Computer Science and Information Systems 18(3):893–925 https://doi.org/10.2298/CSIS200530052R

Class Balancing in Customer Segments Classification Using Support Vector Machine Rule Extraction and Ensemble Learning

Sunčica Rogić and Ljiljana Kašćelan

University of Montenegro, 81000 Podgorica, Montenegro
{suncica, ljiljak}@ucg.ac.me

Abstract. An objective and data-based market segmentation is a precondition for efficient targeting in direct marketing campaigns. The role of customer segments classification in direct marketing is to predict the segment of most valuable customers who is likely to respond to a campaign based on previous purchasing behavior. A good-performing predictive model can significantly increase revenue, but also, reduce unnecessary marketing campaign costs. As this segment of customers is generally the smallest, most classification methods lead to misclassification of the minor class. To overcome this problem, this paper proposes a class balancing approach based on Support Vector Machine-Rule Extraction (SVM-RE) and ensemble learning. Additionally, this approach allows for rule extraction, which can describe and explain different customer segments. Using a customer base from a company's direct marketing campaigns, the proposed approach is compared to other data balancing methods in terms of overall prediction accuracy, recall and precision for the minor class, as well as profitability of the campaign. It was found that the method performs better than other compared class balancing methods in terms of all mentioned criteria. Finally, the results confirm the superiority of the ensemble SVM method as a preprocessor, which effectively balances data in the process of customer segments classification.

Keywords: direct marketing, customer classification, class imbalance, SVM-Rule Extraction, ensemble.

1. Introduction

Direct marketing allows for direct communication with potential customers through various media, such as e-mail, catalogs, social media and the like. It is consumer-oriented, message is sent directly to consumers and, at the same time, it's a "call to action". One of the key issues of this type of marketing is the accurate identification of potential and current customers who will most likely respond to a campaign, i.e. targeting specific customers from an existing database as well as new potential leads. Usually, the customer targeting methods are split up in the literature into segmentation and scoring methods [1, 2]. Segmentation methods, using appropriate explanatory variables, partition the customers into homogenous segments regarding the anticipated response to a direct marketing campaign [3, 4]. Thus, the promotional offers and materials are distributed to such customer segments with highest expected probability of response. On the other hand, scoring methods are used in the customer response models [5, 6], by assigning certain scores to customers, based on the predicted likelihood of the response to the campaign. It is important to state that high probability of response to the campaign does not certainly imply high profits. Hence, methods for customer profitability prediction are included in some of the most important scoring methods [7–10].

Segmentation methods, which are most commonly applied in direct marketing, split a customer data set using Recency, Frequency and Monetary (RFM) attributes [11]. They are based on various techniques, ranging from the simplest cross-tabulation technique, to more complex weighted techniques [4, 12]. These techniques generally require a subjective assessment for the necessary parameters. For this reason, data mining methods, such as K-means or Artificial Neural Network (ANN) clustering, can give more objective results for RFM customer segmentation [13–15].

Recently, classification data mining methods have become very popular, as they can enable the prediction to which segment the customer belongs to, based on the characteristics of the customer [15, 16]. Since the most valuable customer segment is usually the smallest, there is a problem of class imbalance. This problem in most classification methods leads to bias toward small classes and most often to their misclassification [17, 18]. If this problem is ignored, most classification algorithms will not identify the most valuable customer segment at all, or will identify a very small number of customers within that segment, which may lead to an unprofitable campaign.

There are methods in the literature that overcome the class imbalance problem in different ways [19–21]. The main disadvantage of the most commonly used under-sampling method, is that it ignores the large number of examples of the larger class, that may contain significant information for class differentiation. In order to reduce sample bias and minimize the loss of significant information, it can be combined with ensemble techniques (balanced ensemble) [18, 22]. The balanced ensemble approach involves taking random subsets of a larger class (equal in size with a smaller class) in multiple iterations and generating different classification models over those subsets whose results are eventually aggregated to give a final result. Combining multiple classifiers in this way does not only balance classes, but also increases predictive accuracy, reduces sample bias, reduces variance i.e. increases stability of results and avoids overfitting [23, 24].

The previous literature confirms that in case of class imbalance and overlapping the SVM method has a good predictive performance and can be used as a preprocessor that balances classes for other classifiers [25]. However, the SVM classifier is a "black-box", i.e. does not generate a model that can be interpreted, which is very important in the classification of customer segments in order to describe the segments. This deficiency can be solved by a hybrid approach, where the SVM is combined with rules extracting techniques (SVM-RE) [26].

Considering the advantages of the SVM-RE method and ensemble approach noted above, this paper proposes customer segments classification in direct marketing based on a combination of SVM-RE predictive classification and ensemble meta-algorithms. Also, a comparison of the defined method with the standalone data balancing ensemble methods for customer classification was made. Specifically, this study highlights three main research questions (RQ):

RQ1 - Is the ensemble SVM-RE approach adequate for the prediction of customer segments in direct marketing?

RQ2 - Given the unbalanced nature of data in customer segmentation, what is the best class balancing method (ensemble SVM-RE or one of the standalone balancing ensemble methods)?

RQ3 - Is the ensemble SVM-RE method suitable for describing, i.e. explaining segments?

Undoubtedly, the primary success factor of direct marketing predictive models is class imbalance. There are two basic approaches used in previous studies to solve this problem. The first involves modifying the classifier by associating different misclassification costs for each class [27, 28]. Another approach to requires data changes. Balancing is regulated by generating or reducing data using over-sampling or under-sampling techniques [17, 18, 22, 29, 30]. However, both approaches have some drawbacks. Classifier-changing methods require extensive knowledge about the specific learning methods [31], so it is necessary to hire an expert for practical application. With resampling methods, the challenge is removing the data without losing the information necessary to distinguish the classes, as well as knowing whether removing some data would give different results (instability of the solution). When supplementing the data, the main challenge is how to supplement the minority class while maintaining its distribution. Also, the question is what is the optimal class ratio [31]. All this makes the analysis complex in practical applications. A small number of papers for the customer classification use the ordinary SVM method [17, 32, 33], but it has been shown that it is not immune to class imbalance either [17]. The main contribution of the SVM-RE method proposed in this paper is that it automatically eliminates noise and class imbalance. By adjusting the parameters of the SVM as a data preprocessor, the boundaries between the classes are shifted so that the examples of the majority class that are closest (most similar) to the minority class join the minority class. Rule extraction from such balanced data has good classification performance for the minority class as well. The extracted rules provide a description of the segment of the most valuable customers that cannot be obtained if the misclassification of this minority class is not resolved. The performance of SVM-RE methods is further enhanced by combining with ensemble techniques.

Unlike the above-mentioned studies, which mainly use data sets from publicly available repositories, this study uses real-life data that are disordered and have more noise. On such data, the challenge of balancing and achieving good predictive performance is even greater. The final step was using the public dataset to validate the model.

The practical implications of this paper relate to the more accurate and objective planning of direct marketing campaigns, as well as gaining deeper insight into different customer segments, which may lead to more precise targeting and increased profits.

The paper is organized as follows: The second section gives an overview of related papers. Section three shows the proposed methodology, and the fourth section presents the results of the empirical test, which are further discussed in the fifth section. Finally, the sixth section contains concluding remarks.

2. Literature Review

This section provides an overview of previous research related to the customer segmentation and the problem of class balancing in direct marketing.

2.1. SVM Rule Extraction Method

For linearly inseparable classes, Vapnik[34] proposed a SVM method that maps data (viewed as n-dimensional vectors) from the original space into a larger dimension space (feature space), where the classes can be separated by means of a hyperplane. Finding such a hyperplane is realized by minimizing the distance between its end position (so that the gap between the classes i.e. the margin is greater) and the closest points (support vectors). Instead of an explicit mapping function in a larger dimension space, a kernel function is used, which allows calculating the scalar product of the vectors (i.e. the distance of the support vector from the hyperplane) in the original space (kernel trick). Various kernel functions can be used, but Radial Basis Function (RBF) which was used in this paper, is applied most often [35]:

$$K(x_i,x_j) = \exp(-\gamma||x_i-x_j||^2) \quad (1)$$

The SVM algorithm, therefore, strives to maximize the margin in feature space, which boils down to the convex optimization i.e. quadratic programming problem in the original space (2):

$$max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \, \alpha_j \, y_i y_j \, K(x_i, x_j) \quad (2)$$

$$\sum_{i=1}^{n} y_i \alpha_i = 0$$

$$0 \le \alpha_i \le C \, , \;\; i = 1, \dots, n$$

where K is kernel function, αi are Lagrange multipliers, n is the number of training examples and C is a parameter, which is adjusted to trade off margin maximization against classification error minimization.

The training of the SVM classifier comes to the selection of the optimized values of the gamma parameter for the RBF kernel, and parameter C, which represents the boundary for the margin, i.e. empty space between classes. Selecting lower values for parameter C reduces over-fitting and increases the generality of the SVM model, i.e. its predictive performance.

In addition to solving the problem of linear inseparability of classes, the advantage of this method is that, in the case of class imbalance, it exhibits better predictive performance than the standard methods, such as logistic regression [25]. The literature confirms that the SVM can successfully remove the noise, i.e. class overlapping from data. Namely, the parameter C can be set so that a number of examples of a larger class, which are close to the example of a lower class (which means that they are similar), are declared as examples of the lower class. For this reason, the SVM can be used as a preprocessor that balances and purifies data, thus providing higher classification accuracy [25, 36].

However, the SVM does not generate an interpretable model, which is usually very important in application. This problem has been solved in the literature by means of rule extraction techniques that enable generating the rules from the SVM results [26, 37]. According to Barakat and Bradley [26] SVM-RE techniques are grouped into two categories: those based on the components of the SVM model, and those that do not use the internal structure of the SVM model, but draw the rules from the SVM output. When the SVM model is interpreted, or SVM is used as a data preprocessor, authors recommend techniques from the second group because they provide more understandable rules. In line with this recommendation, our research uses rule extraction from SVM output. Namely, customer targeting rules are derived from SVM output using a classification Decision Tree (DT) method [38].

The DT method divides the data set by attributes values, so that subgroups contain as many examples of one class as possible, i.e. their impurity is minimized. The criterion by which division is made (measure of quality of division) can be information gain [39], gain index [40], gini index [41], or accuracy of the whole tree. The attributes that provide the best division according to the given criterion are chosen. During the inductive division, a tree-shaped model is formed. The path from the root to the leaves defines if-then classification rules in the terms of the predictive attributes (tree nodes). The complexity and accuracy of the generated model depends on the depth of the tree, the minimum size of the node by which the division can be made (the number of examples in its subgroup), the leaf size, and the defined minimum gain achieved by the node division. The smaller the depth, the larger the minimum size for the split, the larger the leaf size and the higher minimum gain, lead to a less complex tree, but also a tree with smaller accuracy.

