Applied Artificial Intelligence
An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method

Vladimir Djurisic, Ljiljana Kascelan, Suncica Rogic & Boban Melovic

To cite this article: Vladimir Djurisic, Ljiljana Kascelan, Suncica Rogic & Boban Melovic (2020) Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method, Applied Artificial Intelligence, 34:12, 941-955, DOI: 10.1080/08839514.2020.1790248

To link to this article: https://doi.org/10.1080/08839514.2020.1790248

Published online: 04 Jul 2020.

APPLIED ARTIFICIAL INTELLIGENCE
2020, VOL. 34, NO. 12, 941–955
https://doi.org/10.1080/08839514.2020.1790248

Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method

Vladimir Djurisic, Ljiljana Kascelan, Suncica Rogic, and Boban Melovic

Faculty of Economics, University of Montenegro, Podgorica, Montenegro

ABSTRACT

This paper proposes a predictive approach to segmenting credit card users, based on their value to the bank. The approach combines the Recency, Frequency and Monetary (RFM) method, clustering using the *k*-means method, and predictive classification by the Support Vector Machine (SVM) method. Clustering by non-encoded RFM attributes overcomes the subjectivity in selecting the number of segments and losing information (small differences in the values of these attributes) which are problems of classic RFM segmentation. In order to overcome the problem of class imbalance in predictive classification (which occurs due to the small number of valuable customers), the Support Vector Machine (SVM) method was applied as a pre-processor of data due to its extraordinary generalization capabilities. The end result of predictive classification should be a set of rules that describes the identified customer segments in order to tailor the offer to each segment individually. The extraction of rules from the SVM output was achieved using the Decision Tree (DT) classification method. Using a proposed approach that addresses the issue of the small class, marketing managers can more effectively target the most valuable customers, thereby increasing revenue, but also reducing unnecessary costs due to wrongly targeted valuable clients.

Introduction

In the modern Customer Relationship Management (CRM) concept, systematic analyzes of large amounts of data become a key criterion in the process of identifying and communicating with consumers (Khan 2018).

The primary stage in the development of the CRM concept is market segmentation (Kumar, Chaitanya, and Madhavan 2012). Customer segmentation is important for the banking sector and numerous studies have been done in this direction (Khajvand and Tarokh 2011; Ansari and Riasi 2016; Khobzi, Akhondzadeh-Noughabi and Minaei-Bidgoli, 2014).

The goal of CRM is to maximize Customer Value (CV) for the organization. Measuring CV is the starting point for making marketing decisions and organizing marketing activities. The earlier segmentation approach was

CONTACT Vladimir Djurisic vladodjurisic@yahoo.com Faculty of Economics, University of Montenegro, Podgorica 81000, Montenegro

based on demographic and sociological characteristics, while the modern segmentation process is based on Customer Lifetime Value (CLV), which represents the total net revenue from the customer, including future revenue as well (Mohammadi et al. 2014). However, research has shown that one cannot always predict consumer behavior on the basis of the monetary component alone.

The most commonly used method for customer segmentation is Recency, Frequency and Monetary – RFM (Hughes 1994). This method divides customers into uniform segments according to attributes regarding their purchasing behavior, based on the data recorded in the database. The RFM attributes are defined as follows: R – date of the last transaction, F – the total number of transactions in the considered period and M – the monetary amount spent in the considered period. The attributes are encoded by a standard scheme so that, according to the Pareto principle, the 20% of the most favorable values (Sanderson et al. 2010) receive a score of 5, the subsequent 20% receive a score of 4, and so on. In this way, customers are divided into 125 segments. The number of segments targeting the most valuable customers depends on the manager's subjective judgment. In this respect, the RFM model is a simple tool for analyzing consumer behavior and grouping consumers more precisely.

However, the RFM method has some drawbacks as well. Yang (2004) states that despite its simplicity, RFM can be ineffective due to segment overlap syndrome. In addition, the RFM method creates a uniform division into segments, there is subjectivity when choosing the number of segments, and also there is a loss of information in encoding RFM attributes, because the small differences in their values are lost (Rogic and Kascelan 2019).

More recently, a number of studies have applied data mining techniques i.e. clustering to customer segmentation (Zakrzewska and Murlewski, 2005; Cheng and Chen, 2009; Li et al., 2010). However, the disadvantage of this research is the subjectivity when choosing the number of clusters.

