Bartłomiej Jefmański
Wroclaw University of Economics

Mariusz Łapczyński
Cracow University of Economics

MASKING PROBLEM IN IDENTIFICATION OF SERVICE QUALITY DETERMINANTS WITH AN APPLICATION OF THE CART MODEL – AN EXAMPLE OF THE PUBLIC SERVICES QUALITY RESEARCH IN POLAND

Abstract: An important part of customer satisfaction surveys of the acquired service is to identify the key factors affecting its quality. Since customer satisfaction is a complex category thus its measurement and analysis require the use of multivariate statistical methods. One of them is the CART method (Classification and Regression Trees). Its use in the identification of the key determinants of the quality of services may, however, be associated with the emergence of the so-called variable masking problem, that was characterized in the article. Possible ways of solving it were exemplified in the customer satisfaction survey results of one of the Polish Municipal Offices.

Keywords: CART, masking of variables, customer satisfaction.

1. Introduction

On the basis of the quality service research, one of the most frequently covered issues is to identify the key determinants. The knowledge of them makes it possible to rationalize the management of services, and to make strategic decisions by the organization. The identification of these determinants most often requires the use of multidimensional data analysis methods. A relatively new analytical approach in this regard is the CART method (Classification and Regression Trees), which, as it belongs to the group of non-parametric methods, does not require to fulfill a great deal of troubling assumptions and provides easy to interpret and visualize the results in the form of so-called decision trees.

The researchers who use this approach can, however, meet with the so-called 'variable masking' case, consisting in the fact that some of the independent variables, despite the high rate of improvement measure, do not participate in the split of the

tree. This is shown by their high position in the ranking of the validity of predictors, but also by their lack in the rules describing the model (*if ... then ...*).

The aim of the paper is to characterize the possible solutions in this area involving swapping primary divisions through the use of the best competitive variables and the best supplementary ones. The issue will be discussed on the example of the results of the periodically conducted researches on the quality of services by the Municipal Office of Dzierżoniów, which, as the first office in Poland, has implemented a quality management system and is now a member of the European Foundation for Quality Management (EFQM).

2. Masking of variables' importance in decision trees

The purpose of using decision trees in customer satisfaction research is often to identify those attributes that most influence the overall level of customer satisfaction. The examples of such applications are the following studies: Thomas and Galambos [2004], Nicolini and Salini [2006], Ishwaran [2007], Tutore et al. [2007], Guo and Niu [2009], Atay and Yildirim [2010], Mei-Ping and Wei-Ya [2010], Galimberti and Soffritti [2012], de Oña et al. [2012]. One of the most frequently used recursive partitioning algorithms in this area is CART (Classification and Regression Trees), which was developed by Breiman et al. [1998].

CART is used to build a classification tree if the dependent variable is nominal, and a regression tree if the dependent variable is continuous. Characteristically, at every step of the partitioning procedure it considers two kinds of alternative splits. Apart from the best independent variable, which is used in the partition of the tree, the procedure provides a list of surrogate splits and a list of competitors splits. Surrogate variable gives a split, which reduces the impurity of node almost as accurately as the best predictor does. In general, they are used for replacing missing values, building variable importance ranking, and discovering a masking problem.

It is worth noting, that some variables may not appear in the final tree structure but they can have a high score in the predictors' ranking. This means, that there is a masking problem, i.e. the association of this variable with the dependent one is masked by other variables. This can happen when two independent variables have almost the same values of improvement measure, but only one of them splits the tree. Therefore, the role of the second one is masked by the variable, which can be only slightly better.

