A Comprehensive Integration of RFM Analysis, Cluster Analysis, and Classification for B2B Customer Relationship Management

Ismet Karacan and Inanc Erdogan
AN-EL Anahtar ve Elektrikli Ev Aletleri San. A.S.
Pendik, 34896, Turkey
anelprs142@an-el.com.tr, anel@an-el.com.tr

Ufuk Cebeci
Department of Industrial Engineering
Istanbul Technical University
Macka, 34367, Turkey
cebeciu@itu.edu.tr

Abstract

Customer Relationship Management (CRM) becomes more important day by day as a result of technological developments. Besides improving their internal processes for customer satisfaction, companies especially with a high number of customers seek intelligent and well-designed methods for handling their customers by analyzing their behaviors and revealing their patterns. Companies transform their repetitive processes to Robotic Process Automation (RPA) by integrating systems based on intelligence, data mining, and machine learning techniques to save time, reveal patterns in data more easily, build robust interrelationships within the functions of the company, etc. As a part of these efforts, this study proposes an integrated CRM methodology to manage both Business to Business (B2B) and Business to Customer (B2C) relations. In the integrated approach, Recency, Frequency, and Monetary (RFM) Analysis divides the customers into segments that define their current status, DBSCAN Cluster Analysis verifies RFM Analysis and reveals the anomalies in the data fur further analysis, and Decision Tree Classification Algorithm predicts company-specific custom questions to ease the work of analysts. A Case Study on the data from AN-EL A.S. (an electrical components company) verifies that the methodology produces segments of the customers and predictive responses to custom questions to enable the analysts to carry out their actions easily. According to the results of the case study, 34 % of the customers should be contacted to detail the status of the business in terms of sample approval, new purchase order, arrangement of batch size etc.

Keywords
CRM, Customer Segmentation, RFM, Cluster Analysis, Classification, B2B

1. Introduction

The share of E-commerce within total global retail sales is continuously increasing and expected to be nearly 20 % where it has been 7.4 % in 2015 (Statistica, 2021). In conjunction with many other aspects like technological developments and advances in logistic capabilities, E-commerce has changed the end customer behavior from a classical store purchase to an elective and comparative purchase strategy. Today, customers can easily search for and evaluate many alternatives with the help of Artificial Intelligence (AI) Supported Engines on the internet. Quality alone is no longer a sufficient criterion to satisfy customer demands. Retailers should provide many alternatives to be compared and in case of purchase, the product should be delivered at reasonable times. With a bullwhip effect, all stakeholders of the supply chains have been widely affected by this behavioral change. The Make-to-Stock (MTS) policy has widely lost its dominance on the Make-to-Order (MTO) policy that caused the adaptation of many production plants to Just in Time (JIT) production independent from being Business to Customer (B2C) or Business to Business (B2B). As a holistic result of this progress, the focus of the producers shifted from product to customer. Customer Relationship Management (CRM) is the generally accepted term to define the combination of people, technology, and process that make effort to reveal the properties of customers (Chen and Popovich, 2003). It has become very crucial to understand customers and satisfy their demands by analysis of datasets with well-defined attributes aiming to achieve rational outputs. Today, technological improvements help the suppliers and service providers to analyze their customers more detailly owing to novel applications and methodologies. Machine Learning-based methods ease the process for more robust customer analysis. Moreover, recent applications refer to deep learning algorithms to handle big datasets full of tables with numerical and categorical attributes.

The study has been conducted in and supported by AN-EL A.S., a B2B electromechanical component manufacturer for White Goods Industry. The firm was established in 1972 in Turkey aiming to produce rotary switches for ovens. Today, the firm produces over 45 million components annually as indicators, rotary switches, rocker switches, slide switches, safety switches, and plastic parts. Nearly 55 % of the products are exported directly to plants over 5 continents where the export ratio increases to 85 % by adding exports by White Good producers in Turkey. Since many customers of the firm are global manufacturers and competitors are also global manufacturers, AN-EL A.S. has to carry out CRM with robust methodologies to be able to analyze any critical information hinting for actions to be taken. In this study, we drive a concept methodology for robust customer analysis and adapt it to the framework of AN-EL A.S. CRM strategy. The originality of this study is on the logic of classifying the data. Rather than providing the conditional type of the customer, the methodology responds to company-specific questions for making the actions as clear as possible for the analyst.

1.1. Objectives

This study aims to propose a comprehensive methodology to evaluate current customer behaviors and to reveal actions to be taken in terms of CRM. It has been intended to support the methodology with machine learning algorithms so that all key inferences and results may be deducted by a user-friendly flow. The proposed methodology turns out segmentation of the customers concerning their key parameters and responses two critical questions that are customized according to company-specific requirements. The results should bring the analysts detailed information about customers in respect of membership to segmentation clusters and clear replies to company-specific questions so that each customer can be handled with intense care.

2. Literature Review

CRM is a wide area of practice to manage existing customers, new customers, or customer candidates. Many experimental, analytical, or numerical methodologies have been suggested so far trying to seek responses to defined research questions. This study focuses on the segmentation of customers and the prediction of research customers. Therefore, the literature of relevant methodologies has been reviewed in this section.
Segmentation is the process of separating customers into different and homogenous clusters according to their characteristics so that driving marketing strategies get easier. (Tsiptsis and Chorianopoulos, 2011). Segmentation is a systematic approach to help organizations reveal key features of their customers. Thus, companies can drive distinct policies for targeted groups to improve customer satisfaction. Aside from many proposed techniques, Recency, Frequency, and Monetary (RFM) Analysis and Cluster Analysis are highly encountered methods in the literature.

