Profiling of Customer Data Base through a Sample Survey

A Data Mining Approach for business solution

by

Dr. Furio Camillo,

Professor of Business Statistics and Micro Data Mining

Alma Mater Studiorum Università di Bologna

40122 Bologna, Italy

Email: furio.camillo@unibo.it

Phone: +39 051 2098251

Dr. Caterina Liberati

Researcher of Business Statistics

Alma Mater Studiorum Università di Bologna

40122 Bologna, Italy

Email: caterina.liberati@unibo.it

Phone: +39 051 260187

Dr. Kuriakose Athappilly, Professor

Quatitative Methods and Information Technology

Western Michigan Univeristy

Kalamazoo MI 49008

Email: athappilly@wmich.edu

Phone: 269 387 5405

Profiling of Customer Data Base through a Sample Survey
A Data Mining Approach for business solution

1. Profiling Customers

The customer profiling analysis today represents the target activity of marketing program because it allows businesses to implement a solid marketing strategy. The recent development of the *customer centric* vision is impacting even the Italian market and it is pushing companies to maximize the client's profitability by building a one-to-one relationship with the clients. Such relationship consists of different interactions from client to business and vice-versa. If such interactions are well built and managed by businesses, they will reinforce the company value and the customers trust on one hand, and on the other, they will help businesses to enrich their own clients' information base. This is what the intent and purpose of Customer Relationship Management (CRM) in nutshell. In a literal sense, CRM is defined as system of interactions with clients, which integrates data coming from different channels into a unique data base shared with each business areas that have a contact with final customer. In a brader perspective, CRM is a set of strategies, processes, culture and technologies, which allow organizations to gain competitive advantage and improve their performance through a deeper knowledge of customers needs. Benefits of applying CRM processes are well described in the literature about CRM. They inclue:

1. identifying best business customers - in terms of customer acquisition and retention
2. choosing the appropriate medium to reach company best customers - as media planning
3. prioritizing best business markets and possibilities for expansion - marketing focus
4. selecting a custom model for addressing company marketing strategy towards successful challenges - in terms of analytical service..

CRM Analytics:

To bring the CRM vision into reality, an important process is the data analysis. For data analysis, several technologies and techniques for recognizing and tracking patterns within data are used. This process helps businesses sift through layers of seemingly unrelated data for meaningful relationships, where they can anticipate, rather than simply re-act to customer preferences. Analytical CRM deals with strategic, effective and efficient use of data in order to provide management with good decision-making possibilities. It requires the existence of Customer Data Warehouse (CDW), which integrates data coming from several different fronts and it allows an easy access to the data. Data Warehouse is 'a blend of technologies aimed at the effective integration of operational databases into an environment that enables the strategic use of data' (Fayyad and all, 1996). These technologies include relational and multi-dimensional database management systems, client/server architecture, metadata modeling and repositories graphical user interfaces, and much more. In such context CDW needs a continuous enrichment of information about customer behaviors, their values and purchase models. Therefore, the fundamental way to gather data is to interview clients in order to create a scoring model which produces for each subject the probability of belonging to a specific cluster identified in a data explorative process. Obviously the relevant information obtained is referred only to the customers, who belong to sample used in the survey. Obviously it is a great limitation.

In this paper dealing with a specific case study, we conduct a profiling of the whole Customer Base in order to get what we call *Database Marketing*. It supports a variety of business processes and involves a transformation of the data base information into a knowledge base for actionable business decisions. It will incorporate not only the features associated with the (potential) customers, such as social demographic indicators but also history of purchases and/or behavioural variables. Such customer profiling is realized into two steps: 1) it applies usual explorative data mining techniques such as cluster analysis in order to find groups structures in the sample; 2) we select the the target variables, from the the classifications obtained by initial exploration, for a supervised classification model (such as Discriminant Analysis). Thus it is possible to assign each observation (record) of the Customer Base to a group depending on minimum distance between the single subject and the centroids of the clusters identified. This individual information about the classification of each subject could then be used in the informative process of Customer Relationship Management. Thus it becomes a fundamental tool for planning marketing strategy in order to cultivate clients and approach them on a one-to-one basis. By combining the customer profiling lifestyle and market information, businesses can implement more effective, customized cross-selling and retention strategies. This approach also allows them to adapt to changes at each stage of their customer's relationship spectrum, from customer acquisition to customer retention to overall growth.

