

# Customer Churn Prediction Embedded in an Analytical CRM Model

Ede Lázár

Sapientia – Hungarian University of Transylvania, Romania

## Abstract

This paper presents a practical implementation of an analytical customer relationship (CRM) model, which aims to increase the customer satisfaction, thereby reducing the rate of attrition. The analytical CRM model not only manages and synchronizes customer relationship management processes, but also creates added value regarding to customers by applying mathematical, predictive methods. This presented model was implemented at a Hungarian gas service provider, and estimates the probability of churn for each customer based on the characteristics of former and present customers. The methodological approach is based on econometrical background; the analytical tool is a binomial logistic regression model. As a result this study presents that using logistic regression models as predictive analytic tool we can fulfil multiple CRM goals. Using the theoretical framework of Swift (2001) we can state that the model consists of more CRM dimensions simultaneously. These are the predicted churn probability as a customer retention dimension, and the information about the efficiency of different CRM elements, and CRM channels, as a customer attraction dimension.

**Keywords:** analytical CRM, predictive analytics, churn prediction, logistic regression

**JEL classification:** C53

## Introduction

In this paper a customer relationship model is shown, which does not only implement innovative elements into a Hungarian natural gas supplier's customer relationship management (CRM), but also reorganizes the customer service's procedure. The aim of the model presented in the paper is to integrate and fulfil multiple CRM function with strategic importance: predict the churn probability of the current customers and based on it performs actions which increase customer satisfaction, and furthermore helps customer acquisition. The basic research question is how to develop an existing CRM process to achieve these multiple CRM goals, and which is the adequate analytical tool.

An efficient CRM system contains two parts: the operational CRM and the analytical CRM (Massey et al., 2001). The operational CRM collects customer data through a range of channels such as call centre, mail, sales forces, web page, and the database is the foundation of all customer related processes. The application of predictive analytics in the CRM helps to understand customer behaviours, acquire and retain customers, and also maximize customer value (Mirzaei et al., 2014). Various technologies supports the analytical CRM systems among which "data warehouses", "CRM portals", "predictive and analytical engines", as a result customers are more detailed and effectively segmented, and managers are able to offer products and services based on customer profiles (Xu et al., 2005). Predictive analytics have been extensively used in telecommunication (e.g. churn prediction, maximizing customer value) and banking industries (e.g. credit scoring). In recently

years these techniques are rapidly generalized to other firms as well who are involved in B2C transactions with a large number of customers (e.g. public utilities service providers). The notion of big data and the potential of producing actionable information from the existing databases are the main drivers of predictive analytics application (Halper, 2011). But from the corporate implementation point of view, CRM should not be misunderstood to simply mean a software solution implementation project. Building relationships with customers is a fundamental business of every enterprise, and it requires a holistic strategy and process to make it successful (Parvatiyar et al., 2001).

The spreading practical applications was followed by the academical sphere, Swift (2000) gave one of the most significant theoretical framework to this area. According to Swift (2000) and Parvatiyar and Sheth (2001) CRM consists of four dimensions which share the common goal of increasing customer value to organization. Customer identification is the initial step in CRM life cycle and involves targeting the potential customers. The second stage is the customer attraction which involves direct marketing tools to the targeted customers. Customer retention is the most prevalent concern in CRM which comprises of customer satisfaction, loyalty and churn management. The elements of this dimension include personalization programs, loyalty programs, complaints management, churn analysis (Ngai et al., 2009). Customer development contains elements of up and cross selling and market basket analysis. The books by Srivastava et al. (2002), Miguéis et al. (2012) and Hosmer et al. (2000) were also used during the process of researching and writing this paper.

In the result section we demonstrate that these different steps of the CRM lifecycle could be managed simultaneously by an integrated analytical CRM model. In the next methodology section of the paper, the analytical tool is described.

## Methodology

In this section we present the econometrical base of the analytical tool. In the business intelligence one of the most popular model version, from the Categorical and Limited Dependent Variables (CLDV) models, is the binominal logistic regressive or known as the logit model. Using research methodological terms, the logistic is such a regression model, in which the dependent variable is a dichotomous, categorical variable, and the independent variables can be any type: interval, ordinal, nominal. This "technical advantages" has large significance in empirical researches and in variety of applications.

The logistic regression can be considered as the generalization of the lineal regression by extending its limits. The probability of customer attrition can be described with the following logistic regressive equation:

$$P(Y = 1) = \frac{e^{b_0 + b_1x_1 + \dots + b_kx_k}}{1 + e^{b_0 + b_1x_1 + \dots + b_kx_k}}$$

Where, the vectors of the independent variables ( $x_i$ ) are the properties of the product and customers, and  $b_i$  is the corresponding parameter.

