Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis

Zhiyuan Yao, Annika H. Holmbom, Tomas Eklund, and Barbro Back

Åbo Akademi University and Turku Centre for Computer Science (TUCS), Joukahainengatan 3-5A, 20520 Turku, Finland
{Zhiyuan.Yao,Annika.H.Holmbom,Tomas.Eklund, Barbro.Back}@abo.fi

Abstract. Leveraging the power of increasing amounts of data to analyze customer base for attracting and retaining the most valuable customers is a major problem facing companies in this information age. Data mining technologies extract hidden information and knowledge from large data stored in databases or data warehouses, thereby supporting the corporate decision making process. In this study, we apply a two-level approach that combines SOM-Ward clustering and decision trees to conduct customer portfolio analysis for a case company. The created two-level model was then used to identify potential high-value customers from the customer base. It was found that this hybrid approach could provide more detailed and accurate information about the customer base for tailoring actionable marketing strategies.

Keywords: Customer relationship management (CRM), customer portfolio analysis (CPA), Self-organizing maps (SOM), Ward's clustering, decision trees.

1 Introduction

For a long time, the focus of modern companies has been shifting from being product-oriented to customer-centric organizations. In the industry it is commonly held that maintaining existing customers is more cost-effective than attracting new ones, and that 20% of customers create 80% of the profit [1,2]. Reichheld and Teal [3] also point out that a 5% increase in customer retention leads to a 25–95% increase in company profit. Therefore, companies are focusing attention on building relationships with their customers in order to improve satisfaction and retention. This implies that companies must learn much about their customers' needs and demands, their tastes and buying propensities, etc., which is the focus of *Customer Relationship Management (CRM)* [4]. For CRM purposes, data mining techniques can potentially be used to extract hidden information from customer databases or data warehouses.

Data mining techniques can potentially help companies efficiently conduct *Customer Portfolio Analysis (CPA)*, which is the process of analyzing the existing and potential value of customers, thereby allocating limited resources to various customer groups according to the corporate strategy [4,5]. In this study, we propose a hybrid approach that combines the Self-Organizing Map (SOM)-Ward clustering [6,7] and decision trees for conducting CPA, aiming to create a more informative model for

P. Perner (Ed.): ICDM 2010, LNAI 6171, pp. 292–307, 2010.

focused marketing efforts, compared to using either method alone. First, we use SOM-Ward clustering to conduct customer segmentation, so that the customer base is divided into distinct groups of customers with similar characteristics and behavior. This will allow us to identify the characteristics that separate high-spending customers from low-spending customers. Then, a decision tree technique will be used to further explore the relationship between customers' spending amounts and their demographic and behavioral characteristics. Finally, the trained decision tree model will be used to identify the segments with development potential, as well as customers in the group displaying mid-range spending that have similar characteristics as the high-spending customers. This group thus represents potential high-value customers if correctly activated. By extension, this type of analysis would allow companies to adjust their marketing efforts in order to better fit their customers' needs and demands, not only helping to enhance their relationship with important customers but also cutting down on advertising costs and improving the profitability of the entire customer base.

Although the SOM-based approach and Decision Trees have been used for market segmentation, classification, and data exploration problems individually, these two approaches have not to our knowledge previously been combined to perform CPA. A dataset of more than one million customers was used to create the models.

The remainder of this paper is organized as follows. Section two introduces the methodology (SOM-Ward and Decision Trees) and the data used in this study. Sections three and four document the training and analysis of the SOM-Ward and Decision Tree models respectively. In Section five, the trained Decision Tree model is used to analyze unclassified customers in order to identify their market potential. Section six presents our conclusions.

2 Methodology

2.1 The SOM and SOM-Ward Clustering

The SOM is a well-known and widely used unsupervised neural network that is able to explore relationships in multidimensional input data and project them onto a two-dimensional map, where similar inputs are self-organized and located together [8]. Additionally, as an unsupervised artificial neural network (ANN), the SOM is a data-driven clustering method. In other words, it works with very little *a priori* information or assumptions concerning the input data. Moreover, the SOM is able to compress the input data while preserving the topological relationships of the underlying data structure [8]. For these reasons, the SOM is considered an important tool for conducting segmentation tasks. The algorithm is well-known and will, therefore, not be further presented in this paper. Readers are referred to Kohonen [8] for details concerning the algorithm.

