

Increasing loyalty using predictive modeling in Business-to-Business Telecommunication

Patrick Luciano¹, Ismaïl Rebaï¹, Vincent Lemaire²

(1) Orange Business Services, 27 rue Médéric, 75840 Paris, France

(2): Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion, France

Abstract

Customer Relationship Management (CRM) is a key element of modern marketing strategies. One of the most practical way to build useful knowledge on customers in a CRM system to produce scores to forecast churn behavior, propensity to subscribe to a new service... In AMEA zone (Asia, Middle East and Africa zone), the context of fierce competition may represent a higher percentage, and particularly in B2B market. But by contrast, to our knowledge, no scientific papers were dedicated and published to detail the way to improve loyalty in B2B Telco market. If we can assume at low segments similarities between B2B and B2C, some research is required in order to model B2B user behavior versus B2C behavior. This problematic stands actual as “Bring Your own Device” (BYOD) becomes more and more trendy. Answering business requirements, our team applied some B2C predictive tools with adapting them to B2B in an AMEA country Orange affiliate.

Keywords: value management, predictive modeling, loyalty, CRM, marketing, B2B, Telco.

1. Introduction - Marketing objectives: the interest of CRM tools to improve customer retention

Scholars have questioned the effectiveness of several customer relationship management strategies. The differential effects of customer relationship perceptions and relationship marketing instruments on customer retention and customer share development over time have been investigated. Verhoef [14] has showed that loyalty programs that provide economic incentives positively affect customer retention.

1.1. The role of marketing on a business's performance

One of the main objectives of the marketing function is to increase the value generated by actual and future customers. The customer is key in the company strategy. “The Consumer, not the Company is in the middle” [6].

Literature has shown that a market oriented organization supports the company's performance. Kohli & Jaworski [8] have developed a framework for understanding the implementation of the marketing concept.

Marketing academicians and practitioners have been observing that business performance is affected by market orientation [7]. Narver & Slater [12] developed a valid measure of market orientation. They analyzed its effect on a business's profitability. Using a sample of 140 business units consisting of commodity products businesses and no commodity businesses, they found a substantial positive effect of a market orientation on the profitability.

1.2. Value management

Marketing theory and practice has become more and more customer centered. Managers have increased their emphasis on long-term client relationships because the length of a customer's tenure is assumed to be related to long-run company revenues and profitability [5].

1.3. The role and interest of CRM tools

Customer Relationship Management (CRM) is organized according to the customer lifetime duration because lifetime with a firm generally is not perpetual. Consumers may be dissatisfied and find better value elsewhere [13].

1.4. The effect of loyalty programs / management

Meyer-Waarden [11, 9] examines the impact of loyalty programs on customer lifetime duration. He shows that loyalty schemes have positive effects on customer lifetimes.

1.5. Churn is a key indicator of customer loyalty

The churn, or customer attrition, is defined as the number of disconnections for the period divided by the average base for the period where the average base is the sum of the opening base at the beginning of the period and the closing base at the end of the period divided by 2. We distinguish the commercial churn from the financial churn. The churn is a key indicator, mainly in the financial sector (banking and insurance) and the telecommunication industry.

2. Forecasting / Predicting the churn

2.1. Churn Prediction

As described in the first section of this paper. The churn definition is not the same in postpaid and prepaid context. We focus here on the former: the propensity of customers to switch provider in postpaid mode even if the methodology described share many aspects with prepaid prediction.

2.2. Forecasting aspects

Given a database, the most common task in data analysis is to find the relationships between a set of input or explanatory variables (X) and one target variable (Y). This knowledge extraction often goes through the building of a model (f) which describes these relationships. For all the instances of the database, a probabilistic model allows, face to a classification problem and given the values of the explanatory variables, the estimation of the probabilities of occurrence of each class target. The model learns the relation $f: X \rightarrow Y$.

