroid point. To determine the number of centroid points, researchers use the Davies Bouldin method as described in Chapter II. So from this method, it is found that the division of the number of centroid points for the number of clusters for salesman data is 2 points, which then these 2 points become the number of clusters that will be called clusters with 'good performance', and clusters with 'poor performance'. After processing with RapidMiner, the following modeling results were obtained.

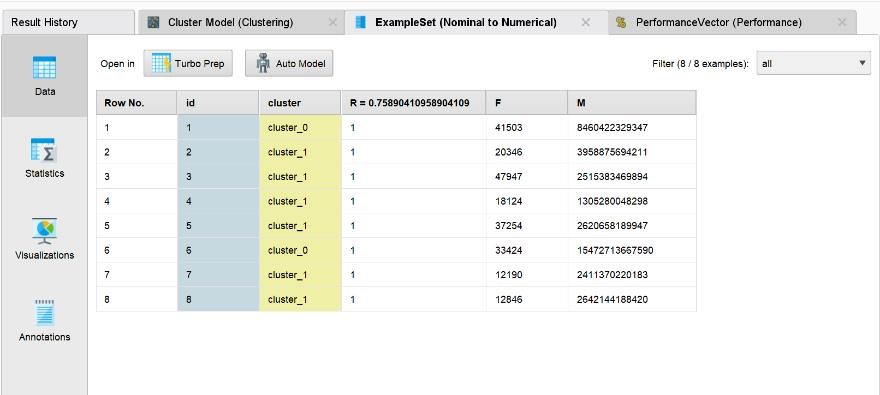


Fig. 4. Modeling Results of K-means Clustering on Salesman Data.

A new yellow column labeled *cluster\_0* and *cluster\_1* appears between the existing columns. This column shows the grouping of data according to the closeness of their respective characteristics*.* So it can be seen which salesmen are included in *cluster* 0 and which salesmen are included in *cluster* 1. Then from the table, the following results are obtained:

* *Cluster* 0 contains salesmen with good performance categories, which are characterized by their higher RFM values.
* *Cluster* 1 contains salesmen with poor performance categories, which are characterized by their lower RFM values.
* From the table, it can be seen that salesmen IV and VII have the least good performance among others, which is characterized by the lowest RFM value among other salesmen.
  1. *Identify the relationship between the results of both types of data*

So far, the research that has been done above reveals specific results between the customer group data studied to show customer groups that have the potential to churn, and the salesman data studied to find out which salesman has the most 'poor' performance. And it is found that the customer group that has great potential to churn is the customer group in the "general service" category, and the salesmen who have the most 'poor' performance are "salesmen IV and VII". So then, to answer the research question "Is the customer group that has the potential to churn caused by the lack of salesman skills?" further research will be carried out from the relevant data from these two results.

Starting with a search of the customer group data in the "general services" segment, the search resulted in the following key points:

* Taken from the raw data provided directly by the company concerned, the customer group in the "general service" category is entirely served by the salesman I. This finding is a little odd, as a salesman I am a salesman who is included in cluster 0 or a cluster with good performance. This finding is a little odd, where salesman I am a salesman who is included in cluster 0 or a cluster with good performance.
* With the same source, namely raw data obtained directly from related companies, with a focus on transaction dates, it was found that the customer group with the category "general services" made the last purchase on September 24, 2021, and there were no more transaction records after that date. This result raises a little question, why transactions were only made until September 24, 2021, and there were no other transactions after that, even though it is known from the data that before September 24, 2021, this customer group consistently made transactions with related companies every month.

Followed by a data search on the data of salesmen IV and VII who have the most 'poor' performance among other salesmen, the search resulted in several conclusions listed in the following points:

* Based on the raw data obtained from the company, it is found that the customer group served by Salesman IV is the majority of customer groups that fall into the category of customer groups that have little potential to churn.
* Similarly, for the data of customer groups handled by salesman VII, the majority of these customer groups fall into the category of customer groups that have little potential to churn.