2. Italian Mobile Phone: A Case Study

Italian Mobile Phone market consists of four big operators. Presently, the company has about 95% penetration rate of service. One of the operators launched on an advanced CRM analytical project. The fundamental idea was to conduct a segmentation study of clients' values about the usage of mobile phone. Such segmentation could be used to enrich the Customer Data Base by soliciting the customers according to their value hierarchy as far as the usage of mobile is concerned. The segmentation will be used in all the surveys directed to the customer base and each client will be assigned a descriptive label about the mobile phone imagery. The innovative part of this project is the realization of a cooperation between offices dedicated to CRM and the marketing folks, when segmentation is the target focus on which all the communication efforts and all the micro and macro business strategies converge. The project has become a platform of many more involved "actors", consisting of the offices of marketing strategies and the CRM for the ordering company, an institute of market researches which prepared and conducted the interviews to a customer sample, a business intelligence company doing the data mining project, and the University of Bologna as scientific supervisor of the statistical part. In this team approach, the definition of the questionnaire, the initial exploration of the collected data and the building of the clusters have been carried out under the supervision of all the staff.

The data miners' task was to extend the initial explorative study into the analytics realm by employing predictive models on data sets consisting of the information about values of the mobile phone obtained with the survey to whole customer population, the so-called *colouring* phase of the whole Customer Base with the information obtained on a sample and an investigation directed to the field by questionnaire**.** In the next sections we will show the technical phases of the project wherein we will focus our attention on the new advanced techniques of data mining, namely, the kernel machines..

3. Analysis Strategy

The analysis strategy has been structured into two alternative and parallel processes which produced independent results, in order to verify the consistency of the classification found. The first process has been conducted by the business intelligence company which applied standard data mining technique to evaluate supervised classification model using tree algorithm. This approach is a widely employed in data mining project because of implementation ease of data integration and evaluation modeling process. The second solution is provided by university researchers applying new and advanced tools of data mining algorithms, known as *kernel machines.* Kernel machines classification patterns are still in the developmental stage in the academic world. Naturally they present a lot of problems since it is built on a semi-automatic process of estimation.. However, they represent the new frontiers of statistical research so we suggest their usage for the relevant performance improvements that it is possible to derived.

The target variable of these discrimination models has been estimated by employing a hierarchical cluster analysis on a customers sample of 10,021 subjects, obtained by a survey composed of 90 items about their lifestyle behaviours, ideas on mobile phone technology and its usage. This process led to a segmentation composed by four distinct groups. They are:

1. “Functional people” (20.8%) generally males between 35-44 years old, with low education, they call and receive calls from people belonging to the same cell brand company, they use widely short message service (sms). They use mobile phone especially for professional needs, choosing it on the base of functional features.
2. “Practical people” (35.9%) generally males and females of 45-64 years old, with medium/high education. Thye use phones especially in the afternoon (3:00-6:00 pm) spending about 30 € per month and use all the services because they have an intense social life.
3. “Techno-funs” (27.59%) generally males of 25-34 years old and highly educated. They live in the North -West area of the Country. They are for the most part professional men or students, who spend about 45 € per month. They buy high technology tools for their mobile phone. They also have an intense social life.
4. “Matures silent” (15.8%) generally over 65 years old females, who live mostly in the big cities of the North-West of the Country. They use the mobile phone only for emergencies. On other occasions, they prefer to call with home phone.