The second type of alternative split is the competitor one [Steinberg, Colla 1997, p. 40]. The difference between competitors and surrogates is that competitors reduce node heterogeneity to the same extent as primary split, while surrogates imitate the primary split. Surrogate copies the way of partitioning provided by the best predictor. It imitates the size and content of child nodes considering particular cases from the parent node. Competitors are also based on improvement measure and they can strengthen interpretation of the model. It is suggested to use competitor variables from the top of the tree as shown at Figure 1.
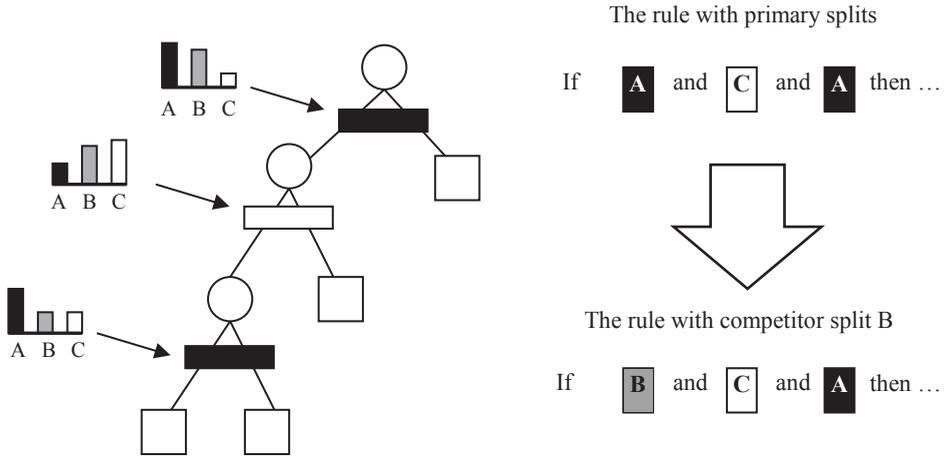
Figure 1. The way of creating an alternative set of rules

Source: own elaboration.

3. An example of masking of variables in the quality services survey

3.1. Description of the survey

Customer satisfaction surveys are now a common practice employed by public authorities in Poland. They are conducted mainly by the "certified offices" and by those institutions that apply the methods of self-estimation. The implementation of satisfaction research is usually based on surveys (available directly in the offices or on internet sites) rather than using telephone interviews [Bugdol 2008, p. 35].

The customer satisfaction survey of the Municipal Office in Dzierżoniów is implemented using the PAPI method (Paper and Pencil Interview) and has been being conducted periodically since 2008. The questionnaire is divided into six sections. The majority of them is assessed on the ordinal scale with the following variants of answers: "very dissatisfied", "dissatisfied", "neither satisfied nor dissatisfied", "pleased", "very pleased".

The analysis was based on data from a customer satisfaction survey in which the interviews were carried out with 488 respondents. Face to face interviews were conducted in the period from February to March 2012 on the premises of the Customer Services Office in Dzierżoniów Town Hall and in the direct customer services posts in the Citizens Affairs Department as well as in the Civil Registry Office. Nineteen variables of the first five sections of the questionnaire concerning satisfaction with selected aspects of the services provided by the Office were adopted as a potential set of independent variables (determinant of customer satisfaction).

Table 1. Names of variables

Symbol	Variable
	Part A – the quality of provided services
x_1	The result of settling the matter by the Office
x_2	Reliability in the realization of the matter
x_3	Correctness of received documents
x_4	Timeliness of settling matters
x_5	Speed of settling matters
x_6	Readability and ease of filling in forms
x_7	Working hours
	Part B – the quality of service
x_8	competence of the employees
x_9	courtesy of the employees
x_{10}	provided information
x_{11}	impartial attempt at settling the matter
x_{12}	assistance provided by the staff
	Part C – the quality of information on services
x_{13}	The information provided on the notice boards
x_{14}	The information provided on the websites of the Office and BIP
x_{15}	Readability of the Services Cards
x_{16}	Availability of the information on the progress of settling matters
x_{17}	Telephone contacts with the Office
	Part D – the terms of services
x_{18}	Marking of the Office
x_{19}	Terms of the Customer Service
	Part E – the overall level of satisfaction
x_{20}	Overall satisfaction with services

Source: own elaboration.

3.2. Findings

Figure 2 shows the final structure of the regression tree that can be replaced with the set of *if ... then ...* rules. A group of people who highly evaluate the quality of service provided by the office (node 5) contains respondents with high rates of the result of settling the matter by the Office (x_1) and high rates of provided information (x_{10}). In the group of people who evaluate relatively low the overall quality of service one can find those with a rate of independent variable x_1 lower than or equal 4.5 (node 2).