Predictive segmentation models are increasingly being used in CRM research (Cheng & Chen, 2009; Khalili-Damghani, Abdi & Abolmakarem, 2018). The task of these models is to predict a client's affiliation to a particular segment based on the client's characteristics. However, since the most valuable customer cluster is most often the smallest cluster, there is a problem of class imbalance, which leads to bias in predictive classification methods, i.e. incorrect classifications of the minor class (Kim, Chae, and Olson 2013; Miguéis, Camanho, and Borges 2017).

In order to overcome the mentioned problems of subjectivity and class imbalance, this paper proposes an approach combining the RFM method with reduced attribute coding, objective segmentation using *k*-means clustering and the predictive SVM classification method, which has been confirmed in literature to successfully balance classes (Farquad & Bose, 2012). However, the SVM classifier does not have the ability to interpret the model, which is its major drawback. Therefore, it is often used in combination with rule extraction (RE) techniques (Barakat and Bradley 2010) and, in this paper, the process of extracting rules from the SVM output was performed using the Decision Tree (DT) classification method which generates explicit if–then rules (Martens et al. 2006).

This approach allows for a significant increase in the accuracy of the segmentation model, an effective prediction of customer affiliation with the segment of the most valuable customers and the adjustment of the offer based on the segments described by explicit rules. Therefore, companies can significantly increase revenue and reduce unnecessary costs with an appropriate offer.

This paper is organized as follows. Following the introductory part, an overview of the current research in this field is presented. The third part describes the methodology used in this paper. Following the methodology, an overview of the results with discussion is given. Finally, the concluding considerations and practical implications are presented at the end of the paper.

Literature Review

Despite the fact that the process of market segmentation has been widespread in the theory and business dealings of organizations from all sectors, a methodology that provides an ideal segmentation process has not yet been developed (Dibb and Simkin 2010).

Recently, intense changes in the banking sector have imposed the need for continuous identification of consumers and analysis of their behavior. Given that CRM helps build long-term and profitable relationships with valuable consumers, its optimization is a potential source of competitive advantage. CRM is the process of maintaining profitable customer relationships based on delivered value and established loyalty (Mohammad, Hasheminejad, and Khorrami 2018). They state that the goal of CRM is to maximize consumer value (CV) for the organization, based on which consumers are classified and ranked. Through identifying customer values based on information on card usage in the previous period, banks are creating different marketing strategies to retain the most valuable customers.

In order to retain existing customers, organizations have focused on relationships with them, as well as monitoring their behavior in the past, and based the market segmentation on the use of data mining techniques (Tavakoli 2018). According to them, the RFM (Recency, Frequency, and Monetary) model is the most widespread customer grouping tool in various forms, but in order for consumer segmentation to be effective, it has to be manually adjusted on a case-by-case basis. Tavakoli (2018) have suggested customer segmentation using a modified RFM model – R + FM. Specifically, they separated R from the other two attributes, since F and M describe consumer behavior, which was the basis of their segment definition. In this regard, the authors performed only customer segmentation, without the possibility of predicting which segment a new customer would be assigned to, based on the client's characteristics.

One of the most important challenges in banks is knowing their clients, understanding the differences between them and ranking them (Khajvand and Tarokh 2011). These authors have concluded, on the basis of research in the banking sector of Iran, that needs-based segmentation has been replaced by segmentation based on consumer behavior, as it is measurable and the valuation of each segment significantly facilitates marketing decision making.

Today, organizations are looking for diverse strategies to adequately organize their customer data and build consumer loyalty on this basis (Sheikh, Ghanbarpour, and Gholamiangonabadi 2019). This research emphasizes the importance and ability to identify specific customer segments within general segments and to define different marketing strategies based on their behavioral patterns. Similar results are found in the research of Ansari and Riasi (2016), which emphasized the importance of customer clustering in the banking sector such that the use of clustering allows banks to identify their most profitable clients and design marketing strategies for each customer group, based on their attributes.

Analysis of customer behavior in the banking sector is quite complex because the databases are multidimensional and consist of monthly and daily transaction records of a large number of clients (Hsieh 2004). In this study, Hsieh (2004) recommends integrating data mining models with a behavioral scoring model, aiming to understand existing credit card users. The results show that the bank's profit is the result of establishing good relationships with consumers whose loyalty and income are increasing. Hsieh (2004) used SOM neural networks for clustering, and associative rules for cluster descriptions. However, while SOM is a predictive method and can be used to predict cluster membership based on consumer characteristics, this method does not provide explicit segmentation rules. To describe the clusters, associative rules are used; however, its semantics can be complex to interpret due to the already mentioned redundancy and overlapping rules.

