Data Mining for Decision Support in Marketing: A Case Study in Targeting a Marketing Campaign

Bojan Cestnik[1, 5], Nada Lavrač[1], Filip Železný[2],

Dragan Gamberger[3], Ljupčo Todorovski[1], Miro Kline[4]

[1] Jožef Stefan Institute, Ljubljana, Slovenia
[2] Czech Technical University, Prague, Czech Republic
[3] Rudjer Bošković Institute, Zagreb, Croatia
[4] Kline&Kline, Ljubljana, Slovenia
[5] Temida, Ljubljana, Slovenia

Abstract. The paper presents a case study in targeting a marketing campaign for a specific non-alcoholic beverage brand name, based on the results of completed questionnaires about brand name recognition in Slovenia. First, we give a motivation for the task. Then, we briefly describe the data and explain the preprocessing steps. In the main part of the paper we highlight the major steps in actionable knowledge generation form the studied task. The paper concludes with lessons learned and directions for further work.

1 Introduction

One of the most important tasks of a marketing expert is how to efficiently target a population for advertising a specific product or service (e.g. Berry and Linoff, 2000). Usually, there is a trade-off between the cost of communicating a message to everybody, on one hand, and a loss due to selecting a too narrow population segment, which may result in missing some of the potential customers, on the other hand. Therefore, the area is particularly suitable for applications of Data Mining (Škrjanc, et al., 2001) and Decision Support methods (Bohanec, et al., 2002; Cestnik and Bohanec, 2002).

Among the available tools that can help marketing experts accomplish their tasks, Data Mining tools have, since recently, gained significant importance. Namely, in most large companies the quantity as well as the quality of data about customers and orders has increased dramatically in the last decade. In addition, there are several global studies and reports about customer behavior published each year. Statistical tools that were successfully applied in the past to study global phenomena can no longer provide sufficient answers to specific, individually focused questions (Berry and Linoff, 2000).

In this paper we present a case study in targeting a marketing campaign for a Slovenian natural non-alcoholic sparkling beverage brand (since we are not allowed to uncover the identity of this brand, the brand name is labeled *X*). The study is based on the data that were collected from a public survey done by the marketing agency

Kline&Kline using dedicated software for questionnaires *QA* developed by the Temida company. More specifically, the task was to identify the characteristics of those consumers that do not yet know and/or use brand *X*. In reality, the target population is further segmented to (A) drinkers of other non-alcoholic sparkling beverages and (B) others that do not drink any beverages of this kind. The latter can, by all means, be excluded from our target since it is fair to assume that they are not inclined to use the product generically. In another words, in our campaign we would like to contact the users of competitive brands and present them with the qualities of new product *X*. However, the marketing expert in our team pointed out one additional constraint. There seems to be one particular brand (labeled brand *Y*) that is so firmly positioned on the market that it would be a wise idea to exclude also the users of this particular brand from our target population. In fact, it is reasonable to expect that the regular drinkers of brand *Y* are very unlikely to change their prevalent behavior no matter what our arguments about the brand *X* are.

The rest of the paper is organized as follows. The next section describes the steps required to accomplish the case study in a more detailed manner. A special emphasis is dedicated to data description, data preprocessing, and actionable knowledge generation. In conclusion we present some lessons learned and give directions for further work.

2 Data Mining for targeting a marketing campaign

This section first describes the data used for the study. Then it presents two major tasks that were accomplished within the study: data preprocessing and subgroup discovery aimed at actionable knowledge generation for the development of a marketing campaign. The marketing problem addressed in this paper is how to target a marketing campaign for a Slovenian natural non-alcoholic sparkling beverage brand. The starting point is a relational database obtained by interviewing potential customers. The targeting task is the problem of selecting potential customer subgroups that can be targeted by advertising campaigns.

Business to be successful have to view their markets as consisting of distinct groups of consumers, each with their own distinct set of requirements. True consumer segmentation has such a profound impact on a business that getting it right cannot be left to chance (McDonald and Dunbar, 2000). Literature on market segmentation underlines the view that markets and their segments are clusters of potential customers (Kotler 1991; Tynan and Drayton 1987). Only some of them are suitable to be selected and approached or targeted to pursue with right offers from a particular marketing agent.

Every segmentation yield several segments and the key question is how to help marketing planners to decide which ones are likely to be most promising. The rule of a thumb is to target the segment with the heaviest users. On surface this makes the most sense; however, there are some strong indications that such approach is not the most promising one (Myers 1996, p.21-22):

- One or more major competitors may have already targeted this group successfully;

- The company product line is not well designed for this group;
- In reality there are no heavy users;
- The company is to small to go after the heavy-user segment;
- The company and its agency want to develop different marketing campaign for its brand for all usage groups.