To estimate the parameters ( $b_i$ ) of the logistic regression equation, the maximum likelihood (MLE) estimation method is used most commonly. Compared to the ordinary least squares (OLS) method used for the linear regression analysis, where the goal is to minimise the summed square distances between the observed and estimated value, the parameter estimation of the logistic regression maximises a probability function which is the likelihood function. The likelihood function is the

probability which estimates a dependent variables value based on the values of the independent variables. The likelihood function of the discrete dependent variable, it can vary between 0 and 1.

Keeping the aims of the analysis in mind, several advantageous features can be listed of the logistic regression based model:

1. It is a parametric method, thus compared to the neural network the effects of the different independent variables can be estimated using the parameters of the logistic regression equation. This way the effects of several factors of the always changing market environment- built in the model –can be investigated using “*what happens, if...?*” simulations. The predicting capability of this parametric method is slightly weaker than the neural network’s, but does not only produce the final output, the churn probability, but it also gives important information about which factor with what weight is responsible for the attrition.

2. The required statistical preconditions are less strict, than other parametric methods, such as the linear regression model, the linear probability model or the discriminant analysis.

3. Compared to the decision tree or neural network the result is a continuous probability value, from which a dichotomous classification (eq. loyal customer, leaving) can be made. This continuous probability variable can be used in numerous ways, the customers can be classified into more than two groups, or the rate of errors resulting from misclassified customers can also be optimised with respect to their expenses, it can be the base of a customer rating indicator.

4. The fourth advantage is not methodical, but it is very important in practical applications. Due to the relatively easy specification, easy understanding and fast learning properties, it is relatively cheap.

The goodness of fit of the model, the accuracy of the parameter estimation and the prediction of the model has high importance; an incorrectly specified model makes false predictions which can result in serious expenses. Therefore the model specification indicators should be chosen carefully. The  $R^2$ , which shows the explanatory power of the linear regression model, cannot be calculated, as the logistic regression model’s dependent variable’s deviation depends from the variable’s distribution also. Greene (2003) says that the essential difference is that, at the ordinary least squares method the criteria of the b parameter estimation is the maximisation of  $R^2$ , whereas during the estimation of the maximum likelihood, not all fitting criteria’s maximisation is aimed. Despite this or because of this several indicators were created which are related to the model fitting goodness. From these indicators the ones relevant for the practice are selected. These indicators can be classified into two groups: indicators based on the likelihood function’s value and the indicators that are based on the models prediction accuracy.