The SOM has been widely applied as an analytical tool in different business-related areas [9-11], including market segmentation [12-14]. In the above mentioned studies, the SOM is used alone, compared with, or used in conjunction with other clustering techniques for conducting market segmentation tasks. Vesanto and Alhoniemi [15] proposed a two-level approach, e.g., SOM-Ward clustering, for conducting clustering tasks. First, the dataset is projected onto a two-dimensional display using the SOM. Then, the resulting SOM is divided into groups. Lee et al. [14], adopting the two-level SOM (using SOM and K-means clustering), conducted a market segmentation of the Asian online game market and found that the two-level SOM is more accurate in classification than K-means clustering or the SOM alone. Samarasinghe [16], comparing two clustering methods (SOM-Ward and K-means clustering) drew the conclusion that SOM-Ward clustering resulted in better representations of the top and middle clusters than K-means alone.

As was previously mentioned, SOM-Ward clustering is a two-level clustering approach that combines the SOM and Ward's clustering algorithm. Ward's clustering is an agglomerative (bottom-up) hierarchical clustering method, which starts with a clustering in which each map node is treated as a separate cluster. The two clusters with the minimum distance are merged in each step until there is only one cluster left on the map [7]. SOM-Ward's clustering is a modification of Ward's clustering which limits cluster agglomeration to topologically neighboring nodes.

2.2 The Decision Tree

A Decision Tree is a supervised data mining technique that can be used to partition a large collection of data into smaller sets by recursively applying two-way and/or multi-way splits [17]. Compared to other data mining techniques, the Decision Tree has many advantages. First, as opposed to "black box" data mining techniques, the Decision Tree produces straightforward rules for classification and prediction purposes. Second, it is relatively insensitive to outliers [17] and skewed data distributions [18,19], and some decision tree algorithms are even capable of dealing with both numeric and nominal variables [19]. Thirdly, the decision tree is also a significant data exploration tool that can potentially be used to unveil the relationship between candidate independent and dependent variables. It can also be used to identify the significant variables for predicting the dependent variable [17]. These advantages make the decision tree applicable to a wide variety of business and marketing problems. For example, Fan et al. [20] adopted the decision tree to identify significant determinants of house prices and to predict these. Abrahams et al. [21] used decision trees to create a marketing strategy for a pet insurance company. Sheu et al. [22] adopted it to explore the potential relationship between important influential factors and customer loyalty. The findings of these studies inspire us to adopt the decision tree to explore the relationship between customers' purchase amounts and customers' demographic and behavioral characteristics, with special attention to the characteristics of high- and low-spending customers.

In this study, the CART algorithm [23] is employed to construct a binary classification tree. It is a tree-based classification method that uses recursive two-way partitioning to split the training records into segments where the records tend to fall into a specific class of the target variable. In other words, the records in the terminal nodes tend to be homogeneous with regard to the target variable. The CART algorithm employs an exhaustive search for the best independent variable for classification purposes at each split. First, the algorithm checks all possible independent variables and their potential values at each split. Then, the algorithm chooses an independent variable that maximizes within-node purity to split the node. We use the Gini index of diversity [23] to measure the improvement in purity at each split. For example, if all the records in a node fall into a specific class of dependent variable, the node is considered pure; however, if the records of each class are proportionate, the node is considered impure. This process is repeated until some user-specified criteria are met, e.g., the maximum tree depth, the minimum number of records in a node or the minimum change in within-node purity improvement [17]. When the tree growing process ends, there is a unique path from the root to each terminal node. These paths can be considered a set of if-then rules that can be used to classify the records.

2.3 The Data

The case company is a national retailer belonging to a large, multiservice Finnish corporation. The corporation uses a common loyalty card system which offers cardholders various discounts and rewards for purchases. The cardholder is required to provide basic personal information in order to register for the loyalty card, and their transactional information is collected and recorded in the system. The dataset, containing 1,480,662 customers, was obtained through the loyalty card system, and contained sales information from several department stores in Finland, for the period 2006-07. The dataset consists of ten variables that fall into two categories: *demographic* and *behavioral* variables.

The demographic variables consist of the following:

- Age
- Gender: 0 for male, 1 for female.
- Mosaic group: The Mosaic group is a socio-economic ranking system that builds upon 250-by-250 meter map grid cells covering all the populated areas of Finland. Each map grid contains an average of seven households. The ranking system combines census data with marketing research data to classify the whole population of Finland into nine groups: A, B, C, D, E, F, G, H, and I. Each map grid can be assigned to one of the nine groups. The households living in the same map grid can then be described in terms of socio-demographics, such as education, lifestyle, culture, and behavior.
- Mosaic class: Based upon the Mosaic group, the Mosaic class divides the nine Mosaic groups further into 33 subclasses.
- Estimated probability of children: This variable divides households into ten groups of equal size, based upon the probability of them having children living in the same household. A higher value in this variable indicates that the family is more likely to have children living at home. Possible values are from one to ten.
- Estimated income level: Predicts customers' income level. The higher the value, the wealthier the household is considered. Possible values are one, two and three.