Observations about the four segmented groups: The following observations about those four segmentations are worth mentioning. Generally when CRM project is implemented in a mobile phone company in order to find profitable clusters and/or to study modifications of a classification, that company already knows that the “young” group is always pre-identified. In other words, such a group is not produced as result of a classification algorithm but it is built *ex ante* depending on subjects’ age (generally under 25). That occurs because marketing practice about this topic has a unique voice the young people represent community behaviours[1] of massive usage of sms, so it has to be treated as a group. In our case study, we employed the cluster algorithm not on the entire sample using *a priori* exclusion of the young customers, according the marketing folks desires. But *ex-post* we re-allocated such subjects to the 4 segmentation clusters. Using a discriminant model, we did a posterior study to evaluate the corresponding probability of each young customer to the 4 clusters. We found very interesting results: the young people could be split very well, according their answers to the survey, into the firsts three groups (Functional people, Practical people and Tecno-funs). The figure below shows that in this case study the young people segment is not reasonable. In fact the posterior probability of each young subject of belonging to one of the 4 clusters is always over 50% even for the bottom percentiles.

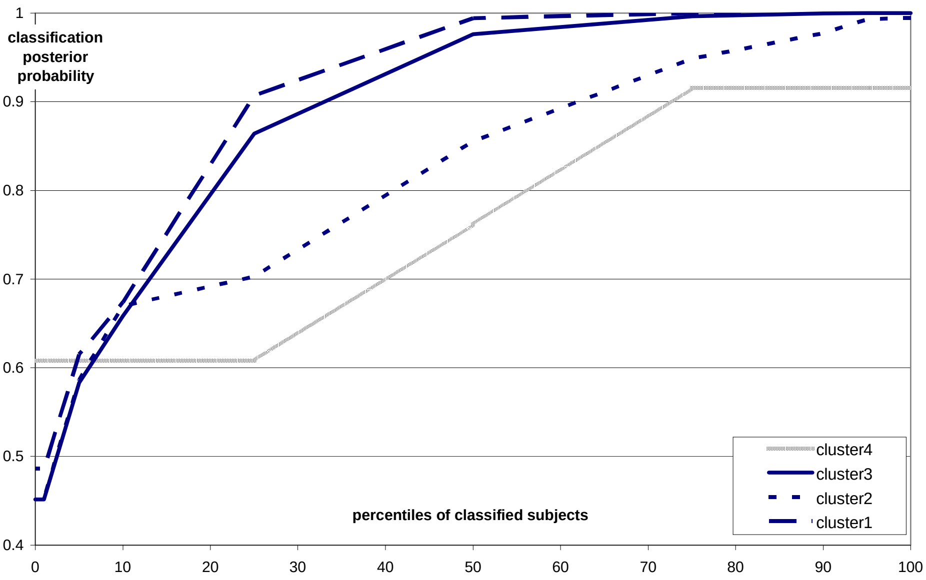

Figure 1: *Cumulative functions of classification posterior probability*

The second step of the study was to build the explicative variables of the discriminant models. The information used in this step, consisted of indicators derived from the Data Warehouse. This information was used in order to score and update the whole Customer Base. There were about 200 covariates in total. Hence a variable selection method was employed to find the best predictors. To this end, we selected those 50 indicators to re-build the partition found with the cluster analysis with an 80% of approximation. On these we applied a factor analysis technique, known as Principal Component Analysis, for obtaining orthogonal variables in order to protect us from the danger of the multicollinearity problem.

We have to point out that the Canonical Correlation between items and the factors obtained are about 70%. This result is not unexpected since the congruency between ideas and behaviours is never perfect. This might present a problem because it could affect the models' performances. The classification tree implemented on this case is a non parametric tool of SAS enterprise-miner so we do not intend to dwell upon the setting and features of this standard algorithm. On the contrary, we think that is important to dwell on the kernel methods which are not widely known at the present time.