With its simplicity, The Recency, Frequency, and Monetary (RFM) analysis that was proposed in the 1920s by a catalog company (Petrison et al., 1997) helps to reveal the posture of the customer with 3 critical indicators. Recency indicates the novelty of customer activities. Frequency is an indicator for showing the repetition of the activities in a certain period. Monetary refers to the total budget volume of the customer in terms of purchasing. Customers that reach high scores in all three variables may be identified as loyal customers (Buttle, F., and Maklan, 2019). Wu and Lin (2005) pointed out that the customers that have high values for both R and F have the potential to produce new business with the company. Moreover, the bigger M is, the more likely the corresponding customers are to buy products or services with enterprises again. Score-based classification tables have been widely proposed for RFM to get more analytical approaches. A two-attribute table with R-F and M score columns may segment customers may divide customers into 11 segments with names “Champion”, “Loyal”, “At Risk”, etc (Kabasakal, 2020). In the literature, RFM has been proposed mainly hybrid or integrated with additional methods to achieve more robust analysis.

Cluster analysis is an unsupervised learning technique with a broad application area to divide entities into groups concerning their attributes. Many K-means is a center-based algorithm where intercluster distances are maximized while intracluster distances are minimized (Hamerly and Elkan, 2004). However, the K-means algorithm suffices to detect outliers since it is a centroid-based algorithm. DBSCAN has been suggested by Ester et al. (1996) as a novel clustering method. With its mentality of iteration, DBSCAN is a satisfactory alternative to identify outlier detection.
Classification is a machine learning method that tries to predict the target variable considering the provided features. Within the entire, Decision Tree Classifiers cover a large portion. The common logic behind Decision Trees is to create a hierarchical tree structure about data to identify the relations that sources the value of a target variable. Among many proposed trees, C4.5 (Quinlan, 1993), Classification and Regression Tree (CART) (Breiman et al., 1984), and Chi-square automatic interaction detection (CHAID) (Kass, 1980) are notable alternatives.

Many CRM-related methodologies with hybrid or integrated structures have been suggested to date. McCharty and Hastak (2007) worked on two datasets to compare RFM, CHAID, and logistic regression methods for marketing segmentation. Cheng and Chen (2009) proposed an integrated model that delivers the output of RFM analysis to the K-means algorithm as input and finalizes the process with rough set (RS) theory. Christy et al. (2018) completed a segmentation study on transactional data by calculating three components of RFM analysis, feeding them to three clustering algorithms, namely K-means, Fuzzy C-means algorithms, Repetitive Median based K-Means, and analyzing the algorithms according to their runtime parameters. Anitha and Patil (2019) studied the real-time transactional and retail dataset by using the K-Means algorithm in cooperation with RFM Analysis. They have analyzed Silhouette Score and RFM at the final step of their methodology. You et al. (2015) proposed a decision-making framework for precision marketing to identify the critical characteristics of different customer categories and update marketing strategies by integrating the RFM Model, CHAID decision trees. The RFM model is used to predict the supply quantity per month by clustering the customers using the K-Means algorithm in their study. Each group is divided using CHAID decision trees based on attribute values. Maryani and Riana (2017) proposed another model that can determine potential customers and suggest necessary action in terms of marketing strategy. They have worked on transactions by creating RFM data, clustering with the K-means algorithm for grouping and profiling customers by a tree model. Bahari and Elayidom (2015) contributed a different framework that helps to predict further customer behaviors. They have analyzed 17 campaigns of a bank to reveal customer intentions. To achieve this, they have integrated multiple tools like Multilayer Perception Neural Network and Naive Bayes.

3. Methods

This study focuses on developing a user-friendly framework to enable users to segment their customers easily by key indicators and create responses to key questions regarding the actions to be taken. To take advantage of multiple algorithms, an integrated methodology of RFM, K-Means Cluster Analysis, and Decision Tree Classification is proposed. Figure-1 demonstrates the complete flow of the proposed methodology and can be explained with 4 steps.

Step 1: Data Preprocessing. Data preprocessing is required to be clean and prepare data for further analysis. Missing customer information should be completed or removed from the data.

Step 2: RFM Analysis: Data should be analyzed with RFM Analysis to segment customers. The RFM scores are utilized to profile customers.

Step 3: Cluster Analysis: Clustering should be conducted on the dataset which is enforced by outputs of RFM Analysis. The purpose of Cluster Analysis is twofold. First, it should be used to compare the results of RFM Analysis. Second, it can be used to determine outlying customers. To identify the outliers easily, the DBSCAN algorithm is preferred within this study.

Step 4: Decision Tree Classification: Custom classification is conducted by the Decision Tree Classifier Algorithm on the data that has additional attributes for classification. The algorithm should be run n times where n is the number of total target variables.