The behavioral variables consist of the following:

- Loyalty point level: Based on the average spending amount per customer in the corporate chain (the case company is one service provider in the corporate chain), this variable divides customers into five classes: zero, one, two, three, and four. A higher value in loyalty point level is an indication of a customer's larger spending amount in the entire corporate chain.
- Customer tenure: Number of years since the customer's registration.

- Service level: Measures how many service providers in the corporate chain the customer has used in the last 12 months.
- The spending amount: Records the total spending amount of each customer during the period 2006-2007.

3 The SOM-Ward Model

3.1 Training the SOM-Ward Model

The training of the SOM-Ward Model was carried out using Viscovery SOMine 5.0 (http://www.viscovery.net/), which is based upon the batch SOM algorithm [8]. SOMine is a user friendly SOM implementation with a number of analytical tools embedded, including automated two-level clustering using three clustering algorithms, i.e., SOM-Ward, Ward and SOM Single Linkage [24].

To begin with, we preprocessed data to ensure the quality and validity of the clustering result. The Mosaic group is a categorical variable that is not orderly ranked. Since the SOM requires numeric input, we converted the Mosaic group into nine binary variables (either 0 or 1). In addition, the Mosiac class variable was excluded from the SOM-Ward Model as each of the 33 sub-classes of the Mosaic class would have required a dummy variable, which would made visualization extremely difficult.

Assigning a higher priority factor (default is 1) to some variables can be used to give them additional weight and importance in the training process [8], while reducing the priority factor can be used to achieve the opposite. If the priority of a variable is set to zero, the variable has no influence on the training process. We assigned the priority factor of spending amount to 1.4, aiming to give it more influence in the training process. In the pilot tests, it was discovered that the Mosaic group binary variables dominated the segmentation result, leading to clusters exclusively defined by a particular Mosaic group. Therefore, the priority factor of the Mosaic group was set to 0.1. Thus, the Mosaic group data had little influence on the segmentation result, but their distributions in the segments can be investigated when the map has been trained. The priority factors for estimated probability of children and estimated income level were set to 0.5, considering that both variables are based upon estimates and might thus involve some uncertainties. In addition, in order to achieve a more interpretable segmentation result, we slightly adjusted the priority factors of the other variables as well, again based upon the results of the pilot tests. The priority factors of age, gender, service level and customer tenure were adjusted to 1.1, 0.9, 0.9 and 0.8, respectively. Finally, the data were scaled in order to make sure that no variable received undue scale-related bias and to ease the overall training process. Total spending amount and customer tenure were scaled according to range while the rest of the variables were normalized by variance. This step was done automatically by the software. No transformation (e.g., sigmoid or logistic) was applied.

SOMine requires very few parameters for training, mainly because of the batch training process used. The user is only required to provide the *map size, map ratio* and *tension* [24]. The default map size is 1,000 nodes. By comparing a set of maps trained in the pilot tests, we chose a map containing 600 nodes to visualize the result. A smaller map is better suited for clustering [25]. The tension parameter is used to specify the neighborhood interaction. A lower tension will result in a map that adapts more to the data space, resulting in a more detailed map. On the other hand, a higher tension tends to average the data distribution on the map. Based upon the pilot test results, a map generated with the default setting (0.5) was chosen.

3.2 Analysis of the SOM-Ward Model

The characteristics of each segment were identified by examining the variables' component planes (displayed in **Fig. 1**), which show the distributions of each variable across the map. The colors of the nodes in the component planes visualize the value distribution of each variable. Cool colors (blue) indicate low values, while warm ones (red, yellow) indicate high values. Values are indicated by the color scales under the component. For instance, high spending customers were mainly found in Segments One and Two, while long-standing customers are mainly found in Segment Five. In addition to the component planes, two bar charts also illustrate the characteristics of each segment (**Fig. 2** and **Fig. 3**). The height of a bar measures the extent to which the mean value of a variable in a segment deviates from that of the entire data set. The unit of the x-axis is the standard deviation of the entire data set. In this way, both the component planes and the bar charts can visually represent the important characteristics of each segment. A description of each segment follows.