Cheng and Chen (2009) use an RFM model and *k*-means algorithm to cluster customers by value, and then implement a rough set LEM2 algorithm to extract classification rules in order to implement the conclusions in CRM. However, although this paper involves predictive classification, the problem of the minor class is still present. To increase predictive accuracy, they used RFM attributes as predictors, which can absorb the impact of consumer characteristics, so the RFM attributes do not appear in the classification rules that describe the segments. In addition, RFM attributes were coded first, leading to some loss of information.

Rogic and Kascelan (2019) have proposed a hybrid SVM-RE method to improve the segmentation method for customer selection in direct marketing. This method increases the targeting efficiency of existing customers and the prediction of new ones that are most likely to respond to a direct campaign. The authors confirmed that this method successfully solves the problem of the minor class in respondent prediction. Unlike Cheng and Chen (2009), they did not use RFM attributes as predictors, but used product data and sales data (discounts), which may also cause the rules describing segments not to contain customer characteristics.

Our research starts from the RFM model and proposes combining DM techniques: *k*-means clustering, predictive SVM classification and DT for rule extraction. Combining these techniques overcomes the prominent shortcomings of the aforementioned research in many ways. Firstly, only the R attribute is encoded during clustering, while the F and M attributes are retained in their original form. Due to automatic clustering, subjectivity is excluded and, because of non-coding of attributes, it enables more accurate segmentation without loss of information. Then, segment prediction for the new user is enabled solely on the basis of its characteristics and the problem of the minor class (most valuable clients) is solved by applying SVM methods as a class-balancing pre-processor. In addition, the problem of the non-interpretability of the SVM model is solved by generating a comprehensive set of classification rules in terms of customer characteristics using DT methods. This allows for easy targeting of the existing segments and more efficient prediction of new, most valuable customers for the bank.

The SVM method has been previously used in the literature as a pre-processor of data to solve problems of unbalanced classes (Martens et al. 2006; Farquad & Bose, 2012; Kascelan et al., 2014; Rogic and Kascelan 2019) but it was applied for the first time in this research to segmentation and more efficient predictions of the most valuable bank customers.

Methodology

This paper proposes a model for the predictive classification (segmentation) of clients in the banking sector.

The original dataset used for this purpose was obtained from the First Bank of Montenegro and relates to transactions performed using the bank's credit cards for the period from October to December 2018. This dataset includes 12 016 credit card users, with the dataset formatted at the customer level without revealing user identities. The obtained database contains data on bank customers, gender, place of residence (region), age, average monthly income, as well as data on transactions, such as the date and value of transactions, on the basis of which calculations of RFM attributes were made for research purposes. The data distribution is given in Table 1.

Table 1. Variable distribution.

Role	Name	Type	Statistics	Range
label	Cluster	polynominal	mode = cluster_0 (7239), least = cluster_1 (31)	cluster_2 (4746), cluster_0 (7239), cluster_1 (31)
regular	client_id	integer	avg = 9188.843 ± 5319.809	[3.000; 18400.000]
regular	client_gender	binominal	mode = 2.0 (5991), least = 1.0 (5961)	1.0 (5961), 2.0 (5991)
regular	client_age	integer	avg = 42.874 ± 14.374	[1.000; 99.000]
regular	avg_monthly_income	numerical	avg = 603.259 ± 4913.199	[0.000; 526616.020]
regular	client_place	polynominal	mode = Central region (6282), least = International (474)	Central region (6282), International (474), Southern region (3262), Northern region (1998)
regular	R	integer	avg = 2.513 ± 1.461	[1.000; 5.000]
regular	F	integer	avg = 24.639 ± 31.686	[1.000; 475.000]
regular	M	numeric	avg = 552.812 ± 1406.128	[0.010; 51370.000]

The RFM attributes (Hughes 1994) are defined as follows: R – date of last transaction, F – total number of transactions in the considered period and M – monetary amount spent in the considered period. The attribute R is coded so that a score of 5 is assigned to the 20% of the most recent dates, the next 20% of the less recent dates get a score of 4, and so on to a score of 1. This is done with the purpose of implementing *k*-means clustering, which operates solely with numerical attributes. The attributes F and M are retained in their original form, i.e. they are not coded. The RFM attributes were normalized to a 0–1 rank transformation before clustering. By means of these transformed attributes, using *k*-means clustering, customers are separated into clusters. It can be seen from the literature review that this data mining technique is gaining more and more application in marketing research.

Generally, clustering is the process of grouping similar objects into groups. This method is effective and often used in this area, especially for target-group definition (Wu & Lin, 2005), since it can identify consumers with similar purchasing behaviors and facilitate the process of market segmentation.