Table 2 shows the importance of independent variables. It is evident that selected variables are highly ranked but are not primary splits. This means that the importance of some predictors is masked and it is worth building an alternative model using surrogate variables.
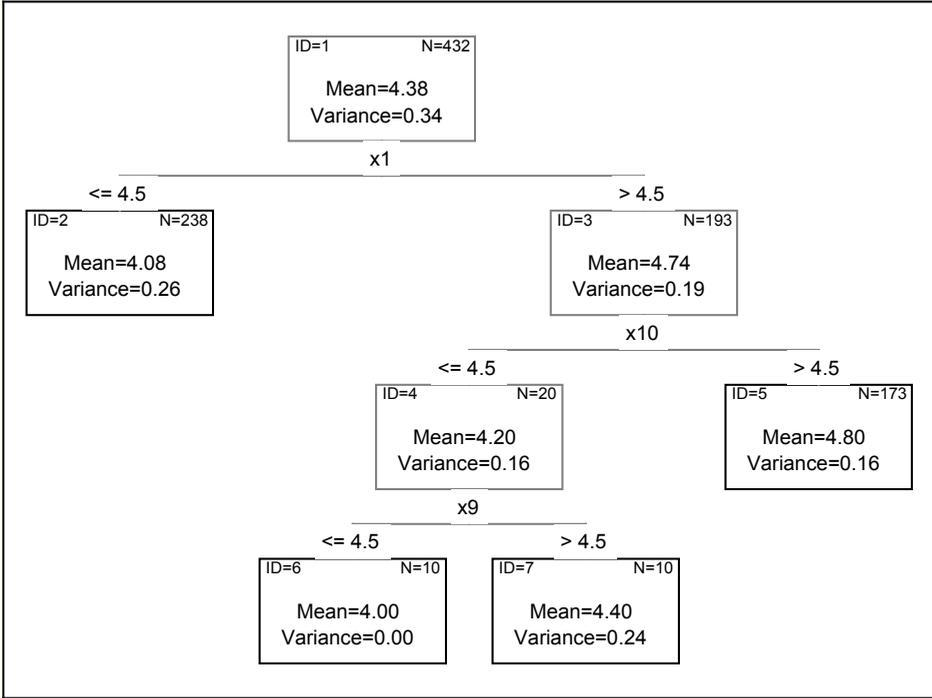
Figure 2. Final tree structure

Source: own computation.

Table 2. Ranking of predictors' importance

Predictor	Relative Importance	Predictor	Relative Importance
x_2	100	x_6	70
x_1	99	x_{19}	57
x_5	91	x_{18}	52
x_8	91	x_{16}	51
x_9	89	x_7	49
x_{10}	86	x_{12}	44
x_4	82	x_{13}	37
x_{11}	75	x_{17}	32
x_3	74	x_{14}	31

Source: own computation.

Figure 3 shows the masked tree structure that was built by using surrogate variables. It turned out that the best surrogate at the top of model (x_2) was also the best competitor which enabled to formulate the set of alternative *if ... then ...* rules.
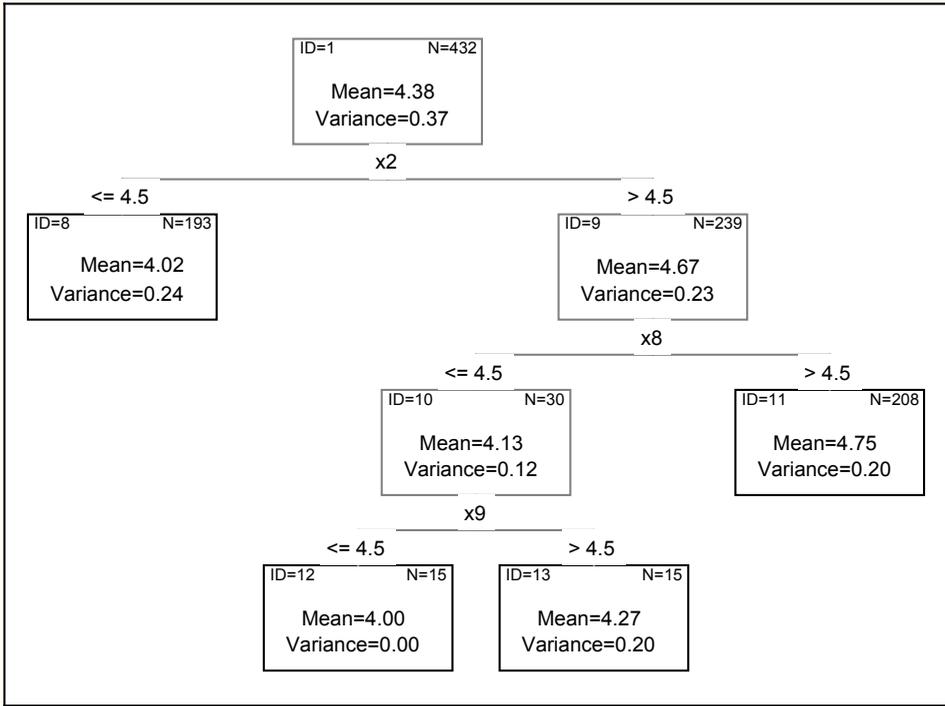
Figure 3. Masked tree structure

Source: own computation.