2.1 Data description

The input data for the targeting task was gathered as a result of a survey done by the Kline&Kline marketing agency about brand name recognition and reputation. The data consist of three relational tables: (1) general customer responses and demographic facts, (2) responses about specific brand names, (3) verification of brand names recognition.

The first table contains customer responses to general questions and demographic facts. Each customer is identified by unique key *Q*. There are 2013 rows (customers) in the table. The customers are described by their answers concerning age, level of education, occupation, area of living, consumer preferences and habits like what TV programs they watch and what newspapers they read regularly.

The second table contains responses about specific brand names. There are 300 different brand names analyzed in this survey; to avoid an overkill each customer is given a subset of 15 brand names to evaluate with respect to their recognition, reputation and usage. Therefore, the second table contains *Q* as a foreign key and *D* as a key for a specific brand. There are in total 30195 lines, representing 2013 * 15 answers.

The third table contains control questions that can be used to estimate the quality of answers in the second table. Here, additional questions about each brand name with respect to its product category are asked. For example, if a customer responds that he knows and uses a specific brand, and at the same time categorizes the brand in a wrong product category, one can reasonably conclude that the customer's knowledge about the brand is questionable. Either the brand is mistakenly mixed-up with some other brand or he simply made a mistake. For the specific task described in this paper we did not take the third table into account.

The most important attribute for our study is the frequency of consumption of a particular brand *D*. It is included in the second table and can have values from 1 to 5, 1 meaning that the customer does not know the brand (and therefore does not use it), and 5 meaning that he regularly uses it. For further analysis we took answers 4 and 5 as positive (uses the brand) and 1, 2 and 3 as negative.

The need for data preprocessing arose from the fact that the concept "drinker of brand *X*" can only be determined for a limited number of respondents: only to those that were originally asked this question. The same holds for two other concepts: "drinkers of brand *Y*" and "drinker of non-alcoholic sparkling beverage brands". Note that when combing the sparse concepts with logical operator AND one can quite easily end up with an empty set. Therefore, the concepts need to be stretched to all respondents in order to obtain a set large enough to produce reliable actionable descriptions.

2.2 Data preprocessing

Every respondent got to evaluate only 15 brands out of 300. This means that only one customer out of 20 was asked about the recognition and reputation of a specific brand. In order to combine several different concepts it is necessary to construct a classifier that can be used to fill in the data about using/non-using a specific brand for the rest of 19 customers out of 20. The process of data preprocessing is described in full detail in (Železný, et al., 2002).

In order to fill in the missing concept assignment, we used the CN2 algorithm (Clark and Niblett, 1989) to learn from the known training cases to produce classification rules that can be used to assign a class to other instances that could not initially be evaluated weather they belong to a given concept. If, for example, the concept is "drinker of brand *X*", then only the respondents that were asked about brand *X* can originally be included in or excluded from the concept. The learned CN2 rules were used to classify all the other respondents.

A common experience of data-mining efforts applied as well in achieving the mentioned goal: several tools had to be applied to obtain a useful result. Besides CN2, we employed the Sumatra Transormation Tool (Aubrecht, et al., 2002) and two Prolog programs, in the sequence described bellow.

Firstly, the transformation of responders' grading (1-5) into positive and negative examples of particular concepts was dictated by the principles given by the domain-expert and in some cases it was not trivial. For example, although the concepts of "X – drinker" and "Y – drinker" are straightforward (satisfied by persons who gave a response equal or greater than 3 to the question regarding the consumption of X or Y, respectively), the concept of "drinker of *other* sparkling drinks" was defined in a more complicated manner. Positive examples were people not asked about brand Y, who responded with confidence equal or grater than 3 for at least one sparkling drink. Negative were those asked about at least one sparkling drink excluding X and Y and the maximum of the answers to the questions on sparkling drinks was equal or smaller than 3. Positives and negatives are thus disjoint, but their union is a proper subset of all responders (not all responders qualify to be examples). Furthermore, the required notion of "other sparkling drink" is itself a pre-defined concept (among all drinks) as well.

To preserve clarity of such transformations, we decided to encode them declaratively in Prolog. The input data thus had to be transformed into Prolog facts and Sumatra TT provided this service.