K-means (MacQueen 1967) clustering is a technique that, for a chosen value of *k*, identifies *k* object clusters, which are based on objects near the center of *k* groups, with the center defined as the average of the n-dimensional attribute vectors of each cluster (Ledolter 2013). Thus, *k*-means is an unsupervised technique for classifying clients into a certain number of clusters, with the clients within the cluster being similar and the clusters different. To define the optimal number of clusters, the Davies–Bouldin index is used (Davies and Bouldin 1979), whereby we look for the lowest absolute value of this index.

The degree of customer loyalty (i.e. belonging to the relevant cluster) is then predicted using a predictive classification method. In this regard, the predictive DT technique is used first.

The Decision Tree method uses a tree structure to define predictive classification rules (EMC Education Services, 2015), with the goal of predicting output based on specific inputs, i.e. the value of the dependent variable. The DT classification method divides the initial dataset up by attribute values so that subsets contain as many examples of one class as possible. A measure of division quality can be either the accuracy of the whole tree, information gain (Quinlan 1986), gain index (Quinlan 1992), or Gini index (Breiman 1984). During the inductive division, a tree-shaped model is formed, with paths from roots to leaves defining if–then classification rules in terms of predictive attributes. The following parameters are used to train the DT model: minimum split, tree depth, minimum gain, and leaf size. A tree of greater accuracy and complexity is obtained by a greater tree depth, smaller leaf size, lower minimum gain, and smaller minimum size for split.

In order to increase the accuracy and precision of the predictive performance of the model, the analysis is continued using the Support Vector Machine method (Vapnik 2010).

This classifier maps data from the original space, which is linearly inseparable, into a feature space of higher dimension, where classes can be separated by means of a hyperplane. Therefore, the aim of this method is to find a hyperplane that maximizes the margin – the empty space between classes (separating them) – that is, which minimizes the distance to the nearest representatives of the class-support vectors. Finding such a hyperplane in the feature space comes down to the optimization problem of square programming in the original space. The kernel function is used to calculate the scalar product of the vector from the feature space in the original space (kernel trick).

Of the many kernel functions, the most efficient and the most commonly used is the RBF – Radial Basis Function (Sanderson, 2010), which is defined by the formula (1), and has also been used in this research.

$$K(xi, xj) = exp(-\gamma||xi - xj||2) \quad (1)$$

The SVM classifier is optimized by finding the optimal gamma parameter combination for the RBF kernel function, as well as a C parameter that defines the margin width. A higher value of the parameter C allows higher classification accuracy in the training set, while a lower value of this parameter contributes to the generality of the SVM model, i.e. better classification accuracy on an unknown dataset.

The SVM classifier has certain advantages over other methods. Namely, it solves problems of linearly non-separable classes (mapping data into higher-dimension space) and unbalanced classes (eliminates class overlap noise in the data, by assigning to a smaller class a number of examples of a larger class that are closest to them; this is realized by moving the margin to a larger class, i.e. by adjusting the parameter C appropriately (Farquad & Bose, 2012). For this reason, SVM can be used as a pre-processor that refines data for other classifiers.

Given that SVM generates a non-interpretable model, the final step in the analysis is the extraction of rules from its output. Barakat and Bradley (2010) state that SVM-RE (Rule Extraction) techniques can be grouped into two categories: techniques that are based on the components of the SVM model, and those that do not use the SVM structure, but derive rules from its output. In our paper the DT method was proposed for the extraction of rules from the SVM output. Thus, on the one hand SVM is a pre-processor of data (balances classes) for DT, and improves its classification performance, and on the other hand DT interprets the results of the SVM classification.

Rule generation is of particular importance for marketing activities, since it allows precise targeting of consumer segments. Hence, it is possible to place differentiated offers to clearly defined segments, according to their specific characteristics and needs.

Summarizing the steps outlined above, the procedure for predicting the classification of bank customers based on their credit card use is as follows: the data is first prepared for analysis by calculating the RFM attributes as described above, and normalization, i.e. 0–1 rank transformation, is carried out. Client segmentation is then performed using *k*-means clustering, based on normalized RFM attributes. The number of clusters is defined based on the DB index. The resulting client clusters (segments) are described in terms of RFM attributes and are assigned the appropriate CV level. For the purpose of predictive client classification, a DT model is generated first. This involves choosing the optimal combination of parameters (the criterion for split evaluation, the minimum size for split, the minimum leaf size, the minimum gain, and the maximum depth) using the Grid-Search technique and *k*-fold cross-validation. The next step is to generate an SVM model for predicting a CV level based on credit card users, finding the optimal combination of C and gamma parameters through the Grid-Search technique. Then, the CV level obtained from the SVM prediction is generated, which is included in the data set as a class label. The next step involves generating a DT model, to predict the CV level based on credit card customer information and the SVM class label, finding the optimal combination of parameters (the minimum size for split, the minimum leaf size, the minimum gain, and the maximum depth). The accuracy achieved in this step is compared with the prediction accuracy of the first generated DT model. Finally, the last step is to generate rules from the DT model described in the previous step. Whether the result of the DT model can be considered a consistent interpretation of the SVM depends on the fidelity parameter, which is the percentage of examples for which the class label provided by DT matches the SVM class label.