The central idea of the kernel method is that we can work on the feature space through kernel functions even without knowing the nonlinear feature mapping or the mapped feature space explicitly, as long as the problem formulation depends only on the inner products between data

---

[1] Community behavior means a specific behavior of subjects which use to call and/or are called by people of the same mobile phone company.

points. This is based on the fact that for any kernel function κ satisfying Mercer's condition (Cristianini Shawe-Taylor, 2000) there exists a mapping ϕ such that

$$< \Phi(a), \Phi(b) >= \kappa(a,b) \tag{3.1}$$

where <,> is an inner product in the *Feature Space* transformed by the κ (Burges, 1998)

The reformulation of DA in the feature space is very similar to the LDA space, where the decision boundary is obtained by maximizing the ratio:

$$J(w) = \frac{w' S_B^{\Phi} w}{w' S_W^{\Phi} w} \tag{3.2}$$

where S_B^{Φ} S_W^{Φ} are Between and Within covariance matrices in the Feature Space (Baudat and Anouar, 2000).

This maximization problem can be resolved by finding the leading eigenvectors of the $(S_W^{\phi})^{-1} S_B^{\phi}$. The number of the discriminant functions we obtain is equal to the number of groups minus 1. The discriminant scores are the linear expansion of the training patterns in the feature space:

$$\alpha = \sum_{i=1}^{n} w_i \phi(x_i) \tag{3.3}$$

Therefore, the improvement employed in this technique is that we have decision functions which are linear in the Feature Space correspond to nonlinear functions in the input space. (Figure 2).

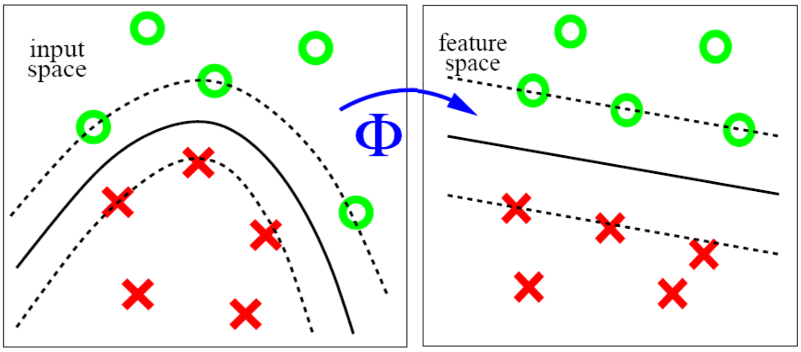

Figure 2: *Decision boundary in the Input Space vs. Feature Space*

4. Results

The results obtained are different and they indicate how effective the two strategies are in terms of performance expressed by misclassification error rate. Both studies have been conducted by splitting the sample into two subsets, the first one with 70% of the subjects sampled, has been used to train the algorithm. The second one with the remaining 30% of the subjects sampled, was used to validate the rules found.

As we pointed out above, he results produced suffer from the lack of perfect correspondence between items of the survey and explicative variables of the classification models. The tree algorithm, as we can see in the confusion matrixes below, which synthesize for each group the percents of correct classifications, recognizes very well those subjects belonging to the second and the third clusters (See Table 1) both in the training and in the test samples. However, the tree alogorithm presents great difficulty in correctly classifying people coming from the other clusters.

	From/to	**1**	**2**	**3**	**4**	
	1	35.8	23.3	36.8	4.0	100
a)	**2**	0.4	71.3	11.4	17.0	100
	3	1.6	36.1	62.2	0.1	100
	4	1.7	30.0	18.8	49.4	100

	From/to	**1**	**2**	**3**	**4**	
	1	15.8	31.2	37.4	15.6	100
b)	**2**	0.8	51.0	15.8	32.5	100
	3	2.6	32.0	42.0	23.4	100
	4	2.1	51.2	12.8	33.9	100

Table 1: *Tree Algorithm: Confusion matrix of the train sample (**a**) and the test sample (**b**)*

This situation poses a problem especially when we make predictions about the updated and extended whole Customer Base, based on this sample case. Kernel Discriminant Analysis (KDA), however, seems to be a superior algorithm when compared Tree algorithm as it has far advanced classification techniques built into it. In general, all these algorithms perform very well on the training set, because during training phase these methods work locally in the subjects space and therefore, they are able to adjust the decision boundary to the non-linear structure of the data. Therefore, the choice of the non-linear transformation[2] is important in order to avoid the over-fitting problem and, at the same time, to test the validity of the generated rules.