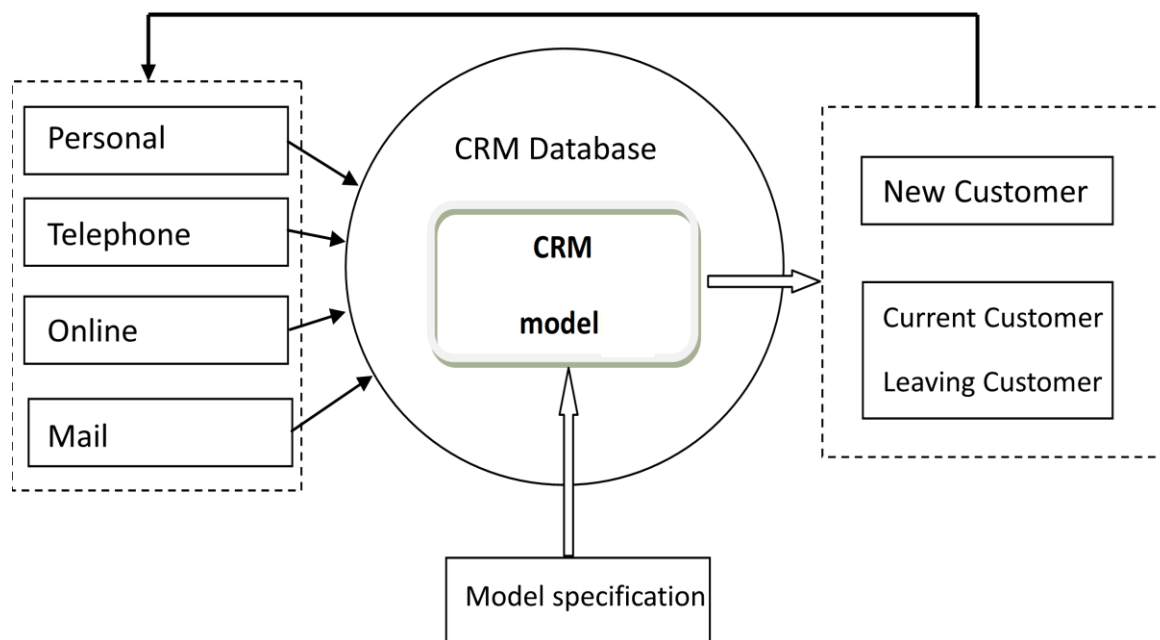
## Results

The model aiming customer relationship innovation is based on the revealed results of an empirical market research ordered by the gas service provider company and exploring all aspects of the customer relationships through quantitative methods, focus groups and mystery shopping. The results are grouped separately for potential and current customers according to the model specification requirements. From these results in the following, those customer relationship characters are determined, which help to reach the CRM model’s aims and can be appropriately quantified for the model specification aims.

The overview of the suggested solution’s model is shown on the following figure.

Figure 1

The conceptual model of the proposed analytical CRM



Source: Author's illustration

The customer relationship process starts by arranging the information, obtained from the personal, telephone, online and using mail service customer contacts, into the uniform database. This database, which absolutely follows the customer relationship activity and provides the broadest information about the customers, is the base of the CRM model. The model that fulfils the previously described properties and aims is created from the specifications made regularly (monthly, quarter yearly) or ad-hoc (marketing campaign). The CRM model is practically two models, because the two aims, customer acquisition and the probability that the current customers stay or leave, use the same method but require different model specifications. The model outputs are looped back to the clients, this way the customer relationship management can be dynamically adjusted according to the requirements.

### 1. Current customers

From the previous market research results we determined the functions need to be fulfilled by the CRM model:

- From the integrated customer relationship database the model makes predictions for every customer about the probability of attrition and determines the factors influencing this probability for the whole database. The explanatory variables of the model are data referring to customers (age, gender, address – geographical situation), data referring to the service (size of the bill, length of the service being used, reporting faults due to low heat value) and the monitoring frequency by the staff of the service provider. Exogenous data referring to prices, marketing activity of competitors are also implemented into the model.

- Even the best CRM model cannot work efficiently, if the staff has no interest in the application. The customer service staff needs to be motivated, and they need to

fell the control. From this point of view the CRM model is a control tool, it also represents the technology generating the awareness of control.

- Recommend the most suitable offer for the customer. Based on the model's analytical properties and integration, the most efficient promotions can be determined. Using segmenting tools further distinction can be made between the promotions. The model quantifies and arranges the following promotions in order according to its customer keeping effect: "Bill-angel", Remote-bill, telephone/online customer service and Premio card. This list of course may be changed by introducing new products.

- In case of breaking contract (churn) the CRM database should know the reason. During the mystery shopping "none of the administrators tried to find out the reason for leaving, breaking up with the supplier." Determining the reason for attrition is a key element for the model, it increases the accuracy of the predictions, but it is a question whether this information request is technically feasible or not.

## 2. Customer acquisition

For customer acquisition a new model is required, which is separately specified, despite that it is similar in many points to the previous model. In this case the output of the model is, whether a potential customer will become a real customer or not. The potential customers are tried to be convinced for changing supplier using the following promotions:

- "complementary, comfort" services
- guaranteed gift
- Premio card or a SP voucher
- gift voucher

As the competitors only provide partial information about administration for the supplier-switch, it is assumed to be a strong reason for those who decided to switch supplier. In case of the telephone customer service this only occurred for 71% of the cases. "The administrators at the service provider could be more helpful in this area". If the customer explains the reason for changing supplier, it can be an important explanatory variable of the model. This type of information needs to be recorded both at personal and at telephone customer service. "The administrators are practically not interested, why the new arriving customers want to switch supplier."

## Discussion

As a result this study presents that using logistic regression models as predictive analytic tool we can fulfil multiple CRM goals, the model consists more CRM dimensions (Swift 2001) simultaneously. These are the predicted churn probability as a customer retention dimension, and the information about the efficiency of different CRM elements, and CRM channels, as a customer attraction dimension.

Regarding to the research goals we determine that the analytical CRM model should have the following characteristics:

Analytical and predictive: the analytical CRM does not only control and synchronises the customer relationship procedures, but also adds customer related values with the use of mathematical methods. Some of these are also suitable for predictive functions, thus the model can predict the behaviour of the customer and the probability of dropout, based on historical data.