This model is trained based on time windows and applied on a sliding window (see Figure 1). Different time periods definitions exist:

- Observation period: period used to observe the customer behavior and find the right explanatory variables (see Section 2.4) to infer their future behavior
- Latency period: period required to compute the score and required by the marketing services to be able to launch the campaign based on the model scores
- Churn observation: period on which the churn is observed and used to train the model
- Churn prediction: period on which one wan to deploy the scoring model
- Churn evaluation: period on which the churn is observed and evaluate the performances of the model applied on the “churn prediction” period (back testing)

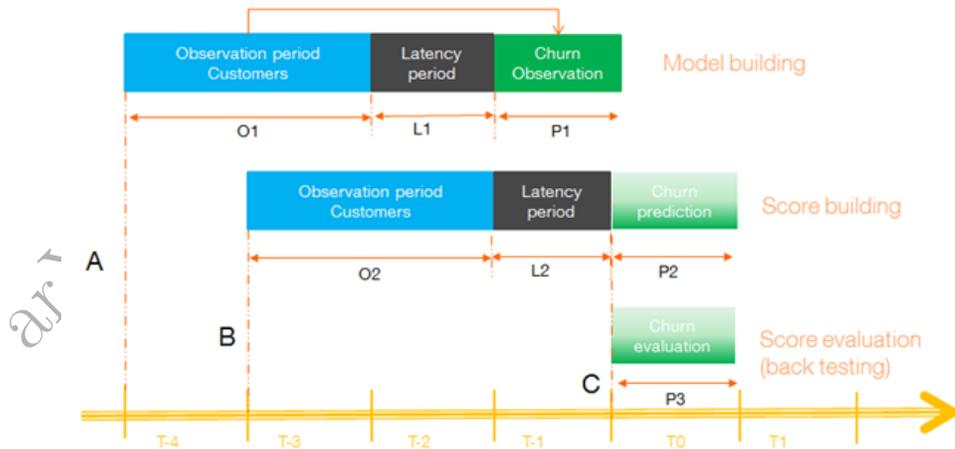


Figure 1: Time periods definition

For example:

- Step 1: at time “A” using an observation period (months T-4 and T-3) “O1” and churn observation “P1” a scoring model is trained. In this step one are in a classical data mining process [REF] in which during the modeling step the data are split in a train part and a test part to evaluate the quality of the model.
- Step 2: then this model is deployed a time “B” using an observation period (months T-3 and T-2) “O2” to predict the churn on the period “P2”
- Step 3: later a time C when the true information about the churn is available the performances of the model deployed at the step 2 are evaluated (back testing).

Note 1: Usually the size of O1, O2 and L1, L2 and P1, P2, P3 are the same respectively. More L is wide more the problem is difficult.

Note 2: A difficult point is to obtain the same performances on the test dataset in the step 1 and the performances in the step 3 due to many aspects such as for example non stationary environment.

Note 3: In step 1 one needs customers which are present in “O1” and remain present until the beginning of “P1”.