# Conclusion

The conclusions in this research are seen from several sides. Starting from the customer group data side, it was found that the majority of the company's customers are loyal customers, with the number of customers in the superstar class totaling 4 customers, for the golden class 4 customers, for the typical class 2 customers, for the everyday class 3 customers, and for the dormant class 1 customer. And for customer groups that have great potential to churn, one customer group was found, namely the customer group that has the dormant label. This is known from the lowest RFM score. And the customer class that has this dormant label is the customer group in the "general services" segment. As for the conclusions for salesmen, based on the RFM and Davies Bouldin method in dividing clusters, it was found that the clusters were divided into 2, which were then named the 'good' performance cluster and the 'poor' performance cluster. With the dominance of cluster charging on 'poor' performance. So from here a new strategy can be formed to improve the overall performance of salesmen. The conclusion from the correlation between the two types of data (data from customer group transactions and data from salesman sales) was that the factor for the high percentage of customer groups to churn was not based on the salesman who handled it. This is proven by the results of research on the customer group in the "general service" segment which is labeled dormant, where this customer group is actually served by salesman I, where salesman I is actually included in the 'good' performance cluster. So from this conclusion it can also be seen that the strategy implemented in the last 2 years has succeeded in retaining customers from quickly switching to other competitors. This research was conducted on the basis of original data obtained directly from the relevant company. However, this research still has shortcomings and limitations in obtaining detailed information because the researcher is not someone who is directly involved in the world of work at the company concerned. So, to achieve more detailed and complete research results, more information can be collected. Such as in-depth interviews with related parties, or collecting more complete supporting data from various accredited sources..

##### Acknowledgment

The researcher would like to thank all parties who helped this research so that it can be published as it should. To the parties of PT Pertamina Patra Niaga Jatimbalinus region, for providing assistance with all the information needed for the success of this research. Also to the university which has facilitated researchers so that they can carry out this research as well as possible.

##### References

1. N. D. Japari, A. Zafrullah TN, and F. R. Djoemadi, "The Role of Pt. Pertamina as a fuel oil supply provider in Indonesia," Calyptra, vol. 7, no. 2, pp. 1-12, 2019.
2. R. Muliono and Z. Sembiring, "Data Mining Clustering Using K-Means Algorithm for Clustering Lecturer Teaching Tridarma Level," CESS (Journal Comput. Eng. Syst. Sci., vol. 4, no. 2, pp. 2502-714, 2019.
3. E. Muningsih, I. Maryani, and V. R. Handayani, "Application of K-Means Method and Optimization of Number of Clusters with Davies Bouldin Index for Clustering Provinces Based on Village Potential," J. Sains dan Manaj., vol. 9, no. 1, pp. 95-100, 2021, [Online]. Available:

https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839.

1. A. A. Syaifuddin and T. Al-Fiyah, "Implementation of Integrated Marketing Communication in Supporting Marketing and Brand Equity of BMT Mandiri Artha Sejahtera East Java," Juornal Econ. Policy ..., vol. 03, no. 02, pp. 37-51, 2022, [Online]. Available:

https://ejournal.uinsatu.ac.id/index.php/jesk/article/view/7126%0Ahttps://ejournal.uinsatu.ac.id/index.php/jesk/article/download/7126/2115.

1. F. Soufitri, E. Purwawijaya, E. H. Hasibuan, and R. N. Singarimbun, "Testing C4.5 Algorithm Using Rapid Miner Applications in Determining Customer Satisfaction Levels," J. Infokum, vol. 9, no. 2, pp. 510-517, 2021, [Online]. Available:

http://infor.seaninstitute.org/index.php/infokum/index.

1. Y. Primawati, I. Verdian, and G. W. Nurcahyo, "K- Means Clustering on Based Classification Method of Sales Agent," J. Comput. Since Inf. Technol., vol. 7, no. 2, pp. 1-6, 2021, doi:

10.35134/jcsitech.v7i2.1.