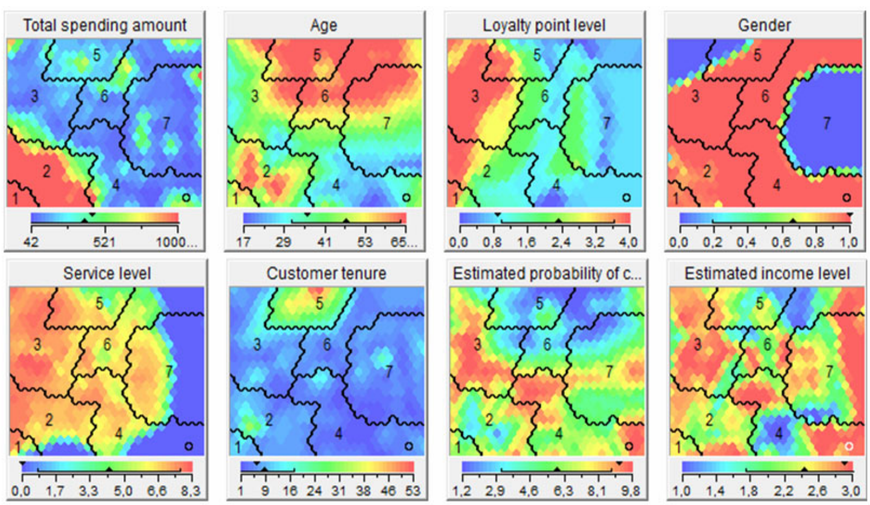

Fig. 1. The component planes of the map

Segment One: Exclusive customers

There are 25,425 customers in Segment One, accounting for 1.7% of the whole customer base. According to **Fig. 1** and **Fig. 2**, the average total purchase amount in this segment is the highest among the seven segments. These customers are mainly female, having a relatively high loyalty point level. **Fig. 3** shows that the customers belonging to this segment are most likely to belong to Mosaic groups A, C, D, and E.

Segment Two: High spending customers
In this segment, there are 177,293 customers, accounting for 12.0% of the customer base. **Fig. 1** and **Fig. 2** show that that most of them, mainly female, are high spending customers. Some of them are around 60 years old, and some have a high loyalty point level. A large percentage of them display a high service level, indicating that they also use many other service providers in the corporate chain. **Fig. 3** reveals that customers belonging to this segment are likely to be from Mosaic groups A, C, D, and E.

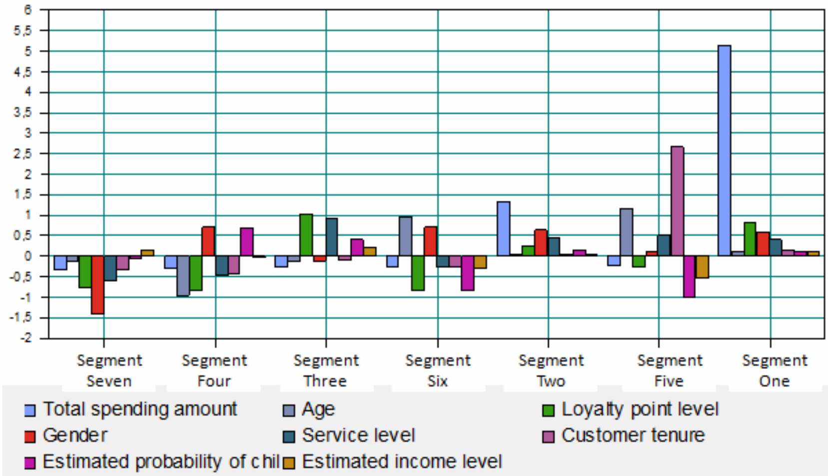

Fig. 2. Bar chart illustrating the characteristics of each segment

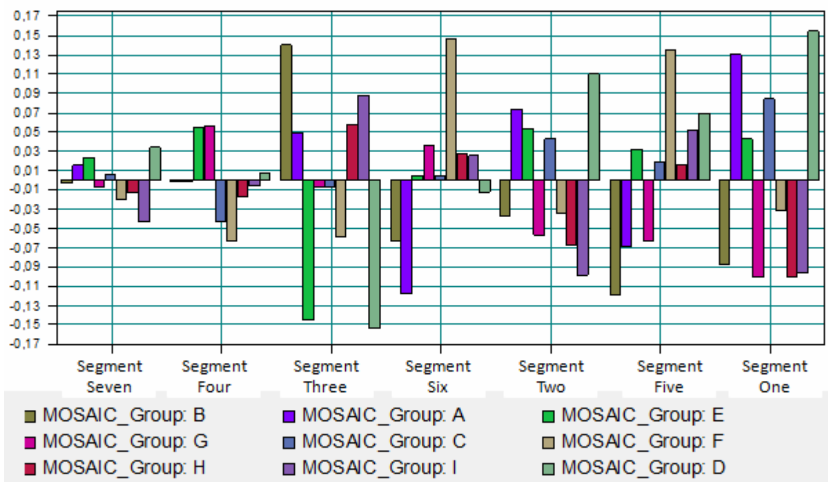

Fig. 3. Bar chart illustrating the proportion of each mosaic group in each segment

Segment Three: Customers with high loyalty point level
There are 283,265 customers in this segment, accounting for 19% of the customer base. **Fig. 1** shows that they have a very high loyalty point level. They use many other service providers in the corporate chain, but their spending amount in the case company is not large. The probability of these customers having children at home is high. **Fig. 3** shows that customers in this segment are likely to be from Mosaic groups B, H, and I.