2.3. Data sources

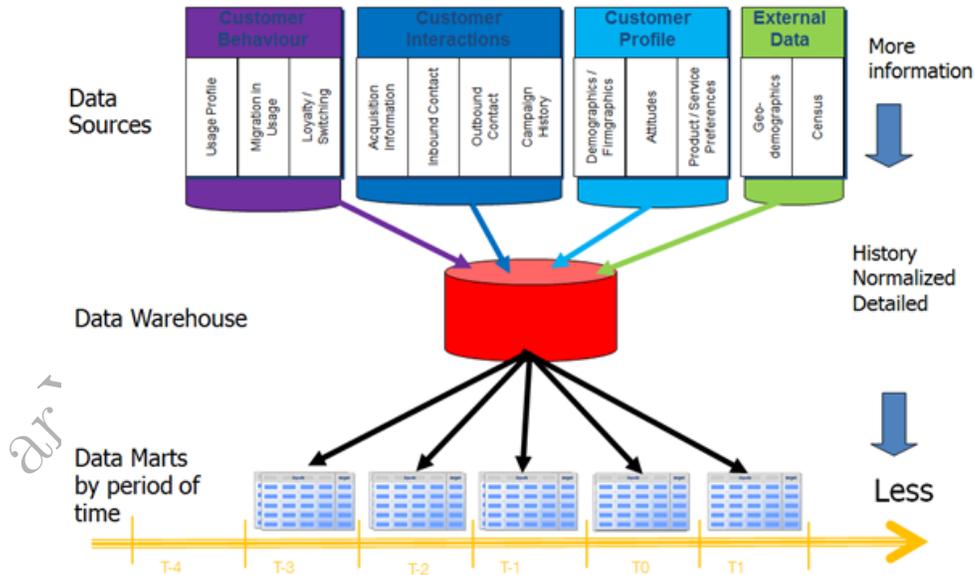


Figure 2: Data sources and data mart by period of time

The score is obtained using the quantitative information and qualitative information available on the user from different sources such as socio-demo, behavior of purchase, preceding answers, ... This sources filled a data warehouse (the red module in Figure 2) which contains tables connected in relational scheme where for examples a “customer table” (the root table) is connected with others “secondary” tables (as Usages, Products, ...).

The root table is related to the marketing domain. For customer data analysis, this table contains all the fields directly connected to the customer, such as his name or address. The secondary tables have a 1-N relationship with the primary table. Each instance of the primary table may be related to a variable number of instances of a secondary table. For telecommunication data for example, the secondary tables contains the list of services, of usages of these services, the call details.

The used of this different sources, different “kinds” of databases, requires the use of “model” able to deal a variety of difficulties: heterogeneous data (numerical and categorical variables), noisy data, unbalanced distributions of predictive variables (very few non-zero values) and target values (only very small percent of the examples belong to the positive class), sparse variables and many missing values.

2.4. Explanatory variables

A score¹ (the output of the model described in section 2.2) is an evaluation for the target variable, here the churn, to explain. The score is computed using customer records represented by a number of variables or features: the explanatory variables. Scores are then used by the information system (IS), for example, to personalize the customer relationship. The explanatory variables come from (at least) three origins:

- Native variables from the IS: these variables are directly present in the root table (see Section 2.3) and can be extracted for example by “Get” instruction
- Aggregates from experts: these aggregates are variables that are elaborated by a formula using the root table and secondary table(s), the formula is given by an expert of the domain who has the expert knowledge that this aggregate is useful for a given problem. For example, to build the number of usages of each service per weekday for all customers, one single language expression needs to be specified, with the use of the “Count” operator on the secondary table “Usage” with two operands “WeekDay(Date)” and “Label(ServiceId)”
- Aggregates automatically constructed: these aggregates are also variables that are elaborated by a formula but without expert knowledge. The idea

¹Technique of hierarchisation of the users which makes it possible to evaluate by a note the probability that a user answers a request or belongs to the required target within the framework of a direct marketing campaign

is not to elaborate randomly potential aggregates but to use a prior distribution over the space of all variables that can be constructed in the basis of a description of a multi-table schema and a set of construction rules. This allows building tens of thousands of features to create a rich data representation space. This point relies on domain knowledge to construct new potentially informative variables [2].

Whatever is the origin a particular explanatory variable, it has to be robust versus the time: i.e. this variable should be:

- useful when training the model (in Step 1 in Figure 1) and useful when using the model (in Step 3 in Figure 1)
- exhibits a good robustness when training the model and a good robustness when using the model (between Step 1 and Step 3 in Figure 1) Note: we call here robustness the performances ratio between the “training dataset” and the “test dataset” during the training process OR the performances between “test dataset” and the “back testing dataset” later. We are interesting in a ratio close to 1.

2.5. Automation and Volume

A good tool should be able to:

- extracts automatically a large number of features from a relational database (relies on domain knowledge to construct new potentially informative variables)
- selects a subset of informative variables and instances [4]
- builds the predictive model for datasets having a very large number of input variables (thousands) and instances (hundreds of thousands)
- allows a robust detection of the most predictive variables
- and efficiently builds an accurate classifier (the model).