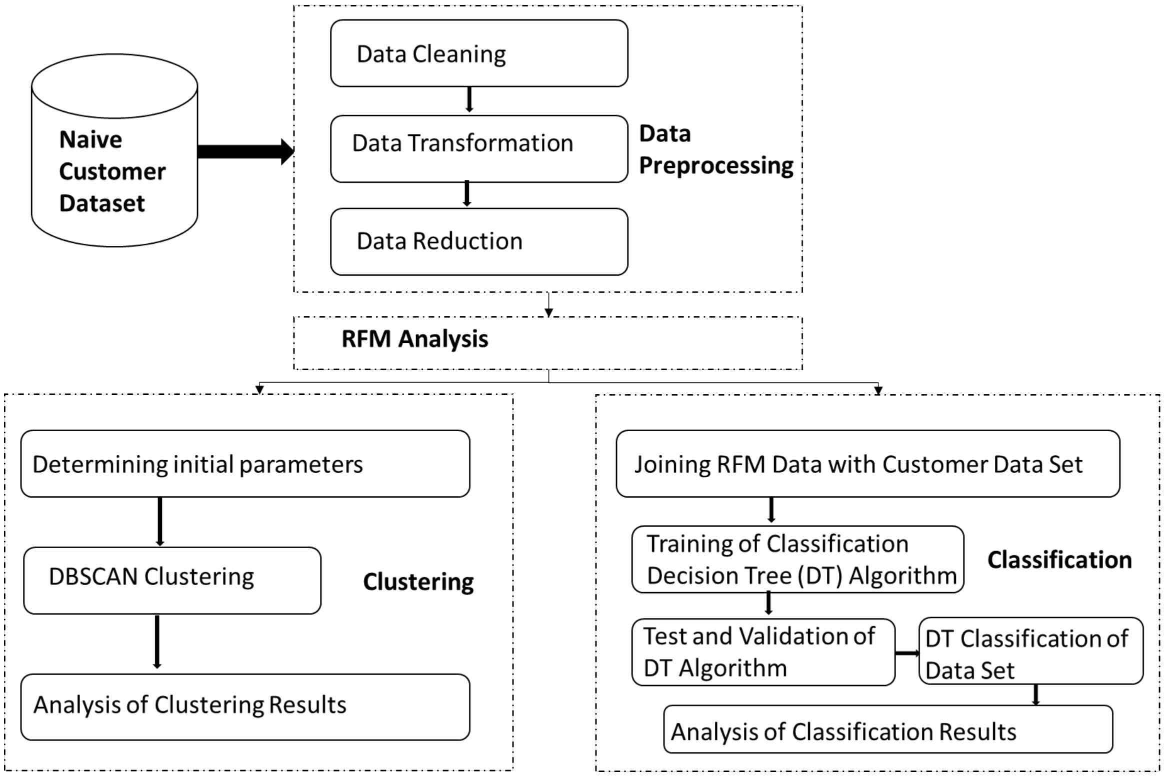

Figure 1. The Flowchart of the Proposed Methodology

3.1 RFM Analysis

The methodology consists of 3 analysis steps. The first analysis can be conducted on the traditional RFM method. R, F, and M values should be calculated before analysis (Nimbalkar and Shah, 2013). R is the time passed from the last activity of the customer. F is the number of activities by the customer. M is the total wealth spent by the customer. After calculation of crisp values, R, F, and M columns should be divided into a defined number of intervals so that the customers can be scored. Usually, the number of segments for each column is preferred as 4 or 5 in the literature. In this study, 5 segments are preferred for a more detailed resolution. The following steps define the RFM Analysis process:

Step-1: Calculate the R-value of each customer by calculating the number of days passed since the last activity.

Step-2: Calculate the F-value by counting the number of activities.

Step-3: Calculate the M-value by summing the costs paid by the customer.

Step-4: Sort the customers according to R values in ascending order. Score each 20 % segment of customers from 5 to 1 beginning from top to down.

Step-5: Sort the customers according to F values in descending order. Score each 20 % segment of customers from 5 to 1 beginning from top to down.

Step-6: Sort the customers according to M values in descending order. Score each 20 % segment of customers from 5 to 1 beginning from top to down.

Step-7: Join each score together to determine the RFM Score.

Step-8 Prepare segmented data according to the defined scale.

RFM results may be analyzed in different ways. RFM scores may be sorted or customers may be divided into groups to discover customer status. In this study, we preferred to apply the scale by Clevertap (2021) since it provides a rational output chart to be examined.

Table 1. RFM Analysis Scale Concerning Recency and Frequency Scores

Cluster	Recency (R) Score	Frequency (F) Score
Champions	5	4-5
Loyal customers	3-4	4-5
Potential loyalists	4-5	2-3
Promising	4	1
Can't lose them	1-2	5
At risk	1-2	3-4
About to sleep	3	1-2
Hibernating	1-2	1-2
New customers	5	1
Need attention	3	3

3.2 Cluster Analysis – DBSCAN

RFM analysis has been preferred for many cases due to its simplicity and accessibility. However, it has shortcomings to interpret the conditional practices of customers completely. For instance, the behavior of a customer that has an R-value of 30 days should be considered suspicious where the average order time is 2 days. However, 30 days rises no alert for a high-volume buyer that prefers to buy every 3 months. Another deficiency is the logic of segmentation. The customers are segmented into the same number of groups. However, there may be cases, in which for instance % 10 of customers have an R-Score of nearly 3 days, and the remaining customers' values are nearly 20 days. In this case, R scoring cannot make sense since every customer except %20 of the total with a recency value of nearly 20 has an R-Score 2 to 5 where % 10 of them get the R-Score 1 with the customers that have recency of 3 days.