Segment Four: Relatively young female customers
There are 303,588 customers in this segment, accounting for 20.5% of the customer base. **Fig. 1** and **Fig. 2** show that these customers are mainly females who are much younger than the average of the customer base, and some have a large probability of having children in the same household. However, their spending amount, loyalty point level, service level and customer tenure are below average.

Segment Five: Long-standing customers of the corporate chain
There are 116,537 customers in this segment, accounting for 7.9% of the customer base. **Fig. 1** reveals that these customers have been customers of the corporate chain for a long time. They widely use other service providers in this corporate chain, but their spending amount in the case company is not high. Compared to other segments, these customers are older and their estimated probability of having children living at home is small. However, some of them have a very high loyalty point level. **Fig. 3** indicates that it is likely that these customers belong to Mosaic groups D, F, and I.

Segment Six: Relatively old female customers
Segment Six has 227,194 customers, accounting for 15.3% of the customer base. **Fig. 1** shows that these customers are comparatively senior to those in other segments. Although they are senior in age, they are not long-standing customers of the corporate chain and their spending amount is not large. They are mainly female, and have a low probability of having children living in the same household. **Fig. 3** shows that these customers often belong to Mosaic group F.

Segment Seven: male customers
There are 347,410 customers in Segment Seven, accounting for 23.5% of the customer base. **Fig. 1** shows that these customers are mainly male. They have a low spending amount, and some of them have a high estimated income level.

4 The Decision Tree Model

4.1 Training the Decision Tree Model

The training of the Decision Tree Model is carried out with The PASW Decision Trees module (http://www.spss.com/statistics/).

A binary target variable, i.e., the variable of high- and low-spending customers, is created for the Decision Tree Model. From the analysis of the SOM-Ward Model, we found that the customers in Segments One and Two, i.e., those who spend much in the company, account for 13.7% of the customer base. Therefore, we arranged the customers in sequence, from the lowest to highest according to their total purchase amounts. The top 13.7% of the customers are labeled high-spending customers, and the bottom 13.7% are labeled low-spending customers. Customers in the middle, whose spending amount are not clearly high or low, were excluded from the training set. These customers will be further analyzed in Section 5.

Compared to that of the SOM-Ward Model, the data preprocessing of the Decision Tree Model is much simpler. First, the CART algorithm is able to construct a decision tree model by training continuous and/or categorical predictor variables [26]. Next, the CART algorithm uses surrogates to handle missing values on independent variables [17]. Thus, observations containing independent variables that include missing values are not excluded from the training process. Instead, other independent variables that are highly correlated with the independent variable containing missing values are used for classification. Lastly, the decision tree is relatively insensitive to outliers and skewed data distributions [17]. The above factors reduce the data preprocessing efforts required for training the decision tree. In the Decision Tree Model we used such variables as Mosaic class, which are not easily used in the SOM-Ward Model.

The maximum number of levels in the tree was limited to five and the minimum number of records in a node was set to 1,000, in order to prevent the Decision Tree from becoming very complex. A complex model, from which too many rules are extracted, would not only make the rules hard to generalize into actionable marketing strategies, but would also increase the risk of overfitting. The final model was chosen based upon ten-folds cross validation.

4.2 Analysis of the Decision Tree Model

Appendix 1 shows the created Decision Tree Model. The tree grows from left to right, with the root node (0) located on the left and terminal nodes on the right, with each non-terminal node having two child nodes. The set of two child nodes represents the answers to the decision rules for splitting the records, and these rules are printed on the lines connecting each node to its child nodes. The variable used to split the root node (node 0) is gender. The algorithm compares the classification results produced by all independent variables, and gender is found to lead to the largest improvement of within-node purity. Female customers are in the upper branch of node 0 and male customers are in the lower branch. At node 2, we find that restricting the records to female customers leads the percentage of high-spending customers to increase from 50% to 60.7%. On the other hand, at node 1, restricting the records to male customers leads the percentage of low spending customers to increase from 50% to 77.2%. After the initial split, the decision tree uses loyalty point level to further divide node 1 and node 2 into nodes 3 and 4, and nodes 5 and 6, respectively. The tree shows that a customer with a higher loyalty point level is more likely to be a high-spending customer. For example, when comparing nodes 5 and 6 (the children nodes of node 2), we find that the proportion of high value customers in node 6, i.e., female customers with a loyalty point level above 1, increases compared to that of node 1, while the proportion of high value customers in node 5 decreases compared to that of node 2. The same pattern also appears at the splits of nodes 3, 12, 28 and 30. In addition, the splits at nodes 4, 6, and 26 clearly show that customers belonging to Mosaic groups A, C, D, and E are more likely to spend more. Moreover, the splits at nodes 9 and 13 indicate that customers belonging to Mosaic classes 11, 12, and 17 are more likely to spend more in the company.