A key requirement is the complete automation of the whole process.

2.6. Evaluation of the model quality

There is value in a CRM system to evaluate the “propensity” of customers to churn. Therefore, tools producing scores are more usable than tools producing only binary classification results. The tool is then asked to provide a score (a discriminant value or a posterior probability $(P(\text{Churn}=1 | X))$). The quality of the model is then judged by the AUC² (the area under the ROC curve), which is a scalar between 0.5 and 1.0 (0.5 corresponds to a random model and 1.0 corresponds to the ‘oracle’ model) and the lift curve (see below Figure 3).

² The AUC is a standard metric in classification see : http://en.wikipedia.org/wiki/Receiver_operating_characteristic

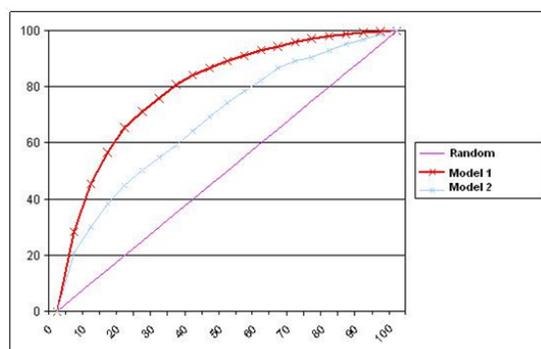


Figure 3: Lift curve: an example

The AUC is related to the lift curve. On the horizontal axis there is the percentage of the total population (of your marketing campaign) and on the vertical axis there is the percentage of the target population, for example the churn population. In the figure here we have two lift curves. The red curve has been obtained using the model 1 and the blue one has been obtained using the model 2. If we use the model 1 and we send an email to 20% of the total population, we will touch only 40% percent of the target. But if we use the model 2 we will touch 60% percent of the population. Therefore a good lift allows a good campaign.

2.7. Life cycle of a score

The 'life' of a score is constituted by 4 main step: the request, the modeling, the deployment and the supervision.

- Request Step:
 - Demand: A business unit wants the computation of a score on Customer
 - Instruction of the demand: the score team studies its advisability and feasibility
 - Validation and prioritization of the demand
 - Specification
- Modeling Step as every data mining process, we have
 - data acquisition
 - modeling (the training of the model)
 - validation
 - economic validation (do the results seem to be sufficient for a marketing campaign ?)
- Deployment Step:
 - cabling

- production
- validation
- Supervision Step

2.8. Key points in B2B - Individual interpretation

2.8.1. Common point with B2C

In churn prediction in B2B and B2C, from the different points listed above in this section we may write a list of requirements:

- Data constraints
 - Heterogeneous
 - Missing values
 - Multiple classes
 - Heavily unbalanced distributions
- Training requirements
 - Fast data preparation and modeling
 - Variable construction
- Model requirements
 - Reliable
 - Accurate
 - Parsimonious (few variables)
 - Robust
 - Understandable
- Deployment requirement
 - Fast deployment
 - Close to real time classification

To achieve this points the Khiops³ software was used. Khiops is a data preparation and modeling tool for supervised and unsupervised learning. It exploits non parametric models to the predictive importance of input variables in the supervised case. These evaluations are performed by the mean of discretization models in the numerical case and value grouping models in the categorical case, which correspond to the search for an efficient data representation owing to variable recoding. The tool also produces a scoring model for supervised learning tasks, according to a naive Bayes approach, with variable selection and model averaging.

The tool is designed for the management of large datasets, with hundreds of thousands of instances and tens of thousands of variables, and was successfully evaluated in international data mining challenges. This tool contains also

³Available on www.khiops.com

a framework aiming at automating variable construction for supervised classification [2].

In B2B as for B2C the purpose is to improve customer experience: leverage each interaction whatever the channel is, to identify the customer and provide him a differentiated treatment. A key point is therefore the interpretation of the results obtained with the scoring model. The level of “interpretation” could not be the same in B2B and B2C as described in the next section.