SVM rule extraction is not a new method in literature and it was applied in some previous economic studies [36, 38, 42], but for the topic of direct marketing, i.e., to solve the problem of the minor class of the most valuable customers in customer classification, it is applied for the first time in our research in this way.

2.2. Ensemble Methods

Ensemble methods use more learning algorithms to achieve better predictive performance than can be achieved with any of these learning algorithms alone.

There are several different ensemble methods. Thus, Bootstrap Aggregating [43] or shortened Bagging, constructs subsets of the training set using bootstrapping, which implies that the same example can be re-selected in the next iteration (sampling with replacement). The subsets thus obtained generate predictive models whose results are aggregated (the result for which most models voted is taken). If bootstrapping from a larger class takes as many instances as there are in a smaller class, then it is a case of balanced Bagging. Unlike Bagging, which generates samples and models simultaneously, Adaptive Boosting [44] generates the following sample and model based on previous results. Specifically, the succeeding sampling is more likely to select those examples that were previously incorrectly classified because they were given more weight. The result is obtained by weighting, i.e. based on the weight attached to the models depending on their accuracy. Random Forest [45] is an ensemble method that combines Bagging with a random selection of predictors. Each sample generates a forest of DTs that take random subsets of predictors, i.e., a random forest of DTs is generated.

Ensemble methods have been used in some previous studies in direct marketing. Hence, Gupta A. and Gupta G. [46] have compared neural networks and Random Forest to predict clients' response to a term deposit offered at a Portuguese bank. Their findings show that Random Forest performs better. Instead of boosting one learning algorithm, Lessmann et al. [47] proposed an ensemble approach that combines multiple different learning algorithms (decision trees, SVM, random forest, logistic regression, etc.) to create predictive marketing models, such as customer response prediction, profit scoring and churn prediction. In aggregating the results to evaluate the best models, they included profit maximization in addition to predictive model performance. The results showed that this ensemble approach outperformed standalone models in terms of profit. Lawi et al. [30] have combined Adaptive Boosting with SVM and achieved better predictive performance compared to ordinary SVM. In the approach proposed in this study, Bagging was combined with SVM to improve data preprocessing performance, i.e. to eliminate class overlapping, and balanced Bagging with the DT classifier is used to further help solve the problem of class imbalance and ultimately improve the performance of customer classification segments as much as possible.

2.3. Customer Segmentation in Direct Marketing

Hughes [11] defined one of the most commonly used customer segmentation method in direct marketing – RFM segmentation. The RFM model is based on the database of previous customers' purchasing behavior. Recency represents the time period since the last purchase, Frequency marks the number of purchases in the stated period of time and Monetary indicates the value of all customer's purchases during that period [48]. The analysis starts by sorting the available data into five equal segments (each containing 20% of customers), according to recency. Most recent customers receive the score 5, less recent score 4 and so on, following the Pareto principle – 20% of customers account for 80% of sales [49]. Following this procedure, customers are sorted according to their frequency, within the formed quintiles, receiving scores 5 to 1, which results in a database with 25 segments, and finally, database is split according to the monetary value by scoring the customers within the defined groups, which, in turn, results in a database with 125 groups based on the RFM values [4], where the best segment will have a 555 score, and the worst will have a 111 score. However, the choice of segments to be targeted in the future marketing campaigns is subjective.

The resulting segments based on the RFM model can be further analyzed using more objective data mining techniques, taking into account the customer features, their buying behavior or product specific variables [16, 50, 51].

Cheng and Chen [16] used k-means clustering [52] for RFM segmentation. They split the data into segments of 20% each (uniform coding) and created 3, 5 and 7 clusters to test the approach. The disadvantage of this approach is that uniform coding leads to the loss of fine differences between the values of RFM attributes (e.g., customers who have 5 or 9 transactions or those whose revenue is 5000 euros or 7000 euros can be placed in the same rank). Also, the pre-assumed number of clusters does not guarantee optimal RFM segmentation.

In order to develop a set of rules for targeting customers based on their features (the region and credit debt), the authors used a rough set and LEM2 rule extraction method. The predictive attributes also include RFM attributes, aiming to achieve high accuracy rate, as clusters are already formed on the basis of RFM. Hence, extracted rules perhaps do not show some significant customer characteristics for targeting, as they may be absorbed by the effect of RFM attributes. In addition, RFM attributes are unknown for new customers, so this model cannot be used for their prediction.

Additionally, the authors in [16] exclusively used accuracy rate (the percentage of precisely predicted examples within all examples) to determine their classification performance. Since there is usually the smallest number of customers with the highest value, clusters do not contain the same number of customers. Hence, there is a problem of class imbalance in the classification, which may lead to low class precision (the percentage of precisely predicted examples within a predicted class) and / or class recall (the percentage of accurately classified examples within the actual class) for the smallest class, which is the most important customer segment in this case.

In order to overcome the shortcomings mentioned above, a new method for customer segments classification based on data mining techniques will be tested in this paper. Customer segmentation by RFM attributes will be performed automatically using a clustering algorithm instead of manual coding and sorting. Clustering will be applied to the original attributes, so that there is no loss of fine differences that arise due to their uniform coding. Instead of a priori determining the number of clusters, an objective indicator of the optimal number of clusters will be used. Predictive classification will not include RFM attributes, therefore, classification rules describing segments will be defined in terms of customer and product characteristics, which is very important for customer relationship management. In this way, it is possible to classify new customers for whom RFM attributes are not known and available. Finally, and most importantly, the proposed method aims to reduce the misclassification of the most valuable customer segment.

2.4. Class Balancing in Direct Marketing

As pointed out, a major difficulty with predictive models in direct marketing is the class imbalance problem. According to the previously mentioned Pareto principle, the segment of the most valuable customers is the smallest (about 20% of the customers), but also the most important for the success of the campaign. The response rate in a direct campaign is often less than 5%, while non-responders make up as much as 95% of the total number of customers. This leads to very unbalanced datasets for training predictive classifiers in direct marketing [17, 22, 53]. Obviously, the problem of class balancing in this area is very topical, and accordingly, much more recent research deals with methods that effectively address this problem.

According to Sun et al.[31] data-level, algorithm-level and cost-sensitive solutions were developed for the problem when using imbalanced classes in classification models. At the data level, the aim is to balance classes with resampling, while solutions include random or targeted under-sampling and over-sampling. At the algorithm level, solutions try to adapt the algorithm to strengthen small-class learning. Cost-sensitive solutions, at both the algorithm and data levels, assign higher misclassification costs to small-class examples. More recently, there have been several papers that deal with this issue [54–56].

Although resampling eliminates class imbalance, this approach has several limitations and disadvantages, such as unknown optimal class distribution, inexplicit criterion in selecting examples for removal, risk of losing information relevant to class differentiation in majority class under-sampling, and risk of overfitting when over-sampling a minority class. Algorithm-level approaches require extensive knowledge of the algorithms and application domains, while cost-sensitive approaches involve extra learning costs for exploring effective cost setups, when real cost values are not available. However, despite the mentioned shortcomings, most of these solutions are used in recent research in the field of direct marketing.

Thus, Kim et al. [17] compared the efficiency of SVM classifiers with decision tree and neural networks on highly unbalanced data sets in direct marketing. They found that only SVM doesn't have a complete misclassification of the minor class, but, that positive sensitivity is very small, which means that the class imbalance is also an issue for SVM method. With random under-sampling of majority class with a ratio of 33% (i.e. the class ratio was 2:1), all classifiers improved their performance, while SVM still outperformed the others. However, with a 1:1 class ratio, the performance of SVM model has weakened, suggesting that by removing a large number of examples of the majority class, data relevant to the learning process may be lost. In that sense, it is good to combine under-sampling with ensemble techniques so that random selection is repeated several times and the probability of significant data being completely excluded is reduced, hence some papers dealing with the class imbalance problem in direct marketing go in that direction.

For example, Kang et al. [22] suggested improving customer response models by balancing classes using clustering, under-sampling and ensemble. First, the instances belonging to the non-response class are clustered. In the next step, under-sampling is performed as part of the ensemble procedure by randomly selecting a number of representatives from each cluster, proportional to the size of the cluster, but with the total number of selected instances equal to the minor class (balanced ensemble). In this way, taking a number of representative members of the larger class is achieved and reduces the loss of information relevant to class differentiation. By performing ensemble procedure in *k* iterations, on *k* of such balanced samples, *k* classifiers are generated and their predictions are combined. The results showed that compared to random sampling methods, this approach has more stable predictive performance that decision makers can trust more.

Migueis et al. [18] compared ensemble balanced under-sampling (the EasyEnsemble algorithm that uses sampling without replacement) with an over-sampling method (the Synthetic Minority Oversampling Technique-SMOTE) for direct marketing response prediction in banking and found the EasyEnsemble method gave better results. The sampling model without replacement can compromise the independence of the classifier in the ensemble procedure because the sampling in the next step depends on the one made in the previous step.

Marinakos et al. [29] tested cluster-based under-sampling and distance-based resampling techniques for the bank customer response model (with 12% of respondents and 88% of non-respondents) with several different classifiers, such as linear discriminant analysis, logistic regression, k-Nearest Neighbor (k-NN), decision tree, neural network and SVM. The highest accuracy of the minority class classification was achieved by the combination of cluster under-sampling and k-NN. Cluster under-sampling combined with SMOTE over-sampling proved consistently well, performing across all classifiers.

Peng et al. [27] proposed a solution based on algorithm adaptation in the form of cost-sensitive learning SVM for segmenting credit card users, and showed that this solution gives better results for the smallest class of high-value users than basic SVM with random under-sampling. This approach requires extensive knowledge of the SVM method in order to include misclassifying costs.

Farquad and Bose [25]tested SVM as a class-balancing preprocessor of insurance customers data and found that when classifiers are applied to such a refined set, a much higher sensitivity is obtained i.e. the number of current examples of the minor class that the model accurately classifies. They also found that data balancing with SVM is more efficient than other balancing techniques such as 100% and 200% SMOTE over-sampling or 25% and 50% under-sampling.

In our preliminary research [57], we tested how successfully a hybrid model that combines SVM and decision trees as a rule extraction technique (SVM-DT) solves the problem of the minor class of the most valuable customers. The results showed that with this approach, the segment of the most valuable customers can be predicted with an accuracy of 77%, which is 44% better than the standalone DT. Thus, SVM as a preprocessor has effectively improved the precision of the minor class. The improvement is even higher for the percentage of existing customers who are recognized as members of the most valuable cluster. Standalone DT identified only 4% of them, while SVM-DT managed to identify 63% of such customers. Although the model performed well on a training data set (obtained by cross-validation), it was not tested on an unknown data set, so its actual predictive power was not confirmed in this study.

In a study by Djurisic et al. [58] authors tested how well SVM preprocesses data and enables CRM optimization in banks. The results showed that during the segmentation of credit card users, this method successfully resolves overlapping and unbalanced classes. In this paper either, the model was not tested on a completely unknown data set.