Accuracy rates, class precision, and class recall (Silberschaz, Korth, and Sudarshan 2011), which were obtained through the *k*-fold cross-validation process, were used to assess the accuracy of predictive classification. The accuracy rate refers to the percentage of accurately predicted examples out of the total number of examples. Class precision represents the percentage of accurately predicted examples within the predicted class, while class recall describes the percentage of accurately classified examples within the current class.

The aforementioned analysis was conducted using Rapid Miner.

Taking into account the perceived shortcomings of previous research and the advantages of the methodology proposed above, this research aims to confirm the following hypotheses:

H1: SVM combined with DT increases the overall accuracy of predictive classification, thereby increasing the segmentation efficiency of existing and new customers in terms of their value to the bank.

H2: SVM as a data pre-processor for DT increases class recall and class precision for the smaller class, thereby increasing the efficiency of targeting existing and new, most valuable customers for the bank and reducing marketing losses due to the lack of recognition of a significant number of such users.

H3: DT, as a rule extractor, consistently interprets the SVM classification (with high fidelity) allowing profiling of the most valuable, medium value, and less valuable users for the bank by their characteristics.

Results and Discussion

As noted in the previous section, the data was initially prepared for analysis by calculating the RFM attributes and normalization (0–1 rank transformation).

Then, *k*-means clustering was performed, the results of which are given in Table 2. Given that, for the three clusters, the absolute values of the DB index are the lowest, we will continue our analysis by describing these three clusters (Table 3).

From Table 3, we can observe that the clients from cluster 0 are of low recency, low frequency and low profitability (CV = low). Cluster 1 covers the

Table 2. *K*-means clustering – selection of number of clusters.

K	2	3	4	5	6	7	8	9	10
DB	−0.9	**−0.751**	−0.757	−0.881	−0.892	−0.892	−0.867	−0.887	−0.94

Table 3. Centroid Cluster model for RFM customer segmentation.

	R	F	M	Items
Cluster 0	−0.69554	−0.4433736	−0.21668	7239
Cluster 1	0.112832	2.16969213	14.58756	31
Cluster 2	1.060167	0.66209878	0.235214	4746

most profitable and high-frequency clients (CV = high), while cluster 2 covers the most recent, medium-frequency and medium-profit clients (CV = medium). As can be seen from Table 3, cluster 1 contains significantly fewer items than the other two clusters, indicating a class imbalance problem, i.e. minor-class problem.

In order to predict the CV level, the next step is to optimize the parameters for the DT model through the Grid-Search technique and 10-fold cross validation. Then, a DT model was generated with optimal parameters. In the next step, the SVM is optimized using the Grid-Search technique and 10-fold cross validation and the optimal parameters are generated. Then the results of the SVM model obtained, i.e. the predicted CV level values, are included in the original dataset as a class label. After the introduction of this class label, a DT (pre-procesed DT) was generated on the SVM output. Table 4 shows the results of the predictive classification procedure.

From Table 4, we can observe that the accuracy of the DT model was improved after conducting the SVM pre-processing, from 63.37% to 74.69%. In this regard, clients from different segments can be more accurately identified and predicted, which positively affects the efficiency of marketing

Table 4. Results of testing the predictive classification procedure.

Model (optimal parameters)	Accuracy	Class recall	Class precision
DT (Gini index, min_size_for_split = 4, min_leaf_size = 2, max_depth = 15, min_gain = 0.1)	63.37%	**47.75**%, 73.82%, **16.13%**	**54.41**%, 68.22%, **27.78%**
SVM (gamma = 100 000, C = 50)	89.15%	82.43%, 93.69%, 58.06%	89.58%, 88.92%, 81.82%
Pre-processed DT (Gini index, min_size_for_split = 4, min_leaf_size = 2, max_depth = 15, min_gain = 0.1)	74.69%	**64.23%**, 80.71%, **63.64%**	**65.55**%, 79.72%, **93.33%**

Since the SVM pre-processes the initial dataset, it shows the predictive performance indicators obtained for the entire dataset (which also contains the examples with which the model was trained), while for the DT models it shows the average indicators obtained over 10 unknown datasets (i.e. the 10-fold cross-validation result). Table 5 shows the primary predictive rules which were derived from the pre-processed DT.