1. P. Algorithm, C. Using, and A. Rapid, "In determining the level of customer satisfaction," vol. 0, no. 2, pp. 510-517, 2021.
2. M. A. Kadafi, "Evaluate the potential bankruptcy of Indonesian oil and gas mining companies in the 2013-2015 period," Forum Ekon., vol. 21, no. 2, pp. 154-164, 2019, [Online]. Available: http://journal.feb.unmul.ac.id/index.php/FORUMEKONOMI.
3. [R. Novendri, R. Andreswari, and ..., "Implementation of Data Mining to Predict Customer Churn Using Naive Bayes Algorithm," eProceedings ..., vol. 8, no. 2, pp. 2762-2773, 2021, [Online].

Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/146 78/14455 sdfdsfdsf.

1. N. L. Widyastuti and H. Nugroho, "The Impact of Covid-19 on the Oil and Gas Industry:

Policy Recommendations for Indonesia," J. Perenc. Pembang. Indonesia. J. Dev. Plan., vol. 4, no. 2, pp. 166-176, 2020, doi: 10.36574/jpp.v4i2.116dfdsfdsfdsf.

1. E. K. Tarigan, "Juridical analysis of retail fuel oil sales according to the oil and gas law (law number 22 of 2001)," J. Lex Justitia, vol. 2, no. 2, pp. 121-134, 2021, [Online]. Available: http://ejournal.potensi-utama.ac.id/ojs/index.php/LexJustitia/article/view/1347.
2. A. Munawar et al., "Cluster Application with K-Means Algorithm on the Population of Trade and Accommodation Facilities in Indonesia," J. Phys. Conf. Ser., vol. 1933, no. 1, 2021, doi:

10.1088/1742-6596/1933/1/012027S.

1. F. Rahman, I. I. Ridho, M. Muflih, S. Pratama, M. R. Raharjo, and A. P. Windarto,

"Application of Data Mining Technique using K-Medoids in the Case of Export of Crude Petroleum Materials to the Destination Country," IOP Conf. Ser. Mater. Sci. Eng., vol. 835, no. 1, 2020, doi: 10.1088/1757-899X/835/1/012058.