The CN2 algorithm was then applied on the example file that was generated by the Prolog program. It then produced a set of rules of the form

```
Body → Class_Assignement
[N1 N2]
```

where `the Body` is a conjunction of attribute value assignments, where attributes are mostly demographic parameters of the responders, their preferred TV channels, journals etc. The numbers N1 (N2) define the quantity of positive (negative) examples complying with the rule's body conditions. An example of the rules is e.g.

```
IF journal_15 = 0 AND tv_1 = 1 AND tv_8 = 1 THEN class = 1
[13 0]
```

Such rules were easily retranslated into a format interpretable by a Prolog machine. Subsequently, to predict a response of a person in the range of 1 to 5, the distributions (*N1*'s and *N2*'s) are summed up for all rules whose bodies are satisfied by the attribute assignments of the given person into a cumulative distribution ($N1_{sum}$ and $N2_{sum}$). Then the probability *P* of that person satisfying the given concept (being a consumer of the drink) can be estimated as

$$P = N1_{sum} / (N1_{sum} + N2_{sum})$$

and this value was used to calculate the most likely response *R* in the original range of 1 – 5, by a linear projection

$$R = 1 + 4\,P\,.$$

In order to assess the expected error of the CN2 induced concepts, we primarily (for evaluation purposes only) split each of the three concepts into 70% of training and 30% of testing data. For obtaining the final concept descriptions, we, however, used 100% of the data for training. The results are presented in Table 1. Although the accuracies on the testing sets are only slightly higher than the majority vote accuracy, it should be noted that this kind of measurement treats the prediction task as pure binary classification, whereas in fact CN2 algorithm produces a probabilistic classification. By measuring the average squared error on probabilities, we obtained the figures in Table 2, which are more favorable for the induced models. Note that the average square error is measured on training examples with known outcomes.

Table 1: Classification accuracy of the induced concepts

Method	Brand X	Brand Y	Other sparkling
Majority vote	65.8	81.0	57.5
Training accuracy	95.1	92.4	83.2
Testing accuracy	66.7	84.8	59.5

Table 2: Average squared error of the induced concepts on the training examples

Method	Brand X	Brand Y	Other sparkling
Random guess	3.64	3.26	3.55
Majority vote	2.76	1.78	3.03
Induced concept	1.02	0.76	1.29

2.3 Discovering important factors for Decision Support

By data preprocessing we managed to obtain the data set that is suitable for further investigation of the concept under study. From Tables 1 and 2 one can observe that the classifications are not performed with high statistical significance. This is mostly due to the fact that the inherent nature of the domain (targeting a population in marketing) is inexact and probabilistic. For example, it makes no sense saying that all the readers of a specific newspaper will buy a certain product; however, it might be fair to conclude that the probability of them buying the product is higher than average.

Indeed, instead of trying to describe the final concept with a rule, we find it more beneficial to present it by listing its supporting factors as well as its opposing factors, which follows the basic principles of Bayesian analysis (e.g. Berger, 1985). These factors were found in such way that they respectively maximize or minimize the conditional probability of the concept. Only the factors with statistical significance higher than 99% were selected as influential and were included in the listings. The marketing expert found such descriptions very intuitive and easy to apply in practice, especially in the cases where the corresponding group can be named with a suitable metaphor. It seems that such a disjunctive approach is particularly suitable in marketing (and possibly related domains), where the task is to increase the probability of a certain event (order, buy, reply) in a target population and not to accurately describe a portion of the target population.

In our case, we first have the concept of "drinker of brand *X*". This group can be characterized by the following supporting factors:

- The customers come from a central Slovenian region,
- Label "Monitored food" is neither important nor unimportant,
- They regularly read *Dnevnik* and/or *Mladina*, and
- Their education degree is higher or equal to university degree.

On the other hand, the non-users of brand *X* can be characterized as follows:

- Availability of a product in different quantities is not important, and
- Product price is not so important.

The next concept is "user of non-alcoholic sparkling drinks". The members of this concept can be characterized with the following descriptions:

- They read regularly *Finance* and *Večer,*
- Their education level if higher or equal to university degree,
- Healthy food is important,
- They watch regularly *Gajba-TV* (local TV station),
- Availability of a product with different tastes is not important, and
- Nice product outfit is important.

In contrast, the opposing factors for the above concept are the following:

- The read regularly *Naš dom*, *Mag*, *Nedeljski dnevnik* or *Gea,*
- They do not watch *POP-TV,*
- Good commercials are not important,
- Their age is 61 and over.