As shown in Appendix 1, the process of recursive partitioning does not stop until the tree grows to the terminal nodes. The paths to these terminal nodes describe the rules in the model. The primary focus of this analysis is to identify the characteristics of high- and low-spending customers. Therefore, among all the terminal nodes, we will select four terminal nodes with the highest percentage of high-spending customers, and two terminal nodes with the highest percentage of low-spending customers. The paths from the root node to the six selected terminal nodes are interpreted as characteristics that identify high-spending and low-spending customers.

High-spending customers – Group One: Node 27

This node has 21,526 customers, out of which 92.3% are high-spending customers. The characteristics of the customers in this node are, in order of importance:

1. They are female.
2. Their loyalty point level is larger than 1.
3. They belong to Mosaic classes 11, 12, and 17.

High-spending customers – Group Three: Node 50

This node has 22,172 customers, 83.4% of which are high-spending customers. Their characteristics are:

1. They are female customers.
2. Their loyalty point level is larger than 3. (We combined the rules applied at nodes 2 and 28, because the loyalty point level is used twice.)
3. They belong to Mosaic classes 2, 3, 9, 10, 13, 14, 15, and 16.

High-spending customers – Group Four: Node 19

This node has 1,205 customers, out of which 82.8% are high-spending customers. Characteristics of the customers in this node are:

1. They are male.
2. Their loyalty point level is larger than 3.
3. They belong to Mosaic classes 11, 12, and 17.

Low-spending customers – Group Five: Node 41

This node has 7,144 customers, out of which 98.3% are low-spending customers. Characteristics of the customers in this node are:

1. They are female.
2. Their loyalty point level is less than or equal to 3.
3. They are less than 18.5 years old.

It is also noted that node 23 (the parent node of node 41) also has a very large percentage of low-spending customers. It restricts the age to less than or equal to 20.5.

Low-spending customers – Group Six: Node 31

This node has 7,624 customers, out of which 96.9% are low-spending customers. The characteristics of the customers in this node are:

1. They are male.
2. Their loyalty point level is less than or equal to 1.
3. Their customer tenure with the corporate chain is less than 4.5 years.
4. They are less than 26.5 years old.

It is also noted that node 15 (the parent node of node 31) also has a very large percentage of low-spending customers. However, it has no restriction on age.

5 Customer Portfolio Analysis

Based upon the results of the SOM-Ward and the Decision Tree analyses, we divide the customer base into three groups: high-spending customers, low-spending customers, and customers with development potential. The SOM-Ward model shows that there are seven segments. They are:

1. Exclusive customers
2. High spending customers
3. Customers with high loyalty point level
4. Relatively young, female customers
5. Long-standing customers of the corporate chain
6. Relatively old, female customers
7. Male customers

The map shows that Segments One and Two are high-spending customers. Our purpose is to now identify which of the Segments Three, Four, Five, Six, and Seven have development potential in terms of spending amounts. We will do this by identifying segments that consist of customers displaying similar characteristics as those in Segments One and Two.

We use the decision tree to identify the characteristics that can tell high-spending customers from low-spending ones. The confusion matrix of correct and incorrect classifications in **Table 1** illustrates the accuracy of the decision tree model.

Table 1. The prediction performance of the decision tree

Observed	Predicted		
	Low-spending customers	High-spending customers	Percent Correct
Low-spending customers	125,743	75,685	62,4%
High-spending customers	34,745	166,683	82,8%
Overall Percentage	39.8%	60.2%	72.6%

Table 1 shows that the overall accuracy of the model is 72.6%. 82.8% of the high-spending customers are correctly classified, while only 62.4% of the low spending customers are correctly classified. As our main objective was to build a model that can identify potential high-spending customers, the cost of incorrectly classifying a high-spending customer as a low-spending one is higher than the cost of incorrectly classifying a low-spending customer as a high-spending one. Therefore, this accuracy rate is acceptable.