Table 5. Primary predictive rules derived from pre-processed DT.

Rule number	Rule	CV level	Rule Accuracy	Rule Coverage
6	avg_monthly_income > 7546.060 and client_place = Central region and client_gender = female and client_age ≤ 61.500	**cluster_1 (0/1/3)**	75.00%	0.03%
7	412.120 ≤ avg_monthly_income ≤ 7 546.060 and client_place = Central region and client_gender = female and client_age ≤ 61.500	cluster_2 (893/518/2)	63.20%	11.76%
8	412.120 ≤ avg_monthly_income > 240.730 and client_place = Central region	cluster_0 (445/745/4)	62.40%	9.94%
31	avg_monthly_income > 13 733.335 and client_place = Southern region and client_gender = male	**cluster_1 (0/0/3)**	100.00%	0.02%
44	0.645 ≤ avg_monthly_income ≤ 237.240 and client_place = Central region and client_age ≤ 53.500	cluster_0 (290/763/0)	72.46%	8.76%

activities and the success of individual strategies. Based on the presented results, we conclude that hypothesis H1 is confirmed.

Regarding the most valuable cluster (cluster 1), which is the smallest cluster, the DT method accurately classified 16.3% of the current clients (class recall), while the pre-processed DT increased the class recall to 63.64%, which represents a significant improvement. Thus, after SVM, DT improved the targeting by 48% (64%–16%) for the most frequent and most profitable existing customers, compared to DT before the SVM. Also, DT, after SVM, targeted 16% more (64%–48%) of the most recent current customers (cluster 2) than DT before SVM. More accurate targeting of the most valuable existing customers at a significantly higher percentage allows managers to establish communication with the primary segment and avoid an unnecessary waste of marketing resources for mis-targeted clients, thereby increasing the efficiency and effectiveness of marketing activities. Respecting Pareto's rule that 20% of customers account for 80% of total profits (Sanders 1987) leads us to conclude that it is of particular importance to accurately identify the most valuable customer segment.

For the most valuable customer cluster, class precision increased from 27.78% for DT, to 93.33% for DT after SVM. The prediction accuracy achieved with the DT model alone (without SVM class balancing) leaves more room for error in the selection of new clients, by incorrectly predicting as many as 70% of them. In this regard, the likelihood of the marketing activities undertaken being successful is significantly lower. High accuracy of predicting the most valuable clients is one of the basic preconditions for the success of the modern marketing concept. Namely, mistakes in this process carry a big risk for a company, because they lead to potentially high costs, as these are caused by an unnecessary waste of marketing efforts. The SVM method, applied in this way, allows for high forecasting accuracy (with fewer than 7% of the most valuable clients wrongly predicted), which significantly reduces the costs of marketing activities aimed at attracting new clients, as well as reducing the likelihood of error in this process.

Taking this result into account, it has been shown that the SVM method effectively solves the problem of class imbalance and increases the accuracy of classification and prediction for a minor class. Hence, significant losses in marketing are prevented, which confirms hypothesis H2.

As noted earlier, SVM does not provide the ability to interpret the model, therefore we used it in combination with DT to perform rule extraction. The rules thus obtained consistently interpret the SVM predictive classification, since their fidelity is about 75%. Some of the 74 rules are given in Table 4, which are significantly accurate and cover a large number of examples.

Based on the rules derived, it can be observed that, for example, those clients who are CV level = high (cluster 1) are those with an average monthly income of more than EUR 7 546.06 and who live in the central or southern region, with gender playing no important classification role (rule 6). Cluster 2, comprised of CV level = medium clients (rule 7), for example, is represented by customers with an average monthly income of between EUR 412.12 and €7 546.06, under the age of 61 and with residency in the central region. This rule covers as many as 11.76% of the total number of clients and its accuracy is 63.20%. Finally, the largest client cluster – CV level = low, is represented by the largest number of rules. For example, rule 8 describes these clients as those with an average monthly income of less than €412.12 and from the central region, while rule 44 describes these clusters based on age as well, and with 72.46% accuracy it defines this segment of clients as being younger than 53 years.