The last simple concept is “user of brand *Y*”. We found the following supporting factors:

- They are younger than 20 years,
- They read regularly *Nedeljski dnevnik,*
- Adequate brand name is very important,
- Availability of a product in different quantities is important,
- They have a free profession (lawyer, architect, artist, …),
- Healthy food is important,
- They watch *TV3*, and
- Good commercials are very important.

The non-users of brand *Y* are characterized by the following factor:

- Good commercials are not important.

Here, let us restate our target population: they are the ones that do not yet know or use the brand *X*, but do drink other non-alcoholic drinks, with the exception of those that regularly drink brand *Y*. To describe it, one might use the combination of the above supporting factors; however, since we stretched the concept in the data pre-processing phase (by learning labels for missing concepts), we can find the following supporting factors for the combined concept:

- Availability of a product in different quantities is not important,
- Good commercials are not important,
- Different tastes of a product are not important,
- Good name of a product is not important,
- Popularity of a product is not important, and
- They read *Večer* regularly.

Here is the list of factors against the combined concept:

- Good commercials are important,
- They read *Dnevnik*, *Nedeljski dnevnik*, *Mladina* and/or *Naš dom*,
- Good name of a product is important,
- They regularly read more than 4 newspapers,
- They are from central Slovenian region, and
- Their education level is higher or equal to university degree.

One important observation to be made is that the supporting factors of the combined concept are not necessarily included in the basic concepts. For example, the concept of reading more than 4 newspapers did not appear in any of the basic concept descriptions. However, there are also some strong factors that can be traced from the combined concept to the basic one. For instance, the consumers from the central Slovenian region tend to be excluded from the combined concept, because they tend to be more then average consumers of brand *X*.

Note that the descriptive factors differ in how actionable they really are. If the description includes readers of a specific newspaper, the information can be used for targeting whereas there is not much that can be changed about the target audience of the newspaper, provided that you are not the editor in chief. Also, if one of the characteristics of the target population is that they don’t value good commercials, you can’t reach them by making bad commercials. On the other hand, if they think that healthy food is important or that the nice product outfit is important, you can address their need by stating the healthy ingredients of your product or introducing its nicer outfit.

When describing subgroups with a set of influential factors it is important to be able to substitute the factors with a proper metaphor. For example, the first five factors from the description of the target population can be, according to the marketing expert, formulated as store-brand consumers. These consumers do not buy established popular brands. They settle for no-brand products that are usually sold under the store brand name, are packed in simple packages and offer good quality for a reasonable price. Such consumers can be addressed by low-profile advertising. According to the marketing expert the discovery of this piece of knowledge is substantial for the marketing analyst when planning and directing a marketing campaign.

3 Conclusions and lessons learned

In symbolic predictive induction, two most common approaches are rule learning and decision tree learning. The goal of rule learning is to generate separate models, one for each class, inducing class characteristics in terms of class properties occurring in the description of examples. Classification rule learning produces characteristic descriptions that are usually generated for each class by repeatedly applying the covering algorithm. In decision tree learning, on the other hand, the rules that can be formed of paths leading from the root node to class labels in the leaves represent discriminating descriptions, formed of properties that best discriminate between the classes. Therefore, classification rules serve two different purposes: characterization and discrimination. They form actionable knowledge, when the action to be performed is classification and/or prediction. This means actionability just in terms of determining class membership of individual non-labeled instances, and not necessarily uncovering the properties of population that can guide a decision maker in directing a targeting campaign.

In a marketing campaign targeting potential clients of a natural non-alcoholic sparkling drink, the target class are people who do not use or know this brand, but do drink other non-alcoholic drinks, except of those who regularly drink beverages of world-famous brands. Why should consumers of world-famous brands be excluded from the target? According to the marketing expert, these consumers are very unlikely to change their habits; therefore it makes no sense to direct a marketing campaign at these consumers. Moreover, in the discussion with a marketing expert it became clear that the negative class should not be formed of all the other consumers. Limiting the population to non-alcohol drinkers makes more sense in uncovering specific properties of the target population. If, for example, alcohol drinkers were to be included in class negative, the subtle differences between people who don't use the Slovenian brand, but do drink other non-alcoholic drinks would be hidden by much stronger regularities discriminating non-alcohol drinkers to those drinking alcohol drinks. Note that even subtler properties could be uncovered if the entire population were limited to consumers of non-alcoholic dark-colored sparkling drinks, since the color of the analyzed brand is dark.