In previous research, in order to overcome the class imbalance problem, balanced ensemble methods in combination with different classifiers, or standalone SVM, as a preprocessor that refines class overlapping and thus balances data, were mainly tested. This paper will test combining ensemble approach and SVM to improve preprocessing performance, as well as balanced ensemble in combination with DT on such a preprocessed dataset to improve rule extraction performance from SVM output, which should ultimately lead to improved performance of customer segments classification.

3. Methodology

The primary goal of predictive customer segmentation in direct marketing is customer value prediction, which determines whether or not a customer is targeted. This section describes the methodological approach for the proposed predictive procedure.

3.1. Data

First step is collection of data on purchasing transactions from previous direct campaigns, which can include customer data (such as: gender, age, region, wealth, etc.), product data (such as: type, category, purpose, etc.), and purchasing behavior data, i.e. recency, frequency and monetary value of purchases. The data were used as a training set for the predictive model.

For the empirical testing in this paper, a data set of on-line purchasing transactions from previous direct campaigns of Sport Vision Montenegro (part of the Sport Vision system - leading sport retailer in the Balkans) was used, for the period from the beginning of September 2018 to the end of January 2019 (fall/winter season). The data set consists of 1605 records (transactions) and has the following attributes: order ID, discount, price, date, gender, product type, product gender, product category, product age and product brand. Product type represents retailer's classification of products into: footwear (sneakers, shoes, boots, etc.), apparel (t-shirts, sweatshirts, joggers, etc.) and equipment (bags, dumbbells, gloves, etc.). On the other hand, product category is another form of classification, based on activities' purpose (for example, running – for running shoes, outdoor – for hiking equipment, etc.). Product gender consists of five values: products for women, for men, for boys, for girls and unisex products. In addition, product age describes the age group that products are intended for (for babies, for kids, for adults, etc.). Finally, product brand splits the products into two major groups – A brands (retailer's distribution brands) and Licence brands (retailer's production and distribution brands) and a small group of "Other" brands. In general, A brands are well known and established sport brands, that are usually more expensive, while Licence brands are more affordable, with not as strong image and brand recognition.

The data was prepared by calculating the RFM attributes as follows: Recency as the date of the last order, Frequency as the total number of orders in the considered period and Monetary as the monetary amount spent by a customer in the considered period expressed in euros. The Recency attribute is encoded so that for 20% of the most recent dates, score 5 is assigned, the next 20% less recent dates are given score 4 and so on until score 1. Attributes Frequency and Monetary are retained in their original form. In the end, all the attributes were normalized with 0-1 range transformation. Table 1 shows the attribute distribution in the starting data set.

For the purpose of testing the predictive performance of the model, the same type of data from the year after were used, but from the same season (fall/winter), when there is a similar sales offer available for consumers. This is because of seasonality, which affects and defines type of current offer. For example, in the fall/winter season, marketing focus is on "back to school" and skiing campaigns, while during the spring/summer season, focus is on summer activities. Hence, it makes sense to only compare the performance of the same seasons and different years, while the same values of attributes are available.

The data was prepared in the same way as the training set (RFM attributes were calculated and all attributes normalized).

Table 1. Attribute distribution in the training dataset

Attribute	Statistics	Range
Order_ID		[42 ; 6278]
Cust_gend	mode = M (891), least = F (714)	F (714), M (891)
Discount	avg = 0.371 +/- 0.107	[0.000 ; 0.500]
Prod_type	mode = Footwear (784), least = Equipment (181)	Footwear (784), Equipment (181), Apparel (640)
Prod_gend	mode = For men (786), least = For girls (67)	For women (399), For boys (210), For men (786), Unisex (143), For girls (67)
Prod_categ	mode = Lifestyle (869), least = Handball (1)	Lifestyle (869), Fitness (231), Running (119), Football (70), Skiing (103), Outdoor (85), Basketball (103), Other (5), Boxing (3), Tennis (12), Accessories (2), Handball (1), Volleyball (1), Skateboarding (1)
Prod_brand	mode = A brands (853), least = Other (74)	A brands (853), Licence (678), Other (74)
Prod_age	mode = For adults (1272), least = For all (23)	For adults (1272), For babies (0-4) (62), For teens (8-14) (127), For younger kids (4-10) (121), For all (23)
R	avg = 3.143 +/- 1.353	[1.000 ; 5.000]
F	avg = 3.616 +/- 3.401	[1.000 ; 17.000]
M	avg = 100.081 +/- 78.211	[9.600 ; 352.000]

3.2. Model Training and Validation

As the first step in model development, a cluster model was generated on the training set and customer segments are identified, i.e. a Customer Value-level (CV-level) for all customers is determined.

As one of the most well-known algorithms for cluster analysis, the k-means method was mostly used for customer segmentation in direct marketing [13–16]and other clustering analysis [59, 60]. This method estimates the centroid cluster model based on the Davies-Bouldin Index (DB) [61], which ensures maximum heterogeneity between clusters and maximum homogeneity within the clusters. DB index calculates the Euclidean distance from the centroid inside and between the clusters. Better quality of clustering is indicated by lower absolute values of the DB index. This study proposes the k-means method because the optimal number of clusters can be determined based on this indicator.

CV-level defines how much the customer is valuable to the company based on purchasing behavior or belonging to the appropriate segment. Thus, the segment of customers who buy most often, who bought the most recently and from whom the largest revenue was made, represents the segment of the most valuable customers for the company. All customers who belong to that segment get the best, that is. first CV level.

In the next step, Bagging SVM is trained, as the data preprocessor. On the training dataset the CV-level is then predicted by the preprocessor. In order to obtain the purest classes possible, with less overlap, and to achieve better class balance, only those results for which more than 90% of the SVM models voted in Bagging procedure are taken, i.e. results for which Confidence is > 0.9. In this way, an under-sampled training set is obtained with a new class label predicted by Bagging SVM. The new class label defines classes that overlap less and are more balanced.

A balanced Bagging DT model is trained on this Bagging SVM output. The model is now trained on much more balanced data and the balanced ensemble meta-algorithm further helps to solve the problem of the minor class (the most valuable customer class) and improves the performance of customer segments classification. In addition, the balanced Bagging DT model extracts rules that better describe customer segments, especially the minor one. Namely, solving the problem of the minor class, significantly more rules are obtained for the most important segment of the customers.

Thus, the final model for customer segments classification was created by combining an ensemble of SVM classifiers and an ensemble of DT rule extractors, so it can be called an ensemble SVM-RE model.

By training the model, the optimal combination of model's parameters is found, which achieves maximum predictive performance. This is attained by combining Grid-Search technique with k-fold cross-validation. More specifically, a grid of possible values is defined for parameters whose combinations are tested using the k-fold cross-validation procedure with stratified sampling. The cross-validation procedure implies that the starting data set is split into subgroups, taking care that percentage of class representation in subgroups corresponds to percentages of class representation in the entire set of data. Then k-1 subsets are used for training the model (training set), while one of the subsets is used for validation, i.e., testing how this model works on an unknown set of data (validation set). The procedure is repeated k times, so that each of the subsets is a validation set. At each iteration, the parameters for classification (accuracy rate, class precision, and class recall), are calculated and finally their average value is found.

For assessment of predictive performance, overall accuracy rate, class precision and class recall are used (these indicators are explained at the end of section 2.3). In addition to predict customer segment with high accuracy, for customer targeting it is important to classify existing consumers more accurately, so class recall is an important indicator of model performance.

3.3. Model Testing

In the testing phase, to assess the accuracy of the model at the test set, the actual CV-level primarily is determined using the cluster model generated in a training phase. Then the CV-level is predicted using the trained Bagging SVM model, while the test set retains examples whose predictions have Confidence > 0.9, and the predicted CV-level now is declared as the actual class label. The trained balanced Bagging DT model is then applied to this preprocessed test set, and thus CV-level predictions are obtained.

In the testing phase, the predictive performance of the model is determined by comparing the CV-level obtained by prediction using trained models with the actual CV-level values in the test set.

3.4. Summary of Predictive Procedure

Figure 1 shows a flow diagram for the training and testing phases of the predictive procedure. The procedure was implemented using Rapid Miner.

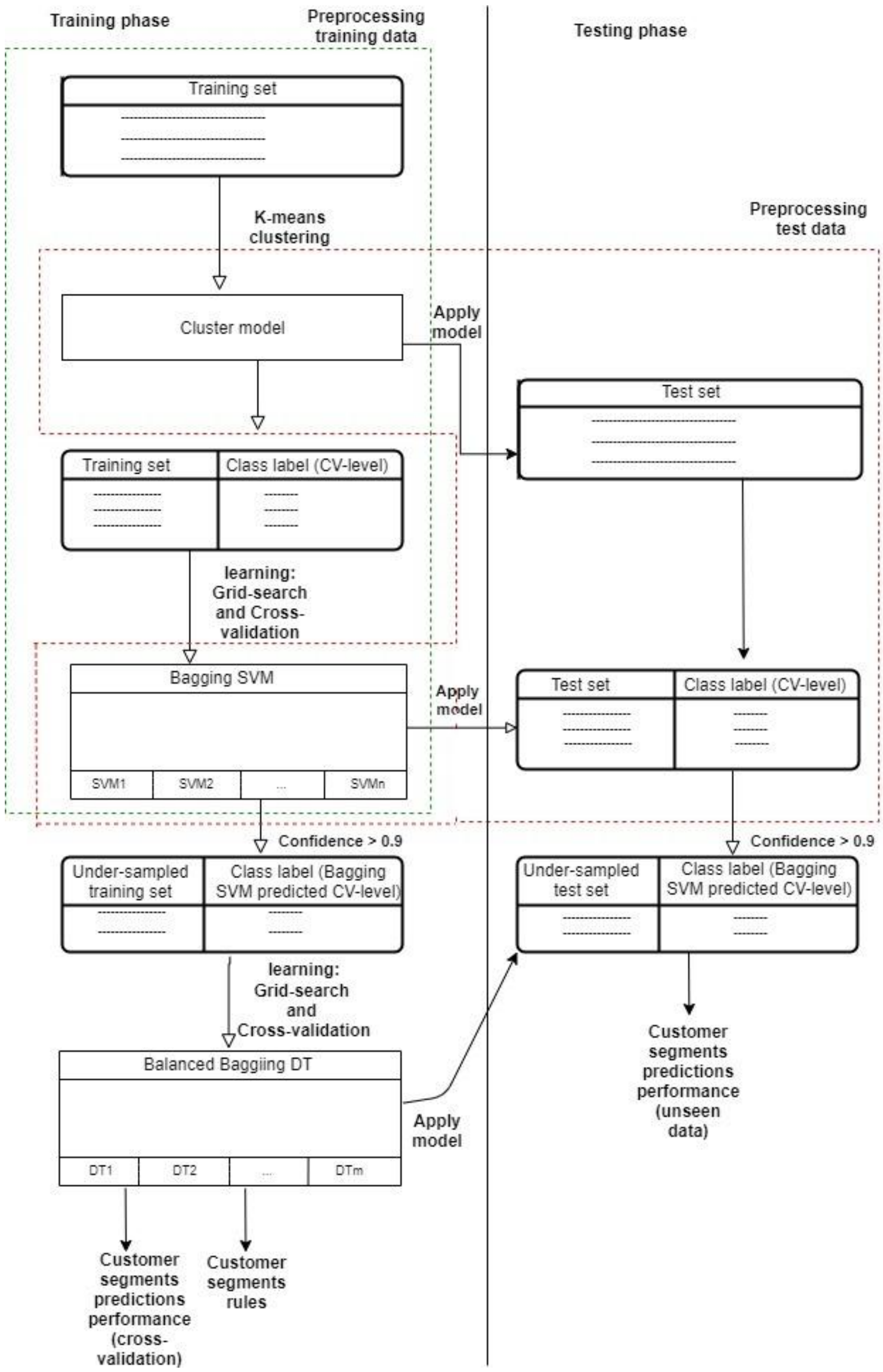

Fig. 1. Predictive procedure

4. Empirical Testing and Results

4.1. RFM Clustering of Training Data Set

By clustering the starting data set using k-means method and normalized RFM attributes, following results are obtained - shown in Table 2. It can be seen that the best DB index (minimum absolute value) is achieved for a 3-cluster model. This cluster model is shown in Table 3.