Thus, a group of people who highly evaluate the quality of service provided by the office (node 11) contains respondents with high rates of the reliability in the realization of the matter (x_2) and high rates of provided information (x_{10}). In the group of people who evaluate relatively low the overall quality of service one can find those with the rate of independent variable x_2 lower than or equal 4.5 (node 8).

4. Conclusions

The use of competitor variables helped to formulate the alternative set of *if* ... *then* ... rules and enrich the interpretation of the model. Initially, two main determinants of customers' satisfaction were recognized (x_1 and x_{10}). After revealing the masked structure of the regression tree the list of determinants should be extended to include variable x_2. It is worth noting that the variance explained by the model is similar for both structures of regression trees (0.63 for the "original" structure and 0.59 for masked structure), so one can treat both sets of rules as complementary ones.

The results of the analysis allowed to identify seven main determinants of the overall level of customer satisfaction with the services provided by the Municipal Office of Dzierżoniów: the reliability in the realization of the matter, the result of settling the matter by the office, the speed of settling matters, the competence of the employees, the courtesy of the employees, the provided information and the timeliness of settling matters.

Literature

Atay L., Yildirim H.M., *Determining the factors that affect the satisfaction of students having under-graduate tourism education with the department by means of the method of classification tree*, „Tourismo: An International Multidisciplinary Journal of Tourism" 2010, vol. 5, no. 1, pp. 73-87.

Breiman L., Friedman J.H., Olshen R.A., Stone C.J., *Classification and Regression Trees*, third edition, Chapman & Hall/CRC 1998.

Bugdol M., *Zarządzanie jakością w urzędach administracji publicznej: teoria i praktyka*, Difin, Warszawa 2008.

De Oña J., De Oña R., Calvo F.J., *A classification tree approach to identify key factors of transit service quality*, "Expert Systems with Applications" 2012, vol. 39, issue. 12, pp. 11164-11171.

Galimberti G., Soffritti G., *Tree-based methods and decision trees*, [in:] R.S. Kenett, S. Salini (eds.), 2012, pp. 284-307.

Guo Y., Niu D., *An analysis model of power customer satisfaction based on the decision tree*, "International Journal of Business and Management" 2009, vol. 2, issue 3, pp. 32-36.

Ishwaran H., *Variable importance in binary regression trees and forests*, "Electronic Journal of Statistics" 2007, vol. 1, s. 519-537.

Kenett R., Salini S., *Modern Analysis of Customer Surveys: with Applications using R*, John Wiley & Sons, Chichester 2012.

Mei-Ping X., Wei-Ya Z., *The analysis of customers' satisfaction degree based on decision tree model*, "Fuzzy Systems and Knowledge Discovery" 2010, vol. 6, pp. 2928-2931.

Nicolini G., Salini S., *Customer satisfaction in the airline industry: The case of British Airways*, "Quality and Reliability Engineering International" 2006, vol. 22, issue 5, pp. 581-589.

Steinberg D., Colla P., *CART. Interface and Documentation*, San Diego, Salford Systems 1997 (http://www.salford-systems.com/).

Thomas E., Galambos N., *What satisfies students? Mining student-opinion data with regression and decision tree analysis*, "Research in Higher Education" 2004, vol. 45, no. 3, pp. 251-269.

Tutore V.A., Siciliano R., Aria M., *Conditional classification trees using instrumental variables*, "Advances in Intelligent Data Analysis VII" 2007, pp. 163-173.

PROBLEM MASKOWANIA ZMIENNYCH W IDENTYFIKACJI DETERMINANT JAKOŚCI USŁUG Z ZASTOSOWANIEM MODELU CART – NA PRZYKŁADZIE BADANIA JAKOŚCI USŁUG PUBLICZNYCH W POLSCE

Streszczenie: Ważnym elementem badań satysfakcji klientów z nabywanej usługi jest identyfikacja kluczowych czynników wpływających na jej jakość. Ponieważ satysfakcja klienta jest kategorią złożoną, jej pomiar i analiza wymagają stosowania wielowymiarowych metod statystycznych. Jedną z nich jest metoda CART (Classification and Regression Trees). Jej zastosowanie w identyfikacji kluczowych determinant jakości usług może się jednak wiązać z wystąpieniem problemu tzw. maskowania zmiennych, który został scharakteryzowany w artykule. Możliwe sposoby jego rozwiązania przedstawiono na przykładzie wyników badania satysfakcji klientów jednego z polskich urzędów miast.

Słowa kluczowe: CART, maskowanie zmiennych, satysfakcja klienta.