In this study, Cluster Analysis is used both to avoid the shortcomings of this segmentation and detecting outliers of the data that need to be evaluated with further attention. Cluster analysis with additional columns is included in the methodology to better understand the customer status. Rather than using the RFM scores, RFM values themselves are included in the dataset for Cluster Analysis to avoid any bias. The attributes of Cluster Analysis are shown in Table 2.

Table 2. Attributes and Their Definitions for Cluster Analysis

Attribute	Definition
Recency	The time passed from the last order of the customer
Frequency	Total number of orders by the customer
Monetary	Total earnings from the customer
$\frac{\text{Monetary}}{\text{Frequency}}$	Ratio showing the average batch value
$\frac{\text{Recency}}{\text{Average Order Time}}$	Ratio showing earliness/lateness of order by the customer (Over 1 shows a late order, below 1 shows an early order)

To be able to both evaluate the routine customers and the outliers, DBSCAN has been preferred for clustering. The algorithm requires 2 parameters in the initialization phase. The first parameter is ε that determines the radius of the neighborhood distance, the second parameter is the minPts that shows the minimum number of entities forming a dense region.

3.3 Decision Tree Classification

Although providing an insight, segmentation alone cannot respond to the company-specific questions. RFM hosts basic indicators for a customer. Besides having more clues about customers, Cluster Analysis requires a more complex process and knowledge about the algorithm to analyze data since it is an unsupervised learning method. Companies seek replies to many questions from their own perspective. Has their customers' business volumes decreased? Is there a decrease in the product variety of the customers? Who are the customers that will gain or lose potential? Are the samples sent to the companies turned to the order? Do I need a change in the order batch sizes or order frequencies of my customers? Are several questions that may be required to be responded for a company according to its CRM strategy.

For even a more powerful CRM capability, the proposed methodology includes a classification phase in which customers may reveal responses to their special questions on customers. Comparing the accuracy of Decision Tree, Support Vector Machine (SVM), and Random Forest classification algorithms during the case study, Decision Tree Algorithm has been preferred to be included for this study. The Decision Tree classifier of the Orange Data Mining Tool developed at Bioinformatics Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Slovenia, together with the open-source community has been applied to predict customized target variables. The flow classification algorithm has been designed to respond to company-specific questions. The questions help the analyst to clearly understand the actions to be taken concerning the state of the customer. Five target variables are defined in Table 3. They have been predicted in distinct runs of the classification algorithm.

Table 3. Target Variables and Their Explanations.

Target Variable	Definition	Possible Values
Business Volume	Indicator for the fiscal change in the business volume	Increased: The volume is increased Decreased: The volume is increased Same: The volume stayed the same New Customer: Since this is a new customer, unable to detect the trend No Order – Only Sample: No paid order so far Losing: The customer is about to be lost. Special action should be taken to convince for continuity. Lost: The customer is lost. Should be analyzed in detail.
State of Orders	Indicator for the requirement of a new order from a customer	Normal: Stays within the time limit for a new order Could be asked: A short time passed after the expected order time. A new order could be asked. Should be asked immediately: A long time passed after the expected order time. A new order should be asked immediately. Not necessary: Can't get a response or customer
State of Samples	Indicator for tracking the samples	NA: No samples have been sent so far Normal: All required feedback regarding samples have been gathered Should be asked: Pending feedback regarding the delivered samples should be asked to the customer. Care before sending samples: No feedback could be gathered from the customer for a threshold time. Further samples should be sent with care
Batch Value	Indicator of the actions to be taken on batch value/size	NA: Not a regular buyer Normal: Batch sizes are in normal value limits. Good: Batch sizes are in preferred value limits. Should be increased: Customer should be encouraged to increase the value of delivered batches.
Product Variety	Indicator for the change in the number of product types	NA: Customer requested only samples so far. Same: The number of types stayed at the same level Increased: The variety has increased. Decreased: The variety has decreased. Lost: Customer buys no products New customer: Since this is a new customer, unable to detect the trend

Additional columns have been added to the data to increase the accuracy of classification. The columns of final data are summarized in Table 4.

Table 4. Final Metadata for Classification Algorithm.

Feature	Definition
Recency	The time passed from the last order of the customer
Frequency	Total number of orders by the customer
Monetary	Total earnings from the customer
Monetary/Frequency	Ratio showing the average batch value
Recency/Average Order Time	Ratio showing earliness/lateness of order by the customer (Over 1 shows a late order, below 1 shows an early order)
Turnover in the Last 12 Months	Turnover from the sales within 12 months
Turnover between the Last X-1th and Xth Years	Turnover from the sales between last X-1th and Xth years. The feature may be reproduced according to the available data.
Yearly Average Turnover	Average of yearly turnovers
Turnover Change in the Last Year (Percentage)	Change of last 12 months turnover in percentage compared to the former year
Cumulative Percentage of Change of Turnover	Sum of yearly turnover percentage changes
# of Samples in the Last X months	# of samples that have been delivered to the customer in the last X months. The threshold X to be determined by the company.
# of Samples responded in the Last X months	# of samples that have been responded to by the customer in the last X months
# of Product Types in the Last 12 Months	# of products types bought by the customer within 12 months
# of Products between the Last X-1th and Xth Years	# of product types between last X-1th and Xth years. The feature may be reproduced according to the available data.
Yearly Average # of Products	The yearly average of # of product types
# of Products Change in the Last Year (Percentage)	Change of # of product types in the last 12 months in percentage compared to the former year
Cumulative Percentage of Change of # of Products	Sum of yearly percentage changes concerning # of products

4. Results and Discussion

As a case study, the customers of AN-EL A.S. have been analyzed the actions to be taken are considered. Being a global supplier, the company should manage a customer portfolio with many different behaviors. Considering the customers' distinct behaviors, the dataset by AN-EL A.S. is a satisfying sample for the case study.