Table 2. Selection of number of clusters (parameter k) for k-means clustering

K	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB	-1.025	**-0.811**	-0.983	-0.958	-0.909	-1.02	-0.98	-0.976	-0.96

Table 3. Centroid Cluster model for RFM segmentation of customers

	R	**F**	**M**	**Items**
cluster_0	0.766807	0.565126	0.677733	238
cluster_1	0.735348	0.088065	0.179499	819
cluster_2	0.137318	0.101734	0.211345	548

Note: Normalized centroid values are shown (0-1 range transformation)

From Table 3, it can be seen that cluster_0 consists of the most recent, most frequent and most profitable customers (CV-Level=1), cluster_1 consists of recent, but less frequent and less profitable customers (CV-Level=2), while cluster_2 is made of non-recent customers, that are less frequent and less profitable (CV-Level=3). The most valuable customer cluster contains significantly less items than the other two clusters (238 versus 819 and 548), so the problem of class imbalance is evident.

Table 4 shows the distribution of customer frequency, recency and profitability across these CV-level segments. It can be observed that customers from the most valuable segment are recent, as well as that they have purchased on average 10 times in the considered period and their average amount of trade is around € 242. In contrast, customers from the CV-Level 3 segment, trade on average at most 3 times, with an average trading volume of around € 82.

Table 4. CV-Level customer segments

CV-Level	**Recency**	**Frequency**	**Monetary**
1	Avg: 4	Avg: 10	Avg: 241.65 €
	Min: 3	Min: 3	Min: 113.4 €
	Max: 5	Max: 17	Max: 352 €
2	Avg: 4	Avg: 2.4	Avg: 71.06 €
	Min: 3	Min: 1	Min: 9.6 €
	Max: 5	Max: 7	Max: 199.5 €
3	Avg: 1.54	Avg: 2.6	Avg: 81.96 €
	Min: 1	Min: 1	Min: 12.5 €
	Max: 2	Max: 9	Max: 239.5 €

4.2. Empirical Testing of Predictive Procedure

Training phase

In order to test proposed predictive procedure, a Bagging SVM model for the CV-level prediction obtained by initial customer clustering was first generated. Using Grid Search parameter optimization and 10-fold cross-validation, optimal combination of parameters for SVM and Bagging is defined as: SVM.C = 400.6, SVM.gamma = 200.006, Bagging.sample_ratio = 0.9, and Bagging.iterations = 10. The model then generated a CV-level prediction which is now taken as the class label of the training set.

In the next step, an under-sampled training set was made by excluding all predictions with confidence <= 0.9, i.e. those results for which only 90% of models and less voted in the Bagging SVM procedure. Table 5 shows the thus obtained new training set.

Table 5. Changes to the training set in the predictive procedure

Training set	Class label	Distribution of class label	Number of examples
Starting	CV-Level	CV-Level 1 (238) CV-Level 2 (819) CV-Level 3 (548)	1605
Obtained at the Bagging SVM output	Bagging SVM predicted CV-Level	CV-Level 1 (132) CV-Level 2 (981) CV-Level 3 (492)	1605
Under-sampled (Conf. > 0.9)	Bagging SVM predicted CV-Level	CV-Level 1 (82) CV-Level 2 (834) CV-Level 3 (360)	1276

Table 6. Results of the training phase (cross-validation performance)

Model	Accuracy	Class Recall	Class Precision
Bagging SVM[4]	61.00%	**21.01%[1]**, 81.07%[2], 48.36%[3]	**50.51%[1]**, 61.37%[2], 62.50%[3]
Balanced Bagging DT on Bagging SVM output[5]	88.71%	**69.51%**, 96.76%, 74.44%	**80.28%**, 87.91%, 93.38%
Standalone DT[6]	60.69%	**4.20%,** 80.34%, 55.84%	**27.78%,** 61.90%, 60.47%

[1] Class performance for CV-Level 1(most valuable customers - minor class);

[2] Class performance for CV-Level 2;

[3] Class performance for CV-Level 3 ;

[4] This model is a data preprocessor;

[5] The performance of this model is actually the performance of the final model for the customer segments classification called the ensemble SVM-RE ;

[6] Standalone DT is generated for comparison purposes.

Following that procedure, the balanced Bagging DT classifier was trained on the training set thus obtained. Grid Search and 10-fold cross-validation determined the optimal combination of parameters: Bagging.sample_ratio= 0.9, Bagging.iteration =

304, balancing_proportion: 82:500:100, split_criterion = gain_ratio, min_size_for_split = 4, min_leaf_size = 2, max_depth = 15, confidence = 0.2, min_gain = 0.01.

For the purpose of comparison, the DT standalone classifier was trained with the optimal combination of parameters: split_criterion = gini_index, min_size_for_split = 4, min_leaf_size = 16, max_depth = 15, confidence = 0.1, min_gain = 0.01.

The classification performance of the trained models are shown in Table 6.

From the Table 6, it can be noticed that the balance Bagging DT on Bagging SVM output model (hereinafter ensemble SVM-RE model) has significantly better classification performance than the standalone DT model. The standalone DT method correctly targeted only 4% of the most valuable customers, while ensemble SVM-RE successfully targeted 69% of them.

Also, all considered classification performances are better with the ensemble SVM-RE model than with DT. The class precision of the most valuable customers for DT is only 28%, which means that the company will have unnecessary campaign costs for 72% of wrongly classified customers. Precision of ensemble SVM-RE model for the class is 80%, which means that only 20% of the offers sent are likely to be unanswered. Therefore, ensemble SVM-RE will, in relation to DT, reduce the cost of the campaign. It can be concluded that, with the high overall accuracy of CV-level prediction (89%), the proposed ensemble SVM-RE method managed to solve the problem of class imbalance.

The results show that Bagging SVM as the preprocessor of data on purchase transactions eliminated noise, so that more precise classification is possible. The DT classification accuracy is increased by 28% - the accuracy for the standalone DT is 61% and for SVM-RE 89%. Mean class recall for standalone DT is 47%, and after data preprocessing and using the ensemble DT it is 80%. Mean class precision has increased from 50% to 87% after preprocessing.

Given the high cross-validation accuracy of rule extraction from the Bagging SVM output (89%), the rules validly interpret the Bagging SVM classification. Table 7 shows some of the 81 derived rules which are recognized as the most important, i.e. rules which cover a large number of examples (Support ~1% and more, except for the minor class where the minimum support is 0.3%), have high accuracy (Confidence > 80%) and good confidence in relation to overall data set (Lift > 1).

On the basis of derived rules, it can be stated that customers with CV-Level = 1 are mostly male customers, who mainly buy: basketball apparel for men from licensed brands (brands for which Sport Vision has licensed production and distribution, such as: Champion, Umbro, Lonsdale, Ellesse, Slazenger, Sergio Tacchini, etc.) and with a discount of 25% to 45%; apparel for adults – men, either for football or lifestyle category from licence brands with a discount between 25% and 35%; as well as men who purchase apparel for teens, with 25-45% discount, or equipment for women with 10-45% discount.

Customers of CV-Level = 2 are mainly women, who either mainly buy clothes from A brands (brands for which the company is a distributor, such as: Adidas, Nike, Under Armor, Reebok, Converse etc.) with a discount from 25% to 35%, or apparel from licensed brands and equipment on a discount from 25% to 45%. Additionally, women who purchase footwear for men on a 25% to 35% discount also belong to this customer segment. Male buyers in this category mainly purchase lifestyle apparel for adult men, either from licensed brands on 35% to 45% discount, or A brands from 25% to 35% discount.

CV-Level = 3 represents the group of least valuable customers. The customers belonging to this group mostly buy products on a discount larger than 45% (sale seekers). Male buyers in this category mainly buy lifestyle or equipment products from A brands. Female buyers in this segment purchase lifestyle footwear for adults from licensed brands.

Hence, with the SVM-RE ensemble, rules are obtained that explain customer segments, which is the answer to the RQ3. Also, solving the problem of the minor class provides a more efficient description of the segment of the most valuable customers with a larger number of important rules.