In the marketing problems where the task is to find significant characteristics of customer subgroups who do not know a brand compared to the characteristics of the population that recognizes the brand, one of the lessons learned is that the ROC space (Flach and Gamberger, 2001) is very appropriate for the comparison of induced models. Only subgroups lying on the convex hull may be optimal solutions and all other subgroups can be immediately discarded. When concrete parameters of the mailing campaign are known, like marginal cost per mailing and the size of the population, they define the slope of the lines with equal profit in the ROC space. Movements in the ROC space along these lines will not change the amount of the total profit, while movements upward or downward will increase of decrease the profit, respectively. The optimal subgroup in a concrete marketing situation is the point on the convex hull that has an equal profit line as its tangent. Additionally, in the direct marketing problem it was detected those optimal subgroups may be combinations of induced subgroups. In order to make use of this possibility, we have induced many potentially good solutions by changing the generalization parameters. In the problem of targeting a marketing campaign for a Slovenian natural non-alcoholic sparkling drink brand, most of already described techniques have been used; additionally, much effort was spent on data preparation.

One of the main requirements for successful application of Data Mining methods in marketing is that the learned concepts are actionable. This is, however, in most cases hard to achieve. If, for example, the learned concept includes customers of a certain age and living in a certain area, there is not much to act about. The only thing one can do is to take it into account when targeting the commercial message. On the other hand, if the learned concept includes customers that were sent a promotional material, then we can actively enlarge the coverage of the concept by sending some additional catalogs.

When describing subgroups with a set of influential factors it is important to be able to substitute the factors with proper a metaphor. For example, the first five factors from the description of the last target concept in section 2.3 can be formulated as store-brand consumers. Those are the consumers that do not buy established popular brands; instead, they settle for no-brand products that are usually sold under the store brand name, are packed in simple packages and offer good quality for a reasonable price. In marketing such consumers can be addressed by low-profile advertising. Although this conclusion is relatively simple and seems rather strait forward, it offers a crucial leverage to the marketing analyst in planning a marketing campaign.

For further work we envisage some more replications of the principle described in this paper on different marketing problem areas. In these new cases we plan to specifically monitor relations between different levels of actionable knowledge and its influence to potential use in practice.

Acknowledgment

The work reported here was in part supported by EU project SolEuNet, IST-11495, and by the Slovenian Ministry of Education, Science and Sport.

References

Aubrecht P., Železný F., Mikšovský P., Štěpánková O.: SumatraTT: Towards a Universal Data Preprocessor. *Proc. 16th European Meeting on Cybernetics and System Research*, vol. 2, p. 818-823, Austrian Society for Cybernetics Studies, ISBN 3-85206-160-1, 2002

Berger J.O., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, 1985.

Berry M.J.A., Linoff G.S.: *Mastering Data Mining, The Art and Science of Customer Relationship Management*, Wiley, 2000.

Bohanec M., Rajkovič V., Cestnik B.: *Report on Decision Support Practical Cases, Phase II*, Jožef Stefan Institute, Ljubljana, Report DP-8512 (2002).

Cestnik B., Bohanec M.: *SolEuNet Report on the repository of problem descriptions*. SolEuNet Report D6.6 (2002).

Clark P., Niblett T.: The CN2 induction algorithm. *Machine learning*, 3(4):261-283, 1989.

Flach P., Gamberger D.: Subgroup evaluation and decision support for direct mailing problem. *Integrating Aspects of Data Mining, Decision Support and Meta-Learning* Workshop at ECML/PKDD 2001 Conference.

Kotler P.: *Marketing Management: Analysis, planning and control*, Prentice-Hall, Englewood Cliffs, 1991.

McDonald M., Dunbar I.: Using structured processes and systems to help managers develop strategic segmentation, *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 9, 2, 109-127, 2000.

Myers J.H.: Segmentation and Positioning for Strategic Marketing Decisions, American Marketing Association, Chicago, 1996.

Škrjanc M., Grobelnik M., Zupanič D.: Insights offered by data-mining when analyzing media space data. *Informatica* 25(3), (2001) 357–363.

Tynan A.C., Drayton J.: Market segmentation, *Journal of Marketing Management*, Vol. 2, 3, 301-335, 1987.

Urbančič T., Križman V., Kononenko I.: *Review of AI Applications*, Jožef Stefan Institute, Ljubljana, Report DP-7806 (1998).

Železný F., Gamberger D., Todorovski L.: *Data Processing for a Marketing Application*, Jožef Stefan Institute, Ljubljana, Report DP-8576 (2002).