Step 1: Data has been preprocessed and prepared for further analysis. The test data consists of 368 active and potential customers.

Step 2: Considering the scale in Table 1, the data has been divided into 11 segments for RFM analysis. The chart in Figure 2 expresses the percentage of customers included in each RFM segment. Although giving an insight, the segments lack to detail individual customer behaviors. It can be suggested that besides a high percentage of liable customers, another high percentage of the customer should be communicated delicately to pull them among loyal customers. 8,5 % of the customers are waiting for further action to be included in the business ecosystem. 7,2 % of them seem to be a risk to be lost and the underlying reasons should be discovered. % 2.9 frequent buyer did not make an order for a long time that means they have probably diverted some of their orders to competitor suppliers. With 16,8 %, the champions form the backbone of the business of AN-EL that make them the most valuable customer. Loyal customers with 19 % are also as valuable as champions, however should be checked for recency. A group with 18,6 % are the candidates for being a loyal customer if they increase their frequency. 5,9 % of the customers should be interested in more intense together with promising 4,2 % customers since they are newcomers. 7,6 % of the customers are about to be lost unless necessary precautions are taken. Another 9,3 % of the customers in the center of the chart waiting to be analyzed to expose their slope.

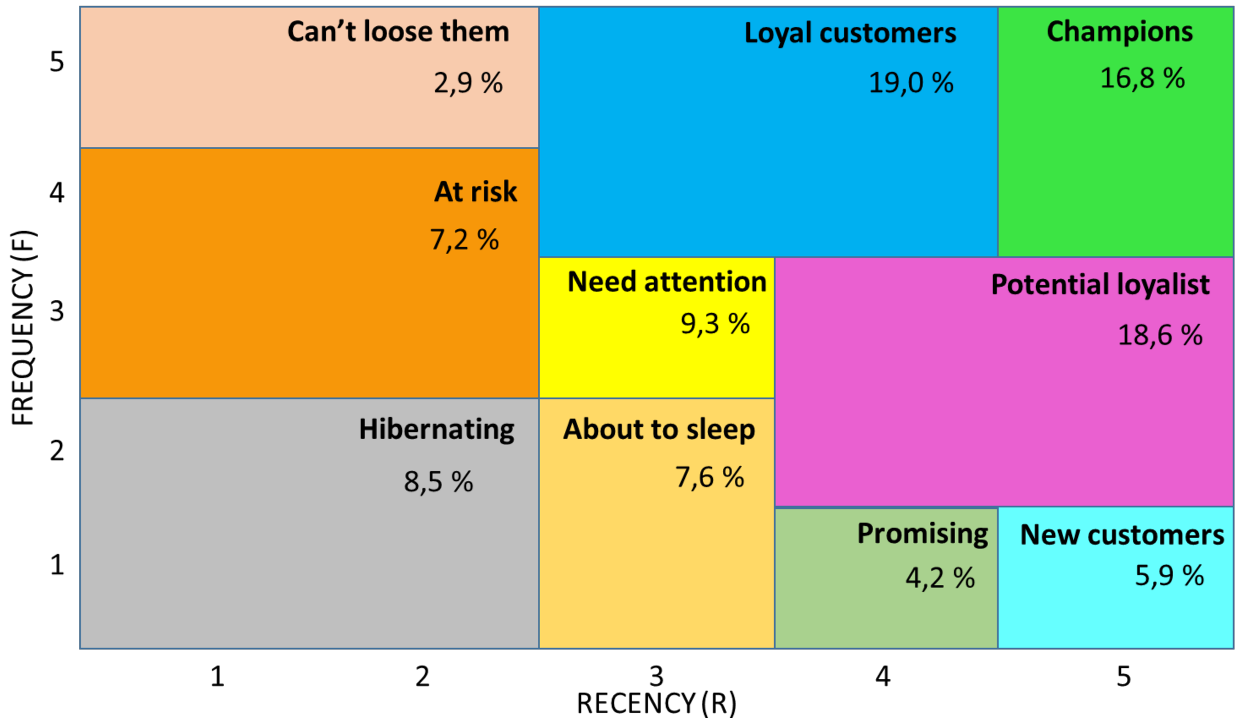

Figure 2. Percentages of RFM Segments

Step 3: DBSCAN Cluster Analysis is especially carried out to detect outliers. Following the intention of detecting outliers, the ε value set to 0,87 where minPts value is set to 3. The algorithm discovered 13 outliers in the dataset. A sample outlier customer revealed by the algorithm is detailed in scatter plots in Figure 3 and Figure 4. The behavior of the customer in the orange circle in both graphs shows that the customer has an extremely high business and ordered recent demands. However, the values of the individual orders are rather low, and a high number of orders have been received in the period. Due to operational difficulties, the customer should be suggested to order some batches together.