Table 7. Most significant classification rules derived by ensemble SVM-RE

CV-level	Rule	Confidence (>80%)	Support (>1%*)	Lift (>1)
CV-Level 1	if Discount > 45% and Prod_category = Football and Prod_age = For younger kids (4-10)	100%	0.3%	8.33
CV-Level 1	if Discount > 45% and Prod_category = Outdoor and Prod_type = Footwear and Prod_brand = Licence	100%	0.4%	8.33
CV-Level 1	if Discount ≤ 45% and Discount >>35% and Cons_gender = F and Prod_type = Apparel and Prod_brand = A brands and Prod_gender = For women	100%	0.3%	8.33
CV-Level 1	if Discount ≤ 45% and Discount > 25% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Basketball and Prod_brand = Licence	100%	0.4%	8.33
CV-Level 1	if Discount ≤ 35% and Discount > 25% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Football	100%	0.4%	8.33
CV-Level 1	if Discount ≤ 45% and Discount > 35% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Lifestyle and Prod_brand = A brands	100%	1%	8.33
CV-Level 1	if Discount ≤ 35% and Discount > 25% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Lifestyle and Prod_brand = Licence	100%	4%	8.33
CV-Level 1	if Discount ≤ 45% and Discount > 25% and Cons_gender = M and Prod_type = Apparel and Prod_age = For teens (8-14)	80%	1%	6.67
CV-Level 1	if Discount ≤ 45% and Discount > 10% and Cons_gender = M and Prod_type = Equipment and Prod_brand = Licence and Prod_gender = For women	100%	1%	8.33
CV-Level 2	if Discount > 45% and Prod_category = Skiing	86%	2%	1.17
CV-Level 2	if Discount ≤ 35% and Discount > 25% and Cons_gender = F and Prod_type = Apparel and Prod_brand = A brands	100%	6%	1.37

CV-Level 2	if Discount ≤ 45% and Discount > 25% and Cons_gender = F and Prod_type = Apparel and Prod_brand = Licence	100%	11%	1.37
CV-Level 2	if Discount ≤ 45% and Discount > 25% and Cons_gender = F and Prod_type = Equipment	94%	3%	1.29
CV-Level 2	if Discount ≤ 45% and Discount > 35% and Cons_gender = F and Prod_type = Footwear	98%	7%	1.34
CV-Level 2	if Discount ≤ 35% and Discount > 25% and Cons_gender = F and Prod_type = Footwear and Prod_gender = For men	100%	2%	1.37
CV-Level 2	if Discount ≤ 45% and Discount > 35% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Lifestyle and Prod_brand = Licence	100%	4%	1.37
CV-Level 2	if Discount ≤ 35% and Discount > 25% and Cons_gender = M and Prod_type = Apparel and Prod_age = For adults and Prod_gender = For men and Prod_category = Lifestyle and Prod_brand = A brands	100%	3%	1.37
CV-Level 2	if Discount ≤ 45% and Discount > 10% and Cons_gender = M and Prod_type = Footwear and Prod_age = For adults and Prod_category = Lifestyle	100%	8%	1.37
CV-Level 3	if Discount > 45% and Prod_category = Basketball	88%	1%	5.83
CV-Level 3	if Discount > 45% and Prod_category = Fitness and Cons_gender = M and Prod_brand = A brands	100%	1%	6.67
CV-Level 3	if Discount > 45% and Prod_category = Lifestyle and Prod_brand = A brands and Cons_gender = M	95%	3%	6.33
CV-Level 3	if Discount > 45% and Prod_category = Lifestyle and Prod_brand = Licence and Prod_age = For adults and Prod_type = Apparel and Prod_gender = For men	100%	1%	6.67
CV-Level 3	if Discount > 45% and Prod_category = Lifestyle and Prod_brand = Licence and Prod_age = For adults and Prod_type = Footwear and Cons_gender = F	100%	1%	6.67
CV-Level 3	if Discount > 45%and Prod_category = Running and Prod_brand = A brands	91%	2%	6.06
CV-Level 3	if Discount ≤ 25% and Discount > 10% and Cons_gender = M and Prod_type = Apparel and Prod_gender = For men	100%	1%	6.67

*note: the criteria for "Support" for chosen rules is ~1% and more, except for the minor class where the minimum support is 0.3%

Testing Phase

In order to determine the predictive performance of the model on the test set, the test set was clustered using a cluster model generated in the training phase (see Figure 1). In this way, each user is assigned an appropriate CV-level to be used to determine predictive performance in the test phase.

The next step is to preprocess the test set using a Bagging SVM trained in the training phase, as well as under-sampling the test set taking examples with a class label voted by more than 90% of the SVM models during the bagging procedure (see Figure 1). The characteristics of the test set before and after preprocessing are shown in Table 8.

Table 8. Changes to the test set in the testing phase

Test set	Class label	Distribution of class label	Number of examples
Starting	CV-Level	CV-Level 1 (286) CV-Level 2 (2906) CV-Level 3 (2027)	5219
Obtained at the Bagging SVM output	Bagging SVM predicted CV-Level	CV-Level 1 (491) CV-Level 2 (3399) CV-Level 3 (1329)	5219
Under-sampled (Conf. > 0.9)	Bagging SVM predicted CV-Level	CV-Level 1 (406) CV-Level 2 (3155) CV-Level 3 (1043)	4604

It is noted that the Bagging SVM complemented the first minor segment so that it has 491 instances after data preprocessing. In addition, this preprocessor has cleared overlaps between segments so that future classification is as accurate as possible.After under-sampling, the trained ensemble SVM-RE model was applied to this test data set and the results shown in Table 9 were obtained.

Table 9. Performance of ensemble SVM-RE model on unseen data

Model	Accuracy	Class Recall	Class Precision
Ensemble SVM-RE	85.79%	**93.84%**[1], 84.44%[2], 86.77%[3]	**79.38%**[1], 94.57%[2], 69.24%[3]

[1] Class performance for CV-Level 1 (most valuable customers - minor class);[2] Class performance for CV-Level 2; [3] Class performance for CV-Level 3;

From the table above it can be seen that the overall accuracy of the ensemble SVM-RE model on unseen data is 85.79% which is good, compared to cross-validation accuracy of 88.71%. Apart from maintaining similar overall accuracy as in model validation, the unknown data also shows good performance for the minor class (class precision of about 80% and class recall of about 94%), which is the most important because the existing and predicted potential most valuable customers are precisely identified.

The ensemble SVM-RE model targets a total of 480 most valuable customers for the campaign. Of these, 381 most valuable customers were correctly targeted (94% of all such customers in the test set) and 99 customers were missed. So about 80% of the offers sent are potentially profitable while for 20% could be in vain (see the confusion matrix shown in Table A1 in the Appendix).

So, as a result of the test phase, it can be concluded that the proposed ensemble SVM-RE model is a quality predictor for CV-level, i.e. adequate model for customer segment classification, so the answer to the RQ1 is positive.

4.3. Comparison with Other Class Balancing Methods

Due to the comparison of the proposed ensemble SVM-DT model with standalone ensemble models, with respect to the efficiency of solving the minor class problem, a combination of DT classifiers with different balanced ensemble techniques were tested. First, on the starting training dataset, a Balanced Bagging DT model was generated with parameters: Bagging.sample_ratio = 0.7, Bagging.iterations = 108, balancing_proportion: 238: 238: 238, criterion = gain_ratio, min_size_for_split = 4, min_leaf_size = 2, max_depth = 15, confidence = 0.2, min_gain = 0.01. Then a balanced AdaBoost DT model with parameters: Ada-Boost.iterations = 3001, balancing_proportion: 238: 238: 238, criterion = gain_ratio, min_size_for_split = 4, min_leaf_size = 2, max_depth = 15, confidence = 0.2, min_gain = 0.01, and finally a balanced Random Forest model with parameters: RandomForest.sample_ratio = 1.0, Random.Forest.iterations = 75, balancing_proportion: 238: 238: 238, criterion = gain_ratio, max_depth = 10, are generated. The optimal parameters were determined using Grid Search and 10-fold cross-validation.

The performance of the class balancing models are shown in Table 10.

Table 10. Classification performance of standalone balanced ensemble models

Model	Accuracy	Class Recall	Class Precision
		Cross-validation performance	
Balanced Bagging DT	56.69%	**44.12%**[1], 51.28%[2], 70.26%[3]	**34.20%**[1], 68.74%[2], 56.04%[3]
Balanced AdaBoost DT	55.32%	**41.60%**, 64.22%, 57.30%	**30.43%**, 69.44%, 57.46%
Balanced RandomForest	51.03%	**53.36%,** 41.76%, 63.87%	**28.93%**, 69.94%, 51.70%
		Performance on test set (unseen data)	
Balanced Bagging DT	49.55%	**26.22%**, 43.53%, 61.47%	**8.35%** , 66.20%, 51.70%
Balanced AdaBoost DT	46.14%	**35.66%**, 46.73, 46.77%	**7.53%**, 65.48%, 52.93%
Balanced RandomForest	44.51%	**34.62%,** 37.89%, 55.40%	**7.12%,** 67.01%, 51.37%

[1] Class performance for CV-level 1 (most valuable customers - minor class);[2] Class performance for CV-level 2; [3] Class performance for CV-level 3

Comparing the results of ensemble SVM-RE method from Table 6 with the standalone balanced ensemble methods Bagging DT, AdaBoost DT and Random Forest in Table 10, it can be concluded that this method outperforms their capabilities in terms of class balancing i.e. solutions to minor class problems. Namely, while for the ensemble SVM-DT the recall and precision for minor class were 69.51% and 80.28% respectively, the best recall of the minor class was achieved by Random Forest (53.36%) and the best precision for the minor class by balanced Bagging DT (34.20%). Also, the maximum overall accuracy of standalone balanced ensemble models (55.69%) is significantly smaller than the ensemble SVM-DT model (88.71%). When comparing the results at the test set (Table 9), the superiority of the ensemble SVM-RE models is even more pronounced.

Finally, it can be concluded that SVM, in combination with the Bagging ensemble meta-algorithm more effectively solves class imbalance problems than other methods used for this purpose.

4.4. Comparison by Profitability Criterion

For model comparisons in terms of the potentially achievable maximum profit from a campaign based on the minor class prediction (i.e. the segment of the most valuable customers), Table 11 shows the calculation of this indicator individually by models. The profit indicator is calculated by the formula (3):

$$Profit = True\ Predicted * R - (True\ Predicted + False\ Predicted) * C \qquad (3)$$

where are: *True Predicted* - number of model's true predicted customers of the most valuable segment; *R*- potential single customer revenue from a campaign; *False Predicted* - number of model's false predicted customers of the most valuable segment; and *C*-estimated campaign cost per single customer.

Table 11. Model comparison by potentially earnable campaign profit

Model	True Predicted[1]	False Predicted[2]	Revenues[3]	Costs[4]	Profit[5]
SVM-RE[6]	150	44	36300	194	36106
Ensemble SVM-RE	165	41	39930	206	39724
Balanced Bagging DT	105	202	25410	307	25103
Balanced AdaBoost DT	126	288	30492	414	30078
Balanced Random Forest	127	312	30734	439	30295

[1]Number of true predicted customers of the most valuable segment
[2] Number of false predicted customers of the most valuable segment
[3] True Predicted *Potential Single Customer Revenue (€ 242)
[4] (True Predicted+False Predicted) * Estimated Single Customer Campaign Cost (€ 1)
[5] Revenues-Costs
[6] Results for standalone SVM-RE are taken from [57]

Potential revenue is assumed to be average revenue generated in previous campaigns at the most valuable segment level (€ 242, see Table 4), while the estimated cost per campaign per customer is € 1. The number of correctly predicted and incorrectly predicted members of the most valuable customer segment is given in proportion to the participation of this class in the initial training set (since the training set obtained at the Bagging SVM output is under-sampled).

Based on the calculation in the table above, it is observed that the maximum profit can be expected based on the ensemble SVM-RE prediction. The improvement of the standalone SVM-RE method by the ensemble meta-algorithm may lead to an increase in profit in campaign for € 3618. From the standalone balanced ensemble method, the highest expected profit of € 30295 is achieved with the RandomForest prediction, which is € 9429 less than the expected profit with the ensemble SVM-RE prediction.