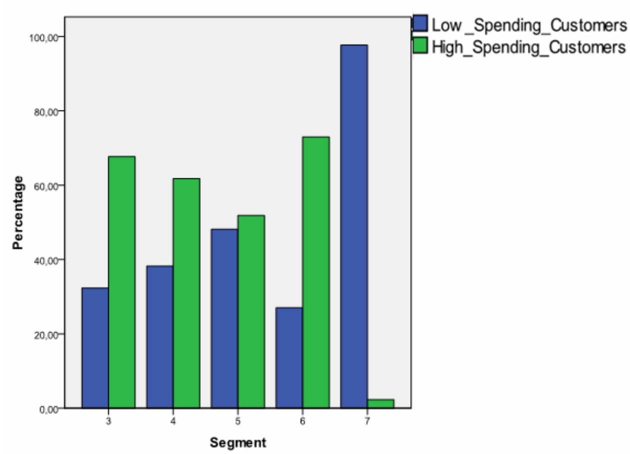

Fig. 4. Percentages of customers in each segment identified as having development potential

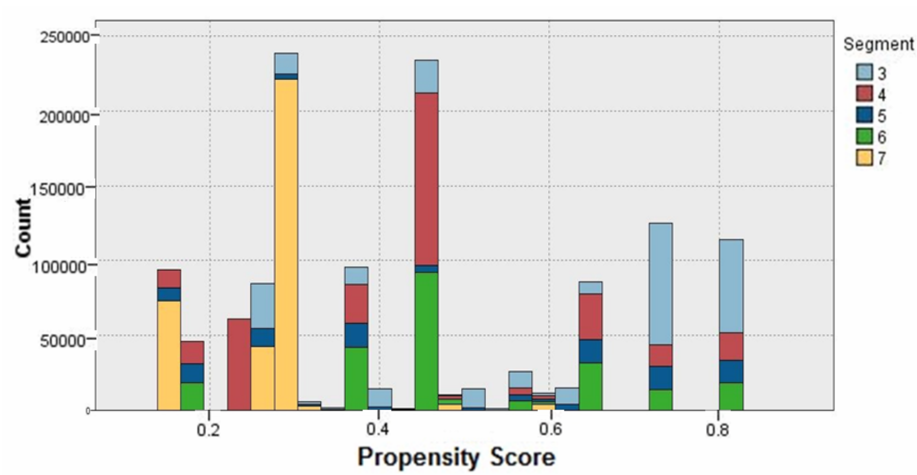

Fig. 5. The propensity scores for the different segments

We then ran all the cases in Segments Three, Four, Five, Six, and Seven through the decision tree model that was created. The ones that we can identify as possessing the same characteristics as the high–spending customers will be our potential group. As the clustered bar chart in **Fig. 4** indicates, the customers in Segments Six and Three have more development potential than the customers in the other segments. Each terminal node is a mixture of high-spending customers and low-spending customers. The predicted value is the category with the highest proportion of cases in the terminal node for each case. Therefore, it is possible to use the model to predict these unclassified cases using propensity scores. The propensity score ranks the likelihood of the prediction from 1 (likely to be a high-spending customer) to 0 (not likely to be a high-spending customer). For example, if an unclassified case has the same characteristics as node 27 in the decision tree model, it will be assigned a propensity score of 0.923, as 92.3% of the cases in node 27 are high-spending customers. The Histogram of the distribution of propensity scores of Segments Three, Four, Five, Six, and Seven is shown in **Fig. 5**.

This figure reveals that the customers in Segment Three are most likely to be potential high value customers, while the customers in Segment Seven are least likely to be potential high value customers. After running all of the data outside of the 27.4% (high and low spending) that were used to train the decision tree, we obtain a list of those customers with their propensity scores or/and predictions appended.

6 Conclusions

A hybrid approach combining unsupervised and supervised data mining techniques has been proposed to conduct customer portfolio analysis. SOM-Ward clustering was first used to conduct customer segmentation. Then, the decision tree was employed to gain insight into whether there are significant determinants for distinguishing between high- and low-spending customers. The results of the two models are then compared and combined to perform customer portfolio analysis, i.e., to identify the high- and low-spending customers, as well as customers with development potential. Each model possesses advantages and disadvantages of its own.

As an unsupervised data mining technique, SOM-Ward clustering is a good tool for exploratory analysis, as is the case when no *a priori* classes have been identified. The SOM is a very visual tool and possesses strong capabilities for dealing with non-linear relationships, missing data, and skewed distributions. However, while the clusters produced using unsupervised methods may be good for gaining an understanding of the customer base, they are not necessarily actionable in terms of marketing strategy as they are not based upon any identified target or aim. In addition, using detailed nominal data (e.g., 33 Mosaic classes) is a problem when using the SOM, as binary variables must be constructed for each potential class. This easily clutters the map and heavily influences training results.

Decision trees, on the other hand, are tailored to a specific purpose by using a supervised learning approach. The decision tree is also a very robust method, easily capable of dealing with difficult data, and requires less data preprocessing and setting of parameters than the SOM. However, the starting point of supervised learning inevitably requires more *a priori* knowledge than unsupervised learning, making the knowledge gained using the SOM potentially very important.