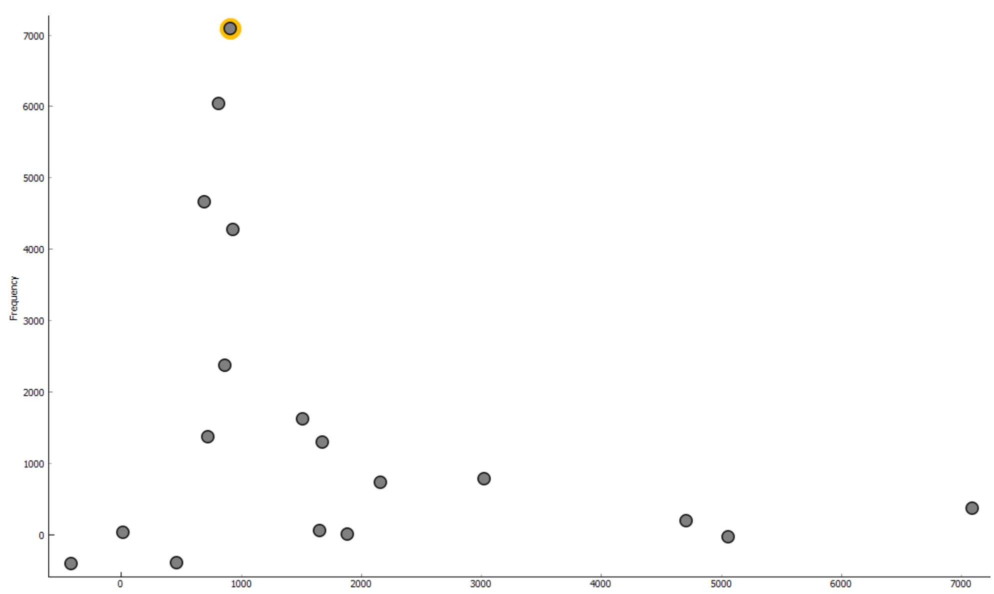

Figure 3. Monetary/Frequency vs. Frequency Scatter Plot

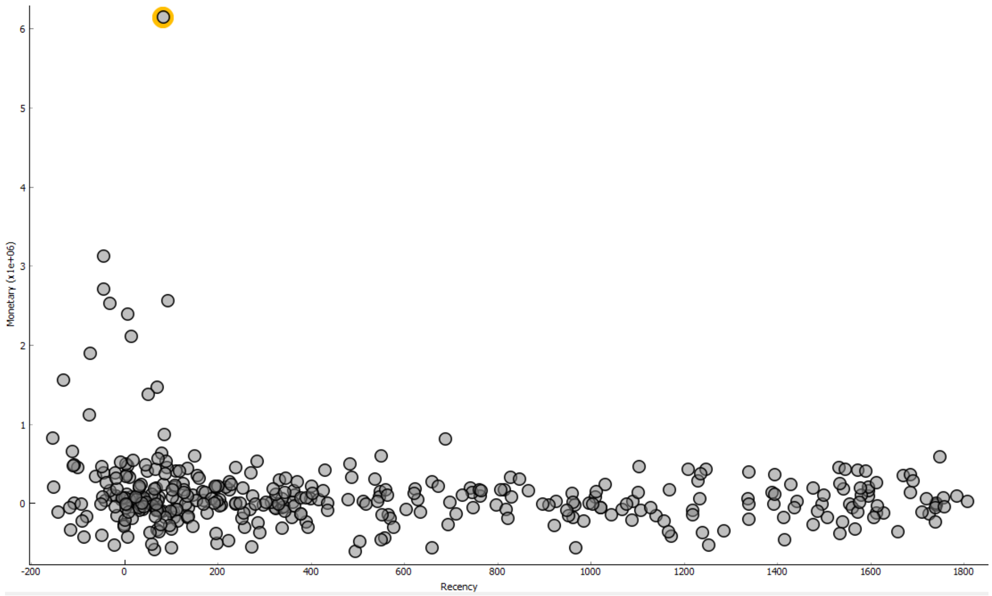

Figure 4. Recency vs. Monetary Scatter Plot

Step 4: The Decision Tree Classification Algorithm should be run five times to disclose custom questions. The outputs of the algorithm are five different decision trees and classification tables. Figure 5 represents a sample tree for classifying the target variable "Status of Order".