Thus, it can be concluded that the proposed ensemble SVM-RE model out-performs other considered models by profitability criterion. Note that the advantages of the ensemble SVM-RE method according to this criterion would be even more pronounced if the comparison was done on unseen data. Since the predictive accuracy on unseen data was not tested in [57], we compared the cross-validation performance of the models.

Taking into account the comparison according to the criterion of predictive performance from the previous section, as well as based on the criterion of profitability, ensemble SVM-RE is a better class balancing method than other considered methods, which is the answer to the RQ2.

4.5. Validation of the method by testing on a public dataset

Method was also tested on a publicly available *Customer_transaction_dataset*, available on *Kaggle* data science repository (available at: https://www.kaggle.com/archit9406/customer-transaction-dataset), which consists of data regarding cycling equipment sales. The data contains 20,000 sales transactions for 3,500 customers in the period from January to December 2017. The data were refined due to missing values, leaving 19765 items in the set, which were divided into training set (70%) and test set (30%). Recency was calculated based on the date of transactions, Monetary based on the total transactions value, and Frequency as the number of transactions in this period, the same way as in the original dataset. The distribution of these and pre-existing attributes in this dataset is shown in Table A2 in Appendix.

Data were first clustered based on RFM attributes and 3 clusters were obtained as the optimal solution (minimum DB index = -0.879) (Table 12).

Table 12. Centroid Cluster Model for the public dataset

	Recency	**Frequency**	**Monetary**	**Items**
Cluster 0	0.668	0.302	0.282	4264.000
Cluster 1	0.969	0.623	0.549	7034.000
Cluster 2	1.000	0.347	0.301	8467.000

Note: Normalized centroid values are shown (0-1 range transformation)

Cluster 1 of the most valuable customers (CV-Level = 1) contains 7034 items, cluster 2 of the medium valuable customers (CV-Level = 2) has 8467 items, while cluster 0 of the least valuable customers (CV-Level = 3) in this case is the smallest and has 4264 customers. Obviously, the problem of unbalanced classes is also present here.

Repeating the same predictive procedure defined in Figure 1, in the training phase by cross-validation and the test phase by testing on an unknown dataset, the results shown in Table 13 were obtained.

Table 13. Predictive performance of the models for the public dataset

Model	Parameters	Cross-Validation Performance	Test Performance
Bagging SVM	SVM.gamma = 0.0325 SVM. C = 1000.0 Bagging.iterations = 10 Bagging.sample_ratio = 0.8	accuracy: 46.15% class recall: 21.94%[1], 44.25%[2], 59.93%[3] class precis.: 33.87%, 46.65%, 49.12%	accuracy: 47.29% class recall: 23.69%[1], 46.16%[2], 60.12%[3] class precis.: 37.04%, 47.56%, 49.85%
Ensemble SVM-RE	DT.criterion = gain_ratio DT.min_size_for_split = 4 DT.minimal_leaf_size = 2 DT.maximal_depth = 15 DT.confidence = 0.1 DT.minimal_gain = 0.01 Bagging.sample_ratio = 0.9 Bagging.iterations = 100 Bagging.Balancig_proportion: 435:1000:2000	accuracy: **88.69%** class recall: **61.61%,** 85.50%, 93.74% class precis.: 74.24%, 88.71%, 90.03%	**accuracy: 90.32%** class recall: **70.07%**, 86.67%, 94.55% class precis.: 72.03%, 91.57%, 91.41%
Standalone DT	DT.criterion = gain_ratio DT.min_size_for_split = 4 DT.minimal_leaf_size = 2 DT.maximal_depth = 15 DT.confidence = 0.1 DT.minimal_gain = 0.01	accuracy: **43.01%** class recall: **0.50%,** 0.35%, 99.87% class precis.: 78.95%, 80.95%, 42.90%	**accuracy: 43.08%** class recall: **0.55%**, 0.33%, 100% class precis.: 100%, 100%, 42.94%

[1] Class performance for CV-Level 3 (minor class); [2] Class performance for CV-Level 1; [3] Class performance for CV-Level 2

The data in the table above indicate that the Ensemble SVM-RE successfully solved the problem of incorrect classification of the minor class on this data set as well. On the training and test set, standalone DT completely misclassified the best customers (class recall is only about 0.3%) and the worst customers (class recall about 0.5%), and the overall accuracy of the model is about 43%. The accuracy of the Ensemble SVM-RE model on unknown data is about 90%, class recall for the best customers about 87% and for the least valuable cluster about 70%. Bagging SVM has a class recall below 50%, not only for the minor (least valuable) class, but also for the non-minor most valuable class. This means that in this data set, besides the problem of the minor class, the problem of class overlap (noise) also exists, which Bagging SVM preprocessor has solved successfully.

5. Discussion

The proposed model aimed to test several improvements of existing methods for predictive classification of customers in direct marketing, such as objective segmentation of customers with an indicator for the optimal number of clusters, description of segments in terms of customer characteristics and products, prediction of value for new customers with unknown purchasing behavior, and finally and most importantly, the reduction of misclassification for the segment of the most valuable customers, i.e. solution of class imbalance problem.

Unlike some previous studies that use hard coding of RFM attributes, sorting based on coded values and subjective selection of segments for the campaign [4, 11, 12], this study suggests a more sophisticated and objective data mining technique - k-means clustering, which achieves segmentation algorithmically using the measure of Euclidean distance, in order to provide maximum similarity within segments and difference between segments. Instead of uniform coding of RFM attributes, which does not treat the customer individually, but identifies them with the group to which they belong, which is a characteristic of many previous studies [13, 14, 16], clustering by un-coded attributes is proposed in this study, because there are numerous values with which the algorithm for clustering works smoothly. This prevents the loss of important information at the level of each individual customer, that may distinguish them from others. Unlike the method proposed in [16], which involves subjective evaluation and testing of the best number of clusters, our method determines the optimal number of clusters objectively based on the DB index, which significantly simplifies the procedure and ensures the accuracy of the model.

Classical RFM segmentation involves the prediction of future customer behavior based only on these three attributes, and is not applicable to the prospecting for new customers because transaction information is not available [4]. In [16], sophisticated data mining techniques are used during customer segmentation, but in addition to customer characteristics, RFM attributes are included as predictive attributes, so the proposed model cannot be used for new customers for whom these attributes are unknown. In our study, only product data and customer characteristics are used as predictive attributes, as it is expected to obtain predictive rules with more suitable information for targeting the potential customers [51].

Furthermore, for predictive customer classification, this study suggests the SVM-RE method in combination with ensemble techniques that enhance the predictive power of the model. The results showed that the SVM ensemble efficiently preprocesses the data, i.e. resolves the noise and class imbalance. First, by moving the margin to the nearest (and therefore most similar) examples of the larger class and classifying them into the smaller class, SVM resolves the noise in the data, i.e. class overlapping and complements the minor class with the most relevant examples. Then, by pooling the results of multiple SVMs in the ensemble procedure, the instances that join the minor class are identified more precisely (the example joins the class that has been voted the most by the SVM model). And in the end, taking only the results for which more than 90% of the SVM models voted, representatives of the classes most likely to belong to the class are selected, i.e. those that are farthest from each other and between which the margin is the widest, leading to maximum separation of classes. Applying a balanced DT ensemble for rule extraction from such pre-processed data set (SVM-RE ensemble) misclassification rate of the most valuable customer segment is reduced by 66%, which is a much better result than the result obtained in [58], where using standalone SVM preprocessor and standalone DT rule extractor this misclassification rate was reduced by 37%.

For the test set, ensemble SVM-RE method achieved Balanced Correction Rate (BCR) (rooted product of class recall of all classes) of 83%, which is 15% better than the best achieved in [22] by applying under-sampling based on clustering and ensemble techniques. Comparing the best class recall of minority class (88%), obtained in [29] using cluster-based under-sampling and k-NN classifiers, with the result achieved by our method (94%), the superiority of our model is obvious. Additionally, the class recall for majority class in [29] is low (63%), while with our method for the other two larger classes it is above 84%. In [17] the best achieved class recall for minority class is 73%, for dataset with moderate degree of class imbalance, and with random under-sampling for class ratio of 2:1, using SVM classifier, which is again lower than our score of about 94% on the test set.

Apart from the proven efficiency, automatic class balancing using the SVM-RE ensemble is less complex for practical application (there are no unknowns regarding the choice of examples to be removed, choice of optimal class ratio, etc.) compared to resampling techniques in similar studies in direct marketing [17, 18, 22, 29, 30]. Balancing the data in this way offers a stable solution that does not suffer from the sampling bias and overfitting that can occur due to resampling [31].

The ensemble SVM-RE method had a misclassification of the most valuable customer segment of about 6% at the test set, which is an excellent result. A similar result, i.e. a misclassification rate of 4% was achieved in [27], where a method based on adapting the SVM algorithm by introducing cost-sensitive learning and random under-sampling was used. However, the advantage of our method is that its application in practice does not require extensive knowledge of the SVM method required for its adaptation.

Comparing the results of standalone SVM-RE method from [57]and the ensemble SVM-RE, it can be seen that the ensemble approach was able to improve overall accuracy by 2.98%, recall for minor class by 6.81%, as well as precision for minor class by 2.83%. While these improvements seem small, considering that identifying the most valuable customers has improved by about 7% and their prediction by about 3%, it can bring about a big increase in the profits generated by the campaign (see Table 11). It should be borne in mind that once an accurately selected or predicted high-profit customer can generate more revenue in a campaign than all other customers combined. Additionally, method was not tested in unseen data in [57], hence, its true predictive power remained unexplored.

The method was additionally tested on a publicly available data set where its superiority was confirmed. The overall accuracy of classification on unknown data was improved from 43% (held by standalone DT) to 90%. The SVM-RE ensemble method at the test set had a misclassification of the most valuable customer segment of about 13%, unlike the standalone DT which had a misclassification of as much as 97% due to the overlap of this segment with the middle value customer segment. As for the problem of the minor class, i.e. least valuable customers in this case, the SVM ensemble reduced its misclassification error from 94.5% to 29% on unknown data. Thus, the method successfully solved the problem of imbalance and class overlap on the validation data set as well.