1. J. R. Riwukore, L. Marnisah, and F. Habaora, “Employee Performance Analysis Based on the Effect of Discipline, Motivation, and Organizational Commitment at the Regional Secretariat of the Kupang City Government,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 12, no. 1, p. 76, 2022, doi: 10.30588/jmp.v12i1.1009.
2. P. A. Sampurna and T. Miranti, “The Effect of Service Quality, Banking Digitalization, and Customer Relationship Management (CRM) on Customer Loyalty,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 12, no. 1, p. 303, 2022, doi: 10.30588/jmp.v12i1.1138.
3. J. F. Sofyan and M. Rianty, “Karakteristik Manajemen dan Kepemimpinan Transformasional sebagai Penentu Kreativitas Karyawan yang Dimediasi oleh Kepuasaan Kerja,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 12, no. 2, p. 448, 2023, doi: 10.30588/jmp.v12i2.1186.
4. F. P. Ferdy Pangestu, N. Y. Nur Yasin, R. C. Ronald Chistover Hasugian, and Y. Yunita, “Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat,” J. Sisfokom (Sistem Inf. dan Komputer), vol. 12, no. 1, pp. 53–62, 2023, doi: 10.32736/sisfokom.v12i1.1461.
5. A. A. Syahidah and M. F. Aransyah, “Pengaruh E-Service Quality dan E-Trust Terhadap E-Customer Loyalty Pada Pengguna Dompet Digital DANA Melalui E-Satisfaction Sebagai Variabel Intervening,” J. Sisfokom (Sistem Inf. dan Komputer), vol. 12, no. 1, pp. 36–44, 2023, doi: 10.32736/sisfokom.v12i1.1593.
6. A. Rosyida and T. Bayu Sasongko, “Deteksi Dini Penyakit Alzheimer dengan Algoritma C4.5 Berbasis BPSO (Binary Particle Swarm Optimization),” J. SISFOKOM (Sistem Inf. dan Komputer), vol. 12, pp. 341–349, 2023.
7. D. E. Sondakh, R. C. Maringka, F. P. Ayorbaba, J. S. C. B. T. Mangi, and S. R. Pungus, “Emotion Mining Review Pengguna Aplikasi Mobile Banking BRImo Menggunakan Algoritma Decision Tree,” vol. 12, pp. 350–355, 2023.
8. F. Warda, F. N. Fajri, and A. Tholib, “Classification of Final Project Titles Using Bidirectional Long Short Term Memory at the Faculty of Engineering Nurul Jadid University,” J. Sisfokom (Sistem Inf. dan Komputer), vol. 12, no. 3, pp. 356–362, 2023, doi: 10.32736/sisfokom.v12i3.1723.
9. R. Vannya and A. Hermawan, “Analisis Performa klasifikasi Kesegaran Daging Ayam menggunakan Naïve Bayes , Decision Tree , dan,” vol. 12, pp. 394–400, 2023.
10. M. W. Sari, N. N. Deswira, and A. Risdwiyanto, “Determinasi Kepuasan Kerja dan Implikasinya terhadap Turnover Intention: Studi pada PT Hayati Pratama Mandiri Padang,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 12, no. 2, p. 346, 2023, doi: 10.30588/jmp.v12i2.430.
11. E. Fatmasari and B. S. Dwiyanto, “Analisis Kinerja Keuangan dengan Metode Economic Value-Added pada Studi Kasus Perusahaan Subsektor Pertambangan Minyak dan Gas Bumi yang Terdaftar di Indeks Saham Syariah Indonesia (ISSI),” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 9, no. 1, p. 17, 2019, doi: 10.30588/jmp.v9i1.435.
12. K. S. Utami, “Green consumers behavior: Consumer behavior in purchasing environmentally friendly products,” Manag. Coop. Entrep., vol. 9, no. 2, pp. 208–223, 2020, [Online]. Available: <http://ejournal.up45.ac.id/index.php/maksipreneur/article/download/499/526>
13. R. B. Prakarsa, W. Yadiati, and N. R. H. Suciati, “Pengaruh Risk Profile, Good Corporate Governance, Earning, Capital terhadap Value of Firm di Bursa Efek Indonesia,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 9, no. 2, p. 137, 2020, doi: 10.30588/jmp.v9i2.530.
14. I. L. Natapermana, W. Yadiati, and E. Nurhayati, “Pengaruh Implementasi Good Corporate Governance dan Strategi Bisnis terhadap Kinerja Perusahaan: Studi Kasus BUMN di Indonesia Tahun 2013-2018,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 9, no. 2, p. 153, 2020, doi: 10.30588/jmp.v9i2.579.
15. A. Junaidi, “Sisfokom-Maret2019-604-1539-1-PB,” vol. 08, pp. 61–67, 2019.
16. W. Marantika and S. Sarsono, “Pengaruh Kualitas Produk, Word of Mouth, dan Store Image terhadap Keputusan Pembelian: Studi pada Pengunjung Toko Amigo Pedan,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 10, no. 1, p. 114, 2020, doi: 10.30588/jmp.v10i1.633.
17. A. Khomsiyah and S. Sanaji, “Pengaruh Loyalitas dan Fanatisme Supporter pada Klub terhadap Keputusan Pembelian Merchandise Orisinal: Studi pada Supporter Persela Lamongan,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 10, no. 2, p. 242, 2021, doi: 10.30588/jmp.v10i2.756.
18. D. M. D. Prama Yanti and I. G. Sanica, “Menelisik Pengelolaan Human Capital di Dunia Bisnis dalam Era New Normal: Studi Kasus pada Generasi Milenial di Bali,” J. Maksipreneur Manajemen, Koperasi, dan Entrep., vol. 11, no. 1, p. 122, 2021, doi: 10.30588/jmp.v11i1.840.
19. M. A. S. Arifin, H. Oktafia, and L. Wijaya, “Deteksi Botnet IoT Menggunakan Autoencoder dan Decision Tree,” vol. 12, pp. 329–334, 2023.