The results of the analysis demonstrate that the combined method of the SOM-Ward clustering and the Decision Tree can potentially be effective in conducting market segmentation. The information provided by the combined model is more detailed and accurate than that provided by either model used alone, thus more actionable information about the customer base for marketing purposes could be retrieved.

Acknowledgements. The authors gratefully acknowledge the financial support of the National Agency of Technology (Titan, grant no. 40063/08) and the Academy of Finland (grant no. 127656). The case organization's cooperation is also gratefully acknowledged.

References

1. Kim, S., Jung, T., Suh, E., Hwang, H.: Customer Segmentation and Strategy Development Based on Customer Lifetime Value: A Case Study. Expert Systems with Applications 31, 101–107 (2006)
2. Park, H., Baik, D.: A Study for Control of Client Value using Cluster Analysis. Journal of Network and Computer Applications 29, 262–276 (2006)
3. Reichheld, F.F., Teal, T.: The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Harvard Business Press, Boston (2001)
4. Buttle, F.: Customer Relationship Management Concepts and Tools. Butterworth-Heinemann, Oxford (2004)
5. Terho, H., Halinen, A.: Customer Portfolio Analysis Practices in Different Exchange Contexts. Journal of Business Research 60, 720–730 (2007)
6. Kohonen, T.: Self-Organization and Associative Memory. Springer, New York (1988)
7. Ward, J.H.: Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58, 236–244 (1963)
8. Kohonen, T.: Self-Organizing Maps. Springer, Berlin (2001)
9. Kaski, S., Kangas, J., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers 1981-1997. Neural Computing Surveys 1, 102–350 (1998)
10. Oja, M., Kaski, S., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. Neural Computing Surveys 3, 1–156 (2003)
11. Deboeck, G.J., Kohonen, T.: Visual Explorations in Finance with Self-Organizing Maps. Springer, Berlin (1998)
12. Holmbom, A.H., Eklund, T., Back, B.: Customer Portfolio Analysis using the SOM. In: Proceedings of 19th Australasian Conference on Information Systems (ACIS 2008), pp. 412–422 (2008)
13. D'Urso, P., Giovanni, L.D.: Temporal Self-Organizing Maps for Telecommunications Market Segmentation. Neurocomputing 71, 2880–2892 (2008)
14. Lee, S.C., Gu, J.C., Suh, Y.H.: A Comparative Analysis of Clustering Methodology and Application for Market Segmentation: K-Means, SOM and a Two-Level SOM. In: Zhong, N., Ras, Z.W., Tsumoto, S., Suzuki, E. (eds.) ISMIS 2006. LNCS (LNAI), vol. 4203, pp. 435–444. Springer, Heidelberg (2006)
15. Vesanto, J., Alhoniemi, E.: Clustering of the Self-Organizing Map. IEEE Trans. Neural Networks 11, 586–600 (2000)
16. Samarasinghe, S.: Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition. Auerbach Publications, Boca Raton (2007)
17. Berry, M.J.A., Linoff, G.S.: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley Publishing Inc., Indianapolis (2004)
18. Murthy, S.K.: Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. Data Mining and Knowledge Discovery 2, 345–389 (1998)
19. Rokach, L., Maimon, O.: Data Mining with Decision Trees: Theory and Applications. World Scientific Publishing, Singapore (2008)
20. Fan, G.Z., Ong, S.E., Koh, H.C.: Determinants of House Price: A Decision Tree Approach. Urban Studies 43, 2301–2315 (2006)
21. Abrahams, A.S., Becker, A.B., Sabido, D., D'Souza, R., Makriyiannis, G., Krasnodebski, M.: Inducing a Marketing Strategy for a New Pet Insurance Company using Decision Trees. Expert Syst. Appl. 36, 1914–1921 (2009)

22. Sheu, J.J., Su, Y.H., Chu, K.T.: Segmenting Online Game Customers - the Perspective of Experiential Marketing. Expert Syst. Appl. 36, 8487–8495 (2009)
23. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Wadsworth, Pacific Grove (1984)
24. Deboeck, G.J.: Software Tools for Self-Organizing Maps. In: Deboeck, G.J., Kohonen, T. (eds.) Visual Explorations in Finance using Self-Organizing Maps, pp. 179–194. Springer, Berlin (1998)
25. Desmet, P.: Buying Behavior Study with Basket Analysis: Pre-Clustering with a Kohonen Map. European Journal of Economic and Social Systems 15, 17–30 (2001)
26. Hill, T., Lewicki, P.: Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining. StatSoft, Inc., Tulsa (2006)

Appendix 1. The Decision Tree Model