Contributions to theory/knowledge/literature

Given the above comparison and the highlighted advantages of the proposed ensemble SVM-RE method in relation to previously applied methods, it can be concluded that this study contributes to the existing theory and knowledge in the field of predictive analytics in direct marketing in several ways:

1. Instead of judgment based RFM segmentation, objective k-means based RFM segmentation and estimation of the optimal number of clusters based on DB index is proposed, which simplifies the application and guarantees higher accuracy of the model.
2. Instead of classifying customers into uniformly coded groups, clustering is performed at the level of an individual customer, thus preventing the loss of significant segmentation information.
3. Instead of RFM attributes, customer and product characteristics are used as predictors, so the method can also be used to classify unknown customers.
4. Instead of resampling or adapting the learning algorithm, it is proposed to automatically balance and remove class overlaps using the ensemble SVM method, which leads to a stable solution free of sampling bias, overfitting and extensive knowledge of the learning method by marketers.
5. Instead of preprocessing data using standalone SVM, an ensemble SVM has been proposed that increases the efficiency of balancing and class separation.
6. Instead of rule extraction using the standalone classifier, this study suggests rule extraction using the DT classifier combined with a balanced ensemble meta-algorithm which gives better predictive performance, compared to using standalone DT as the rule extractor.
7. The proposed ensemble SVM-RE method has a smaller misclassification of minority class (segment of the most valuable customers) than the standalone balanced ensemble method, as well as the methods of resampling and adaptation of algorithms used in previous studies, while maintaining high overall accuracy.
8. The proposed method extracts rules that effectively describe user segments (including the smallest one with the most valuable customers, for which rules may be omitted if the minority class misclassification is not addressed). These rules are semantically richer because they contain customer and product characteristics, and are more suitable for targeting existing and new customers in the campaign.
9. Unlike most previous studies, the method is tested on a real-life data set in this study that has not been refined and specially prepared for analysis.Then, the method was validated by testing on a publicly available dataset.

Implications for practice

In addition to the theoretical contribution, the proposed method is important for practical applications and can significantly help marketers in planning direct campaigns. A very creative and innovative offer can result in a low response rate if the targeting is not done precisely, while, on the other hand, a poorly formulated and medium creative offer to the right target group can reduce, but not eliminate, the desired consumer response [62]. Therefore, understanding the preferences and needs of consumers is a more important factor in creating a campaign, than the creative process and the way of communicating the offer. In addition, business intelligence and data mining can enhance the competitive advantage for the companies in contemporary markets [63]. This is in line with the current and ongoing trend of digital transformation in companies, conducted with the aim of keeping up with the competition [64] and improving customer experience [65].

Using the proposed model, it is possible to overcome the impersonal nature of traditional marketing, as it allows companies to treat similar groups of customers in a unique way. The benefits that this model provides to practitioners are reflected through precise targeting, minimizing message waste, and more profitable campaigns. In this way, they are enabled to objectively segment the market, adapt the content to individual segments, and to build a reliable and loyal relationship with customers. In this paper, it is shown that using ensemble SVM-RE model prediction for the most valuable segment results in the highest number of true predicted customers, as well as the lowest number of false predicted customers of that segment. In that sense, this proposed model in direct marketing practice can achieve the highest profit, compared to other considered models and reduce the waste of marketing resources.

Based on the insights from this predictive model, a more elaborate segmentation strategies can be created and more effective targeting can be applied. The rules extracted from our model enable marketers to learn about their most valuable consumers, which is of high importance, having in mind that keeping the current customers is often six to ten times more cost-effective than acquiring new ones [66, 67]. Additionally, explicit rules that describe the most valuable consumers allows for acquisition of precisely those customers that are the most similar to this group, through various targeting strategies. Hence, customers from different clusters and of different value for the company can be targeted in customized and tailored promotional activities. In other words, targeting can be conducted in an objective and precise manner, which improves the profitability of each campaign, as well as the overall effectiveness of direct marketing activities.

Another advantage for the practitioners of direct marketing is the ease of use of our method. There is no need for complex resampling procedures to be carried out, since automatic data balancing is used. Also, practitioners do not have to know the details of the learning algorithm or hire additional experts for that purpose.

6. Conclusions

In this paper an efficient method for customer classification in direct marketing is designed. The presented predictive procedure implies the classification of customer clustering. Using the k-means clustering, customers are divided into segments based on their RFM attributes (past purchasing behavior). Different clusters have different customer value levels, as well as different probability of responding to a marketing campaign. Following this procedure, customer's affiliation with one of the clusters (as well as consumer's appropriate CV-level) is predicted using the ensemble SVM-RE method, using the data on the purchased products and the customer characteristics.

The results of our empirical testing indicate that the class imbalance problem can be overcome, which improves the classification of the minority, and most valuable class.

Combining multiple SVM models with an ensemble meta-algorithm can improve data preprocessing and separate customer segments more efficiently than standalone SVM. Applying balanced ensemble classifiers on such a preprocessed training set improves the predictive indicators for the smaller class and, consequently, the effectiveness of predictive segmentation (especially for the segment of the most valuable customers), as well as the chances of making greater profits in the campaign. Combining ensemble method, based on random (with replacements) under-sampling of larger classes, i.e. on bootstrapping, with SVM data preprocessing and rule extraction, balances classes better than standalone balanced ensemble methods, in customer segment classification.

The main contribution of this study is that the proposed method better deals with the problem of class imbalance that occurs when classifying customers in direct marketing, than the methods of resampling and algorithm adaptation applied in previous papers in this field. Namely, in comparison with the previous results, a smaller misclassification of the minority class (segment of high-value customers) with high overall accuracy was achieved. The class balancing procedure is automated by data preprocessing, thus overcoming the shortcomings of previously applied methods (sampling bias, overfitting, the need for extensive knowledge of learning methods). Ultimately, application is simplified.

In addition to the scientific contribution, this study is of practical importance because the proposed method can significantly help marketers to increase the efficiency and profitability of direct campaigns and to maintain good customer relations. The results of the method can help decide if existing (new) customers should be targeted in the following direct marketing campaigns, as two key elements of the customer relationship management are customer attraction and retention [68]. This method draws out and generates classification rules, which can be used in improving relationships with existing customers and targeting new potential customers, based on their characteristics and the products offered. Ultimately, the method can notably increase the campaign revenues, as well as decrease its costs.

However, this study also has several limitations and drawbacks. First, training sets with relatively small number of instances were used in this study. For a large training set, training of SVM learners, i.e. setting the appropriate parameters, requires high computation time [33]. Secondly, as a preprocessor, SVM is combined only with Bagging, although it is possible that it would balance classes better with some other ensemble technique. Third, the proposed model was tested on only one real data set, so it is unknown what the results would be on another set with a different class distribution. Fourth, as a rule extractor from the SVM ensemble of the preprocessed dataset, only a combination of Bagging and DT classifiers was tested. The bagging technique uses random under-sampling with replacement, which may be less effective than some other techniques such as cluster-based under-sampling. Thus, the question remains whether the rule extraction would yield better results with an ensemble using cluster under-sampling as in [22, 56], or by combining ordinary cluster under-sampling with different classifiers as in [22, 29], as well as by combining some other ensemble technique (e.g. Adaptive Boosting) with different classifiers similar to [30]. In the end, the success of the data mining method largely depends on the quality of the data. The data set used here includes only some customer characteristics. By including more customer attributes, clearer rules for targeting new customers can be obtained.

In future research, this method can be tested on other data sets to verify or improve its efficiency. In this study, the method was tested in the classification of customer segments where the minority class share is about 15%. It would be interesting to test its performance in the customer response model where minority class participation can be significantly lower, even below 5%.

To extract the rules from the ensemble SVM preprocessed dataset, some other ensemble techniques could be tested, such as Random Forest, as well as algorithm-level techniques. Since cluster-based under-sampling was previously confirmed in the literature as a successful class balancing technique [22, 29], a combination of this technique could be tested as a rule extractor (independently or within an ensemble procedure) with different classifiers. Also, although in this study Bagging SVM was confirmed as a preprocessor that successfully balances data from the domain of direct marketing, in future research its results could be compared with cluster-based under-sampling on the same data set. It would be useful to test whether some other ensemble technique, such as Adaptive Boosting, in combination with SVM, would pre-process better, i.e. balance the data more effectively.

References

1. Jonker, J.J., Piersma, N., Van Den Poel, D.: Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. Expert Syst. Appl. 27, 159–168 (2004). https://doi.org/10.1016/j.eswa.2004.01.010
2. Kaymak, U.: Fuzzy target selection using RFM variables. Annu. Conf. North Am. Fuzzy Inf. Process. Soc. - NAFIPS. 2, 1038–1043 (2001). https://doi.org/10.1109/nafips.2001.944748
3. Hughes, A.M.: Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program. McGraw-Hill (2005)
4. McCarty, J.A., Hastak, M.: Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. J. Bus. Res. 60, 656–662 (2007). https://doi.org/10.1016/j.jbusres.2006.06.015
5. Olson, D.L., Cao, Q., Gu, C., Lee, D.: Comparison of customer response models. Serv. Bus. 3, 117–130 (2009). https://doi.org/10.1007/s11628-009-0064-8
6. Olson, D.L., Chae, B.: Direct marketing decision support through predictive customer response modeling. Decis. Support Syst. 54, 443–451 (2012). https://doi.org/10.1016/j.dss.2012.06.005
7. Cui, G., Wong, M.L., Wan, X.: Targeting High Value Customers While Under Resource Constraint: Partial Order Constrained Optimization with Genetic Algorithm. J. Interact. Mark. 29, 27–37 (2015). https://doi.org/10.1016/j.intmar.2014.09.001
8. Kim, D., Lee, H. joo, Cho, S.: Response modeling with support vector regression. Expert Syst. Appl. 34, 1102–1108 (2008). https://doi.org/10.1016/j.eswa.2006.12.019
9. Otter, P.W., van der Scheer, H., Wansbeek, T.: Optimal selection of households for direct marketing by joint modeling of the probability and quantity of response. (2006)
10. Malthouse, E.: Ridge Regression and Direct Marketing Scoring Models. J. Interact. Mark. 13, 19–23 (1999)
11. Hughes, A.M.: Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program. Irwin, Chicago (1994)
12. Drozdenki, R., Drake, P.: Optimal database marketing. Sage Publications, Thousand Oaks, CA (2002)
13. Hosseini, S.M.S., Maleki, A., Gholamian, M.R.: Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Syst. Appl. 37, 5259–5264 (2010). https://doi.org/10.1016/j.eswa.2009.12.070