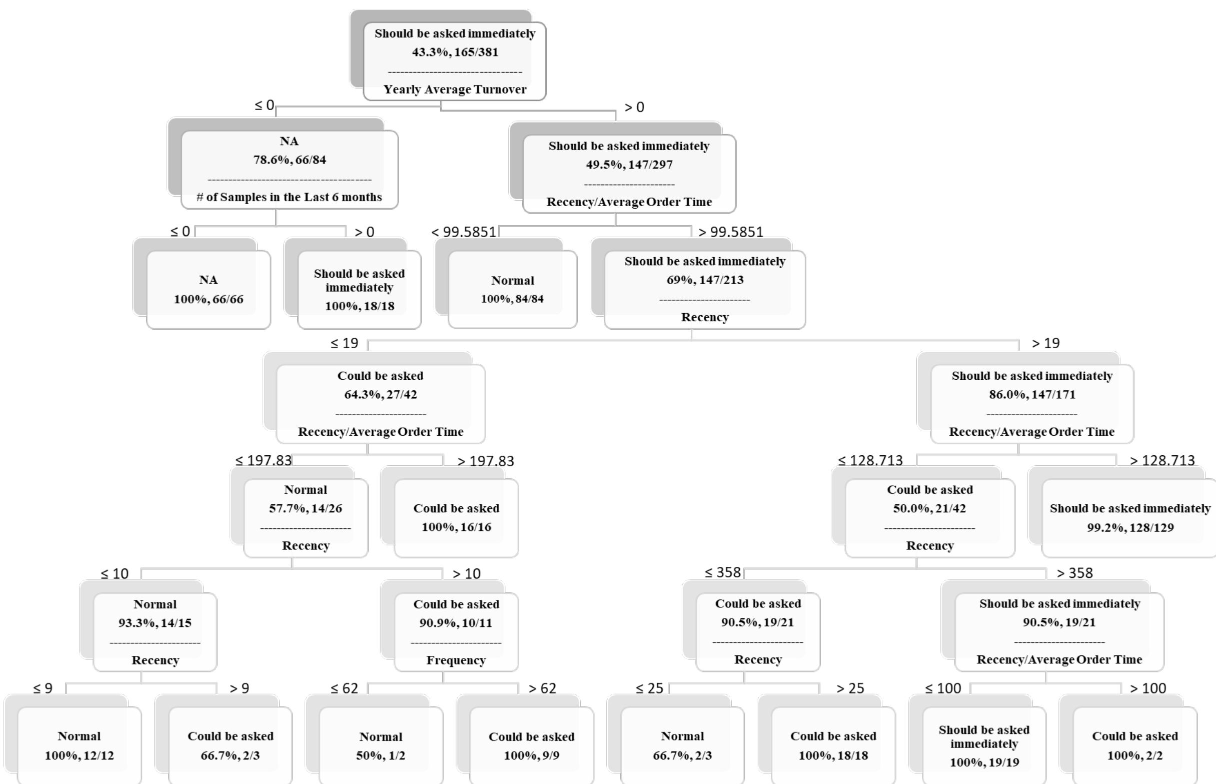

Figure 5. Decision Tree for Target Variable "Status of Order"

The algorithm predicts the target variable considering several features in the data. The algorithm has been run five times to respond to the custom questions. Accuracy alone stays insufficient to indicate the performance of a classification algorithm. The classifier should be evaluated considering its capability to identify True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions. We have compared the performances according to the measures in Table 5.

Table 5. Comparison of Classification Algorithms

Measure	**Definition**
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1	$2 * \frac{Precision * Recall}{Precision + Recall}$

According to the performance measures in Table 5, three different classification algorithms have been compared. Results of the classifications concerning Decision Tree, Random Forest, and SVM algorithms have been shared in Table 6. After training with 600 data points, the algorithms have been tested on 368 data points. Cumulative results suggest that the Decision Tree Classifier performs best in all five prediction processes.

Table 6. Performance Measures of Classification Algorithms on Test Data

Target Variable	**Algorithm**	**Precision**	**Recall**	**F1**
Business Volume	Decision Tree	**0.989**	**0.992**	**0.990**
	Random Forest	0.978	0.981	0.979
	SVM	0.809	0.789	0.799
State of Orders	Decision Tree	**0.995**	**0.995**	**0.995**
	Random Forest	0.967	0.967	0.967
	SVM	0.849	0.851	0.850
State of Samples	Decision Tree	**1.000**	**1.000**	**1.000**
	Random Forest	0.989	0.989	0.989
	SVM	0.954	0.954	0.954
Batch Value	Decision Tree	**1.000**	**1.000**	**1.000**
	Random Forest	0.978	0.978	0.978
	SVM	0.966	0.962	0.964
Product Variety	Decision Tree	**1.000**	**1.000**	**1.000**
	Random Forest	0.959	0.948	0.953
	SVM	0.874	0.842	0.858

After completing all steps, the results are aggregated to determine the detailed course of action. To verify the model, aggregated results are evaluated manually for consistency. A 5-sample table is introduced in Table 7. Besides acquiring the segment of the customer by RFM Analysis, the analyst can easily consider the action by checking the predictions by the classification algorithm. Since CRM is a sensitive process, the accuracy of the results is highly important to avoid any improper activity with the customer. With well-defined training data, the Decision Tree performs a high accuracy on all five target variables. As a result, the membership data by RFM analysis supported by adding additional responses to research questions. For instance, besides knowing that Customer 5 is a promising customer, the analyst can easily consider that the customer should be asked for feedback on newly delivered samples. As another example, besides being a potential loyalist, Customer 1 should be encouraged to increase the size of its orders since the low batch sizes create additional load on the company's production process. Although having a good trend as a loyal customer, the product portfolio of Customer 4 has been decreased. Possible reasons may be obsolescence of some products or change of supplier. In both cases, the customer should be offered to work on new product projects. This analysis capability of the methodology is the distinguishing part of the study.