Dynamic: the model changes according to the recently collected customer information. This dynamic property means continuity at the data recording level, but at the CRM model level it means that model specification should be performed at

predefined intervals or ad-hoc. The model is not only dynamic from methodological point of view, but in its functionality also, the output as a feedback could continuously improve the customer service activity. For instance the churn probability as a model output is an input of information for that CRM sub process which aims to increase customer satisfaction and loyalty.

Integrated: the model integrates all information coming from the customer relationship channels (form personal, telephone, online, mail sources), and from the customer relationship activities (inspection, error corrections, etc.).

Modular: it is customisable according to customer segments. After the CRM data base, data structure and model specifications problems are solved and done, an analogue CRM model can be created cost efficiently along the customer segments. This type of "macro-segmentation" results from the model functionality, as there are separate models determining the most attracting promotions for the potential customers, the factors increasing satisfactions and decreasing attrition among the current customers. Furthermore there is possibility to apply techniques, which result a range of more accurate "micro-segmentations" within the potential and current customers.

## Conclusion

The paper discuss the logical structure and the methodological base of an analytical CRM model planted in practice, but it has the limitation that could not specify in detail the model specifications because of privacy reasons.

Answering the main research question we could affirm that analytical CRM models can fulfil more marketing functions. One of them is the customer loyalty measure, which directly determines the customer relationship, and furthermore the efficiency of the different promotions, CRM elements and CRM channels, is shown by the analysis of the estimated explanatory variable's parameters. According to this, the customer relationship management can be optimized and reviewed periodically.

Further research areas could focus on the use of different analytical tools for different functions in the same CRM system, for instance neural network for churn probability prediction, logistic regression for the estimation of different CRM channels and tools, decision tree for segmentation.

## References

1. Greene, W. (2003), "Econometric analysis", Fifth Edition, Prentice Hall, New Jersey.
2. Halper, F. (2011), „The top 5 trends in predictive analytics", Information & Management, Vol. 21 No. 6, pp. 16-18.
3. Hosmer, W.D., Lemeshow, S. (2000), "Applied Logistic Regression", 2<sup>nd</sup> edition, Wiley, New York.
4. Kumar, A., Vithala, R.R., Harsh S. (1995), "An Empirical Comparison of Neural Network and Logistic Regression Models," Marketing Letters, Vol. 61 No. 4, pp. 251-264.
5. Massey, A.P., Montoya-Weiss, M.M., Holcom, K. (2001), "Re-engineering the customer relationship: leveraging knowledge", Decision Support Systems, Vol. 32, pp. 155-170.
6. Miguéis, V.L., Van den Poel, D., Camanho, A.S., Falcão, J. (2012), "Modeling partial customer churn: On the value of first product-category purchase sequences", Expert Systems with Applications, Vol. 39 No. 12, pp. 11250-11256.
7. Mirzaei, T., Iyer L. (2014), "Application of Predictive Analytics In Customer Relationship Management: a Literature Review and Classification", Proceedings of the Southern Association for Information Systems Conference, Macon, GA, USA.



8. Ngai, E.W. T., Xiu, L., Chau, D.C.K. (2009), "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, Vol. 36 No. 2, pp. 2592-2602.
9. Parvatiyar, A., Sheth, J. (2001), "Customer Relationship Management: Emerging Practice, Process, and Discipline", *Journal of Economic and Social Research*, pp. 1-34.
10. Srivastava, J., Wang, J.H., Lim, E.P., Hwang S.Y. (2002), "A Case for Analytical Customer Relationship Management", *Computer Science*, Vol. 2336, pp. 14-27.
11. Swift, R.S. (2001), "Accelerating Customer Relationships", Prentice Hall, New Jersey.
12. Xu, M., Walton, J. (2005), "Gaining customer knowledge through analytical CRM", *Industrial Management & Data Systems*, Vol. 105 No. 7, pp. 955 – 971.

## About the author

Ede Lázár has got Master of economy diploma in 1998 Budapest University of Economic Sciences (today: Corvinus University), and PhD in 2011 at Szent-István University in Hungary, Management and Business administration sciences. He was working as analyst at Central Statistical Office, (Budapest, Hungary), and as market researcher at TNS Hungary and TNS Telecoms (US). From 2003 he has been teaching as associate professor at Sapientia – Hungarian University of Transylvania and work as a free professional. His courses are information-management, research methodology, market research. Author can be contacted at [lazarede@sapientia.siculorum.com](mailto:lazarede@sapientia.siculorum.com)