14. Sarvari, P., Ustundag, A., Takci, H.: Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. Kybernetes. 45, 1129–1157 (2016)
15. Khalili-Damghani, K., Abdi, F., Abolmakarem, S.: Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. Appl. Soft Comput. J. 73, 816–828 (2018). https://doi.org/10.1016/j.asoc.2018.09.001
16. Cheng, C.H., Chen, Y.S.: Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst. Appl. 36, 4176–4184 (2009). https://doi.org/10.1016/j.eswa.2008.04.003
17. Kim, G., Chae, B.K., Olson, D.L.: A support vector machine (SVM) approach to imbalanced datasets of customer responses: Comparison with other customer response models. Serv. Bus. 7, 167–182 (2013). https://doi.org/10.1007/s11628-012-0147-9
18. Miguéis, V.L., Camanho, A.S., Borges, J.: Predicting direct marketing response in banking: comparison of class imbalance methods. Serv. Bus. 11, 831–849 (2017). https://doi.org/10.1007/s11628-016-0332-3
19. Huang, K., Yang, H., King, I., Lyu, M.R.: Learning classifiers from imbalanced data based on biased minimax probability machine. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2, (2004). https://doi.org/10.1109/cvpr.2004.1315213
20. Wu, G., Chang, E.Y.: Class-boundary alignment for imbalanced dataset learning. ICML Work. Learn. from Imbalanced Data Sets II. 49–56 (2003)
21. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)
22. Kang, P., Cho, S., MacLachlan, D.L.: Improved response modeling based on clustering, under-sampling, and ensemble. Expert Syst. Appl. 39, 6738–6753 (2012). https://doi.org/10.1016/j.eswa.2011.12.028
23. Dietterich, T.G.: Ensemble learning. Handb. brain theory neural networks. 2, 110–125 (2002)
24. Zhang, C., Ma, Y.: Ensemble machine learning: methods and applications. Springer Science & Business Media, Boston, MA (2012)
25. Farquad, M.A.H., Bose, I.: Preprocessing unbalanced data using support vector machine. Decis. Support Syst. 53, 226–233 (2012). https://doi.org/10.1016/j.dss.2012.01.016
26. Barakat, N., Bradley, A.P.: Rule extraction from support vector machines: A review. Neurocomputing. 74, 178–190 (2010). https://doi.org/10.1016/j.neucom.2010.02.016
27. Zou, P., Hao, Y., Li, Y.: Customer value segmentation based on cost-sensitive learning support vector machine. Int. J. Serv. Technol. Manag. 14, 126–137 (2010). https://doi.org/10.1504/IJSTM.2010.032888
28. Al-Rifaie, M.M., Alhakbani, H.A.: Handling class imbalance in direct marketing dataset using a hybrid data and algorithmic level solutions. Proc. 2016 SAI Comput. Conf. SAI 2016. 446–451 (2016). https://doi.org/10.1109/SAI.2016.7556019
29. Marinakos, G., Daskalaki, S.: Imbalanced customer classification for bank direct marketing. J. Mark. Anal. 5, 14–30 (2017). https://doi.org/10.1057/s41270-017-0013-7
30. Lawi, A., Velayaty, A.A., Zainuddin, Z.: On identifying potential direct marketing consumers using adaptive boosted support vector machine. Proc. 2017 4th Int. Conf. Comput. Appl. Inf. Process. Technol. CAIPT 2017. 2018-Janua, 1–4 (2018). https://doi.org/10.1109/CAIPT.2017.8320691
31. Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: A review. Int. J. Pattern Recognit. Artif. Intell. 23, 687–719 (2009). https://doi.org/10.1142/S0218001409007326
32. Bhadani, A., Shankar, R., Vijay Rao, D.: A computational intelligence based approach to telecom customer classification for value added services. Adv. Intell. Syst. Comput. 201 AISC, 181–192 (2013). https://doi.org/10.1007/978-81-322-1038-2_16

33. Cui, D., Curry, D.: Prediction in marketing using the support vector machine. Mark. Sci. 24, 595–615 (2005). https://doi.org/10.1287/mksc.1050.0123
34. Vapnik, V.N.: The nature of statistical learning theory. Springer, New York (2010)
35. Sanderson, M.: Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. Nat. Lang. Eng. 16, 100–103 (2010)
36. Martens, D., Huysmans, J., Setiono, R., Vanthienen, J., Baesens, B.: Rule extraction from support vector machines: An overview of issues and application in credit scoring. Stud. Comput. Intell. 80, 33–63 (2008). https://doi.org/10.1007/978-3-540-75390-2_2
37. Diederich, J.: Rule Extraction from Support Vector Machines: An Introduction. In: Diederich, J. (ed.) Rule Extraction from Support Vector Machines. Studies in Computational Intelligence, vol 80. pp. 3–31. Springer, Berlin, Heidelberg (2008)
38. Martens, D., Baesens, B., Gestel, T. Van, Vanthienen, J.: Comprehensible Credit Scoring Models Using Rule Extraction From Support Vector Machines Credit Risk Modelling , Group Risk Management , Dexia Group. Decis. Sci. 1–21
39. Quinlan, J.R.: Induction of decision trees. Mach. Learn. 1, 81–106 (1986)
40. Quinlan, J.R.: C4.5 - programs for machine learning. Kaufmann, San Mateo, CA (1992)
41. Breiman, L.: Classification and regression trees. Wadsworth International Group, Belmont, CA (1984)
42. Kašćelan, L., Kašćelan, V., Jovanović, M.: Hybrid support vector machine rule extraction method for discovering the preferences of stock market investors: Evidence from Montenegro. Intell. Autom. Soft Comput. 21, 503–522 (2015). https://doi.org/10.1080/10798587.2014.971500
43. Breiman, L.: Bagging Predictors, URL: https://link.springer.com/article/10.1007%2FBF00058655. Mach. Learn. 24, 123–140 (1996). https://doi.org/10.1007/BF00058655
44. Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 55, 119–139 (1997). https://doi.org/10.1006/jcss.1997.1504
45. Breiman, L.: Random Forests. Mach. Learn. 45, 5–32 (2001)
46. Gupta, A., Gupta, G.: Comparative study of random forest and neural network for prediction in direct marketing. Springer Singapore (2019)
47. Lessmann, S., Haupt, J., Coussement, K., De Bock, K.W.: Targeting customers for profit: An ensemble learning framework to support marketing decision-making. Inf. Sci. (Ny). (2019). https://doi.org/10.1016/j.ins.2019.05.027
48. Wang, C.H.: Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques. Expert Syst. Appl. 37, 8395–8400 (2010). https://doi.org/10.1016/j.eswa.2010.05.042
49. Marshall, P.: The 80/20 Rule of Sales: How to Find Your Best Customers
50. Hsieh, N.C.: An integrated data mining and behavioral scoring model for analyzing bank customers. Expert Syst. Appl. 27, 623–633 (2004). https://doi.org/10.1016/j.eswa.2004.06.007
51. Tsai, C.Y., Chiu, C.C.: A purchase-based market segmentation methodology. Expert Syst. Appl. 27, 265–276 (2004). https://doi.org/10.1016/j.eswa.2004.02.005
52. MacQueen, J.: Some methods for classification and analysis of multivariate observations. Proc. fifth Berkeley Symp. Math. Stat. Probab. 1, 281–297 (1967)
53. Bose, I., Chen, X.: Quantitative models for direct marketing: A review from systems perspective. Eur. J. Oper. Res. 195, 1–16 (2009). https://doi.org/10.1016/j.ejor.2008.04.006
54. Liu, J., Zio, E.: Integration of feature vector selection and support vector machine for classification of imbalanced data. Appl. Soft Comput. J. 75, 702–711 (2019). https://doi.org/10.1016/j.asoc.2018.11.045
55. Lopez-Garcia, P., Masegosa, A.D., Osaba, E., Onieva, E., Perallos, A.: Ensemble classification for imbalanced data based on feature space partitioning and hybrid

metaheuristics. Appl. Intell. 49, 2807–2822 (2019). https://doi.org/10.1007/s10489-019-01423-6
56. Wong, M.L., Seng, K., Wong, P.K.: Cost-sensitive ensemble of stacked denoising autoencoders for class imbalance problems in business domain. Expert Syst. Appl. 141, (2020)
57. Rogic, S., Kascelan, L.: Customer Value Prediction in Direct Marketing Using Hybrid Support Vector Machine Rule Extraction Method. Commun. Comput. Inf. Sci. 1064, 283–294 (2019). https://doi.org/10.1007/978-3-030-30278-8_30
58. Djurisic, V., Kascelan, L., Rogic, S., Melovic, B.: Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method. Appl. Artif. Intell. 00, 1–15 (2020). https://doi.org/10.1080/08839514.2020.1790248
59. Kramarić, T.P., Bach, M.P., Dumičić, K., Žmuk, B., Žaja, M.M.: Exploratory study of insurance companies in selected post-transition countries: non-hierarchical cluster analysis. Cent. Eur. J. Oper. Res. 26, 783–807 (2018). https://doi.org/10.1007/s10100-017-0514-7
60. Bach, M.P., Juković, S., Dumičić, K., Šarlija, N.: Business client segmentation in banking using self-organizing maps. South East Eur. J. Econ. Bus. 8, 32–41 (2014). https://doi.org/10.2478/jeb-2013-0007
61. Davies, D.L., Bouldin, D.W.: A Cluster Separation Measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1, 224–227 (1979). https://doi.org/10.1109/TPAMI.1979.4766909
62. Stone, B., Jacobs, R.: Successful Direct Marketing Methods. McGraw Hill (2008)
63. Bach, M.P., Jaklič, J., Vugec, D.S.: Understanding impact of business intelligence to organizational performance using cluster analysis: Does culture matter? Int. J. Inf. Syst. Proj. Manag. 6, 63–86 (2018). https://doi.org/10.12821/ijispm060304
64. Furjan, M.T., Tomičić-Pupek, K., Pihir, I.: Understanding Digital Transformation Initiatives: Case Studies Analysis. Bus. Syst. Res. 11, 125–141 (2020). https://doi.org/10.2478/bsrj-2020-0009
65. Rekettye, G., Rekettye, G.: The Effects of Digitalization on Customer Experience. SSRN Electron. J. 340–346 (2019). https://doi.org/10.2139/ssrn.3491767
66. Colgate, M.R., Danaher, P.J.: Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. J. Acad. Mark. Sci. 28, 375–387 (2000). https://doi.org/10.1177/0092070300283006
67. Verbeke, W., Martens, D., Mues, C., Baesens, B.: Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Syst. Appl. 38, 2354–2364 (2011). https://doi.org/10.1016/j.eswa.2010.08.023
68. Lazar, E.: Customer Churn Prediction Embedded in an Analytical CRM Model. SSRN Electron. J. 24–30 (2018). https://doi.org/10.2139/ssrn.3281605

Ljiljana Kašćelan is a full-time professor at the Faculty of Economics at the University of Montenegro. She graduated in Computer Science at the Faculty of Natural Sciences and Mathematics, obtained a M.Sc. in Computer Science from the Faculty of Electrical Engineering and a Ph.D. in Business Intelligence from the Faculty of Economics (all at the University of Montenegro). She teaches courses in Business Informatics and Business Intelligence. Ljiljana is the author of several papers in international journals and conferences in the field of data mining and applications.

Suncica Rogić, MSc, is a teaching assistant at the Faculty of Economics at the University of Montenegro. She graduated and obtained a MSc from the same Faculty, and currently is a PhD candidate. The topics of her research interest include direct and digital marketing analytics, customer segmentation and data mining applications.

Received: May 30, 2020; Accepted: December 02, 2020.