---

[2] It is used a sigmoid kernel $\kappa(x,z) = \tanh[a(x_i, z_i) + b] \quad a, b \in \Re^+$

Table 2 highlights the inability of such techniques to recognize subjects coming from cluster 1 with subjects of cluster3 as well as subjects from cluster 2 with those of cluster 4.

	From/to	1	2	3	4	
	1	81.6	0.0	18.4	0.0	100
a)	2	0.0	74.3	0.0	25.7	100
	3	17.2	0.0	82.8	0.0	100
	4	0.0	38.0	0.0	62.0	100

	From/to	1	2	3	4	
	1	41.2	23.5	23.5	11.8	100
b)	2	9.3	62.5	8.6	19.6	100
	3	17.4	21.7	51.3	9.6	100
	4	13.3	26.7	13.3	46.7	100

Table 2: *Kernel approach: Confusion matrix of the train sample (**a**) and the test sample (**b**)*

Table 2 shows conflicting results. When we examine groups 1 and 3 as well as group 2 and group 4 they show differences in their ideas, life concepts and values but they are very similar when we use their behaviors as discriminant predictors. Such evidence determines an increase of noise when the rule has to classify people in prediction process (Table2 b). That occurs especially for borderline subjects. If we preserve for each client the belonging (customer loyalty) probability to different clusters we can observe that the misclassification happens among those subjects which present low value of such probability. To avoid the misclassification error rate of the test sample increases, we may decide to assign only those subjects who present a belonging probability to a specific cluster higher than 60% in order to build reliable segmentation.

5. Conclusion and further developments

It is hard to conclude which solution is the best. It is fundamental to specify what we mean by the best solution in a marketing strategy. Often marketing campaign or personalized offers are the results of automatic or semi-automatic processes of analysis realized by the company on its Customer Base. Subsequently, first of all, the solution has to be fast and integrated with the technology framework of the company. The KDA algorithm is to be brought into a business data mining process environment. There is no commercial software available in the market that produces such solution. In fact, it is developed and run still in a mathematical programming language. Moreover, the greatest disadvantage for users to employ this technique is the limited computing capacity of calculator. KDA solutions are very time consuming for a standard workstation.

Therefore, in business projects it could be used only as a benchmark for comparing data mining methodologies.

References:

- Baudat, G. & Anouar, F., “Generalized discriminant analysis using kernel approach”. In Neural Computation, 2000.
- Bozdogan H., Camillo F., Liberati C. “One the choice of the kernel function in KDA using Information Complexity” in “Data Analysis, Classification and Forward Search” Zani S., Cerioli A., Riani M., Vichi M. (Eds.), Springer-Verlag Heidelberg, 2006
- Breiman L., “Classification and Regression trees” New York; London : Chapman & Hall, 1993
- Fayyad U., Piatetsky-Shapiro G., Smyth P. “From data mining to knowledge discovery: an Overview." In Advances in Knowledge Discovery and Data Mining, Fayyad et al (Eds.) MIT Press, 1996.
- Hastie, T, Tibshirani, R., and Friedman, J “The elements of statistical learning ” Springer New York, 2001.
- Liberati C., Camillo F. “Kernel Discriminant Analysis and Information Complexity: advanced models for micro-data mining and micro-marketing solutions,” . in “Data Mining VII: Data, Text and Web Mining and their Business Applications”, A. Zanasi, TEMIS S.A., Italy, C.A. Brebbia, Wessex Institute of Technology, UK and N.F.F. Ebecken, Universidade Federal do Rio de Janeiro, Brazil (Eds.), WIT press, 2006.
- G. Lugli e C. Ziliani , “Micromarketing. Creare valore con le informazioni di cliente”, *UTET*; Torino Italy, 2004.
- Schölkopf B., Smola A. “Learning with kernels” MIT Press, Cambridge, MA, 2002
- Shawe-Taylor J., Cristianini N. “Support Vector Machines” Cambridge University Press, 2000