A differentiation strategy in the banking sector based on the derived rules is a source of competitive advantage. Objective profiling of different categories of clients makes it easier to create different offers for different segments, placing offers with high precision to the relevant segments through marketing activities. Given the complexity of supply in the banking sector today, the problem of creating different offers poses a challenge. However, by understanding the characteristics of clients and the segments they belong to, as well as their value to the bank, they provide the conditions for successful implementation of the differentiation strategy. Based on the above, we conclude that hypothesis H3 is confirmed.

Conclusion

This paper proposes a model for efficient predictive segmentation of clients (credit card users) based on their value to the bank. Most standard methods for predictive classification (logistic regression, DT, neural networks, etc.) lead to the misclassification of the minor class, which becomes an issue, since the segment of the most valuable clients is the smallest. Thanks to SVM's ability to declare a number of large class instances most similar to those of the minor class as actual minor class examples, this method can serve as a data pre-processor that balances classes and increases the predictive performance of the standard classifiers mentioned above. Using this feature of SVM, a predictive classification model has been proposed that combines SVM pre-processing (which solves the minor class problem) with the DT method, by which explicit classification rules are extracted. The proposed model was tested on empirical customer data and card transactions taken from a Montenegrin bank. The results confirmed the effectiveness of the proposed model because the obtained predictive performance showed that the problem of minor class misclassification was overcome and the accuracy of prediction was generally increased.

Based on such a model, clients from different clusters can be targeted by different promotional strategies, while avoiding subjective profitability estimates for particular segments. In other words, a bank can more accurately target customer groups in terms of their value to the bank, which positively affects the effectiveness of marketing activities and the success of individual strategies.

Due to the fact that the proposed model successfully overcomes the problem of the minor class, significant losses in marketing are prevented due to misclassification and prediction of the most valuable clients. Successful implementation of CRM entails a significant reduction in the unnecessary costs that can occur if a model wrongly declares a client as "valuable" that actually is not or, vice versa, if we neglect a very valuable client because the model declared it as "not valuable" to the bank. On the other hand, accurate classification and prediction of the most valuable clients result in higher revenues. In addition, by creating interaction between the bank and the most valuable customers, their loyalty is strengthened, their rate of loss is reduced, and CRM is significantly improved.

Creating an objective profile of different categories of clients makes it easy to create different offers by differentiated segments, with a high precision of placing offers to the appropriate segments while implementing marketing activities. By understanding the characteristics of clients and the segments to which they belong, as well as their value to the bank, the results provide the conditions for successful implementation of the differentiation strategy.

For future research, the question remains as to whether the effectiveness of this model can be validated in sectors other than banking, as well as by using different predictors.

References

Ansari, A., and A. Riasi. 2016. Taxonomy of marketing strategies using bank customers' clustering. *International Journal of Business and Management* 11 (7):106–19. doi:10.5539/ijbm.v11n7p106.

Barakat, N., and A. P. Bradley. 2010. Rule extraction from support vector machines: A review. *Neurocomputing* 74:178–90.

Breiman, L. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Cheng, C., and Y. Chen. 2009. Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems With Applications* 36 (3):4176–84. doi:10.1016/j.eswa.2008.04.003.

Davies, D. L., and D. W. Bouldin. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1 (2):224–27.

Dibb, S., and L. Simkin. 2010. Judging the quality of customer segments: Segmentation effectiveness. *Journal of Strategic Marketing* 18 (2):113–31. doi:10.1080/09652540903537048.

EMC Education Services. 2015. Data Science & Big Data Analytics. Indianapolis, IN: John Wiley & Sons, Inc. doi:10.1002/9781119183686

Farquad, M. A. H. and I. Bose. 2012. Preprocessing unbalanced data using support vector machine. *Decision Support Systems 53*(1): 226-233. doi:10.1016/j.dss.2012.01.016

Hsieh, N. 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems With Applications* 27 (4):623–25. doi:10.1016/j.eswa.2004.06.007.

Hughes, A. M. 1994. Strategic database marketing: The masterplan for starting and managing a profitable, customer-based marketing program. Irwin, Chicago.

Kašćelan, L., V. Kašćelan, and M. Novović-Burić. 2014. A Data Mining Approach for Risk Assessment in Car Insurance: Evidence from Montenegro. *International Journal of Business Intelligence Research 5*(3): 11-28. doi:10.4018/IJBIR

Khajvand, M., and M. J. Tarokh. 2011. Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science* 3:1327–32.

Khalili-Damghani, K., F. Abdi, and S. Abolmakarem. 2018. Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. *Applied Soft Computing* 73, 816-828. doi:10.1016/j.asoc.2018.09.001

Khan, D. A. 2018. Information technology in supply chain management. *Journal of Supply Chain Management* 3 (3):35–46.