Table 7. Sample Table For Aggregated Results

Customer	Segmentation	Business Volume	State of Orders	State of Samples	Batch Value	Product Variety
Customer 1	Potential Loyalist	Increased	Normal	Should be asked	Should be increased	Increased
Customer 2	About to sleep	Losing	Should be asked immediately	Should be asked	Should be increased	Same
Customer 3	New customer	New customer	Normal	Normal	Normal	New customer
Customer 4	Loyal Customers	Increased	Normal	Normal	Good	Decreased
Customer 5	Promising	No order – Only sample	Could be asked	Should be asked	NA	NA

5. Conclusions and Future Research

In this study, we have proposed a comprehensive methodology for customer segmentation and revealing responses to company-specific questions concerning customer patterns. RFM Analysis provides plain segmentation information that constitutes an insight regarding the membership status of the customers. Without any further detail, the customer may be grouped in terms of status. DBSCAN Cluster analysis improves the accuracy of RFM analysis and detects outliers in the dataset. Outliers are unordinary customers that behave distinctly compared to remaining customers. Distinct behaviors of outliers should be analyzed with further efforts to better understand the customers. At the final stage, the Decision Tree Classification Algorithm eases the work of analysts by revealing replies to custom questions. These specific questions clearly state the actions to be taken concerning each customer. The novelty of this study is providing a holistic approach to B2B and B2C CRM management by a methodology of segmentation, outlier detection, and responding to customer-related custom questions. Revealing all these data may be possible with intense manual analysis. However, this process would consume a meaningful time every time the analysis is carried out. The methodology can currently provide membership and classification data considering the features of the study. CRM is widely applied to B2C and not common for B2B. The study also proves that CRM may be applied easily to B2B analysis. For future study, the methodology may be enhanced by adding forecasting techniques to suggest possible future actions that may be carried out by the customers.

References

Anitha, P., and Patil, M. M., RFM model for customer purchase behavior using K-Means algorithm, *Journal of King Saud University-Computer and Information Sciences*, 2019.
Bahari, T. F., and Elayidom, M. S., An efficient CRM-data mining framework for the prediction of customer behaviour, *Procedia computer science, 46,* 725-731, 2015.
Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A., Classification and regression trees. *CRC press*, 1984.
Buttle, F., and Maklan, S., Customer relationship management: concepts and technologies. Routledge., 2019.
Chen, I. J., and Popovich, K., Understanding customer relationship management (CRM), *Business process management journal,* 2003.
Cheng, C. H., and Chen, Y. S., Classifying the segmentation of customer value via RFM model and RS theory, *Expert systems with applications*, 36(3), 4176-4184, 2009.
Christy, A. J., Umamakeswari, A., Priyatharsini, L., and Neyaa, A., RFM ranking–An effective approach to customer segmentation, *Journal of King Saud University-Computer and Information Sciences*, 2018.
Clevertap, Available: https://clevertap.com/blog/rfm-analysis/, Accessed on Mar 08, 2021
Ester, M., Kriegel, H. P., Sander, J., and Xu, X., A density-based algorithm for discovering clusters in large spatial databases with noise, *In Kdd (Vol. 96, No. 34, pp. 226-231)*, 1996.
Hamerly, G., and Elkan, C., Learning the k in k-means. *Advances in neural information processing systems*, 16, 281-288., 2004.
Kabasakal, İ., Customer segmentation based on recency frequency monetary model: A case study in E-retailing, *Bilişim Teknolojileri Dergisi*, 13(1), 47-56., 2020.

Kass, G. V., An exploratory technique for investigating large quantities of categorical data*, Journal of the Royal Statistical Society: Series C (Applied Statistics), 29(2), 119-127*, 1980.

Maryani, I., and Riana, D., Clustering and profiling of customers using RFM for customer relationship management recommendations, *In 2017 5th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE,* 2017.

McCarty, J. A., and Hastak, M., Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research*, 60(6), 656-662, 2007.

Nimbalkar, D. D., and Shah, P., Data mining using RFM Analysis. *International Journal of Scientific & Engineering Research (IJSRE),* 4(12), 940-943., 2013.

Petrison, L. A., Blattberg, R. C., and Wang, P., Database marketing: Past, present, and future, *Journal of Direct Marketing, 11(4), 109-125*, 1997.

Quinlan, J. R., C4.5: Programs for Machine Learning, *Morgan Kaufmann Publishers,* 1993.

Statistica, Available: https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/, Accessed on Mar 08, 2021.

Tsiptsis, K. K., and Chorianopoulos, A., Data mining techniques in CRM: inside customer segmentation, *John Wiley & Sons.*, 2011.

Wu, J., and Lin, Z., Research on customer segmentation model by clustering, *Proceedings of the 7th international conference on Electronic commerce*, pp. 316-318, 2005.

You, Z., Si, Y. W., Zhang, D., Zeng, X., Leung, S. C., and Li, T., A decision-making framework for precision marketing, *Expert Systems with Applications*, 42(7), 3357-3367, 2015.

Biographies

Ismet Karacan graduated from the Naval Academy with a Bachelor's Degree in Industrial Engineering in 2003 and received a Master's Degree in Operations Research from the Naval Sciences and Engineering Institute in 2014. He is a Ph.D. candidate at the Marmara University. He is currently Manager of the R&D Center at AN-EL Anahtar ve Elektrikli Ev Aletleri A.S. His research interests include optimization, data mining, machine learning, scheduling, and simulation.

Inanc Erdogan graduated from the Istanbul Technical University with a Bachelor's Degree in Electrical Engineering in 1989. He is General Manager of AN-EL Anahtar ve Elektrikli Ev Aletleri A.S since 1996. His research interests include metallurgy on electrical components, project management, lean manufacturing, and data mining.

Ufuk Cebeci graduated from the Istanbul Technical University with a Bachelor's Degree in Industrial Engineering in 1983. He received a Master's Degree in 1986 and a PhD degree in 1993 from the Istanbul Technical University. He is a professor in the same university in the Istanbul Technical University, Department of Industrial Engineering. His research interests include quality systems, lean manufacturing, industry 4.0, artificial intelligence.