Khobzi, H., E. Akhondzadeh-Noughabi, and B. Minaei-Bidgoli. 2014. A new application of rfm clustering for guild segmentation to mine the pattern of using banks'e-payment services. *Journal of Global Marketing 27*(3): 178-190. doi:10.1080/08911762.2013.878428

Kim, G., B. K. Chae, and D. L. Olson. 2013. A support vector machine (SVM) approach to imbalanced datasets of customer responses: Comparison with other customer response models. *Service Business* 7 (1):167–82. doi:10.1007/s11628-012-0147-9.

Kumar, V. M., V. M. Chaitanya, and M. Madhavan. 2012. Segmenting the banking market strategy by clustering. *International Journal of Computer Applications* 45 (17):10–15.

Ledolter, J. 2013. *Data mining and business analytics with R*. Hoboken, New Jersey: John Wiley & Sons.

Li, G., Z. Ma, C. Huang, Y. Chi, and H. Wang. 2010. Segmentation of color images of grape diseases using K_means clustering algorithm. *Transactions of the Chinese Society of Agricultural Engineering 26*(1): 32-37.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California* 1: 281–97.

Martens, D., J. Huysmans, R. Setiono, J. Vanthienen, and B. Baesens. 2006. Rule extraction from support vector machines: An overview of issues and application in credit scoring. *Studies in Computational Intelligence* 80:33–63.

Miguéis, V. L., A. S. Camanho, and J. Borges. 2017. Predicting direct marketing response in banking: Comparison of class imbalance methods. *Service Business* 11 (4):831–49.

Mohammad, S., H. Hasheminejad, and M. Khorrami. 2018. Data mining techniques for analyzing bank customers : A survey. *Intelligent Decision Technologies* 12 (3):303–21. doi:10.3233/IDT-180335.

Mohammadi, R., B. Bidabad, T. Nourasteh, and M. Sherafati. 2014. Credit ranking of bank customers (An integrated model of RFM, FAHP and K-means). *European Online Journal of Natural and Social Sciences* 3 (3):564–71.

Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1 (1):81–106.

Quinlan, J. R. 1992. *C4.5 - programs for machine learning*. San Mateo, CA: Kaufmann.

Rogic, S., and L. Kascelan. 2019. Customer value prediction in direct marketing using hybrid support vector machine rule extraction method. In *New trends in databases and information systems. ADBIS 2019. Communications in computer and information science*. In T. Welzer,J. Eder, V. Podgorelec, R. Wrembel, M. Ivanovic, J. Gamper, M. Morzy, T. Tzouramanis, J. Darmont, & A. K. Latific (Eds) vol. 1064. Cham: Springer: 283-294.

Sanders, R. 1987. The pareto principle: Its use and abuse. *Journal of Services Marketing* 1 (2):37–40. doi:10.1108/eb024706.

Sanderson, M., D. Christopher, R. Prabhakar, and S. Hinrich. 2010. Introduction to Information Retrieval. *Cambridge University Press. Natural Language Engineering* 16 (1):100–03.

Sheikh, A., T. Ghanbarpour, and D. Gholamiangonabadi. 2019. A preliminary study of fintech industry: A two-stage clustering analysis for customer segmentation in the B2B setting. *Journal of Business-to-Business Marketing* 26 (2):1–11. doi:10.1080/1051712X.2019.1603420.

Silberschaz, A., H. F. Korth, and S. Sudarshan. 2011. *Database system concepts (6th ed)*. New York, NY: McGraw-Hill.

Tavakoli, M. (2018). Customer segmentation and strategy development based on user behavior analysis, RFM model and data mining techniques : A case study. *2018 IEEE 15th International Conference on E-Business Engineering (ICEBE)*, Xi'an, China:119–26. 10.1109/ICEBE.2018.00027

Vapnik, V. N. 2010. *The nature of statistical learning theory*. New York: Springer.

Wu, J., and Z. Lin. 2005. Research on customer segmentation model by clustering. In *Proceedings of the 7th international conference on Electronic commerce*: 316-318. doi:10.1145/1089551

Yang, A. X. 2004. How to develop new approaches to RFM segmentation. *Journal of Targeting, Measurement and Analysis for Marketing* 13 (1):50–60. doi:10.1057/palgrave.jt.5740131.

Zakrzewska, D., and J. Murlewski. 2005. Clustering algorithms for bank customer segmentation. In 5th International Conference on Intelligent Systems Design and Applications: 197–202.