2.8.2. Key point in B2B

The scores are not always directly usable. For example if a scoring model identifies a customer interested in churning, the score does not say anything on the action needed to avoid his cancellation. To prevent him from churning, his reasons have to be identified.

The reasons of the churn could be given at three levels of granularity:

- Global: the reasons of the churn are given in average for the entire population
- Local: the reasons of the churn are given in average for a segment of the population
- Individually: the reasons of the churn are given customer per customer.

In B2B because of the volume (i.e. the number of customers concerned by the future retention campaign to launch) the reasons of the churn are mainly described at the global level. This is done often by simply giving the importance of the explanatory variables placed at the input of the scoring model for the global population. The objective of the measure of the importance of the input variables is four-fold: improve the prediction performance of the predictors, provide faster and more cost-effective predictors, allow a better understanding of the underlying process that generated the data and simplify the understanding of the results delivered by the model. For the Khiops software this is done by giving the weights of the variables. These weights allow to say that in average a given variable is the most important to predict the churn.

In B2C since the volume is lowest it is possible to go down from one or two levels. For the individual level one will be interested by: a methodology which would, for every customer, (i) identify the importance of the explanatory variables; (ii) identify the position of the values of these explanatory variables; and (iii) propose an action in order to change the customer response. The first point should answer to “Why THIS customer is predicted to go to churn” and the third one “How is it possible to keep THIS customer”. The reader interested by this methodology may read [10].

The Khiops software has therefore a complementary module, the add-on “Khiops Interpretation” that allows to obtain this couple of information. Understanding the score given by a model for a particular instance can for example lead to an immediate decision in a Customer Relationship Management (CRM) system.

3. Use Case: Churn prediction in B2B

3.1. Preliminaries and context

3.1.1. Definitions used

In this section, and before modelling churn and using customers' value analytics, we fixed some preliminaries:

- A customer is: a person using a card and identified by its MSISDN (Mobile Station Integrated Services Digital Network) commonly the GSM phone number. We will focus on his mobile consumption, for problem simplification. We distinguish between his in net and off net calls, as well as his international calls. All usages will be included: voice, SMS and data. As we are studying the B2B market, we will include all segments, i.e. low SoHo (Small office, Home office) segment, mid SME (Small & Medium Enterprise) segment, as well as high LA (Large Account) segment.
- A churner is someone (end-user) who: has disconnected from one month to another. This is the TARGET VARIABLE. A pre-paid churner is a user who did not refill in the previous two months. A post-paid churner is the one who stops the contract. We excluded bad debts as they do not have the same behavior as potential churners.
- A churner is someone (account) who: has a decrease of at least 25% of the total number of MSISDNs compared to previous month.

Our aim is to identify the end-users presenting the risk to leave the company, in order to retain them. What should be underlined is that in B2C churn modelling concerns the end-user. However, in B2B, we need to link the end-user to the account that identifies the company of the end-user.

Another declination of the churn modelling concerns pre-paid offers, post-paid offers and hybrid offers. In order to simplify the problem, we did not make difference for churn forecasting between the type of the offer. One of the reasons is that pure post-paid and post-paid part of hybrid offers correspond to B2B usage and pre-paid is a more B2C usage.

3.1.2. Explanatory Analysis of the used data

- What to fetch: The objective is to identify correlations between different factors (past activity, business demographics, etc.) and future activity, i.e. to estimate a model of customer activity.

The modeling will focus on customers' churn probability. It is equally critical to prioritize anti-churn actions based on customer revenues:

- Statistical analysis and forecasts
- Draw insights from the model to support next action to take
- Document model and model findings
- Produce scoring for each customer

- How to fetch: The data collected had to be analyzed, cleaned and modeled.

Initial received data summarized 500 information per MSISDN (e.g. voice, SMS, sales representative in charge of the account etc.). We followed a methodology based on a three steps close to the classical data mining process (see Figure 4).

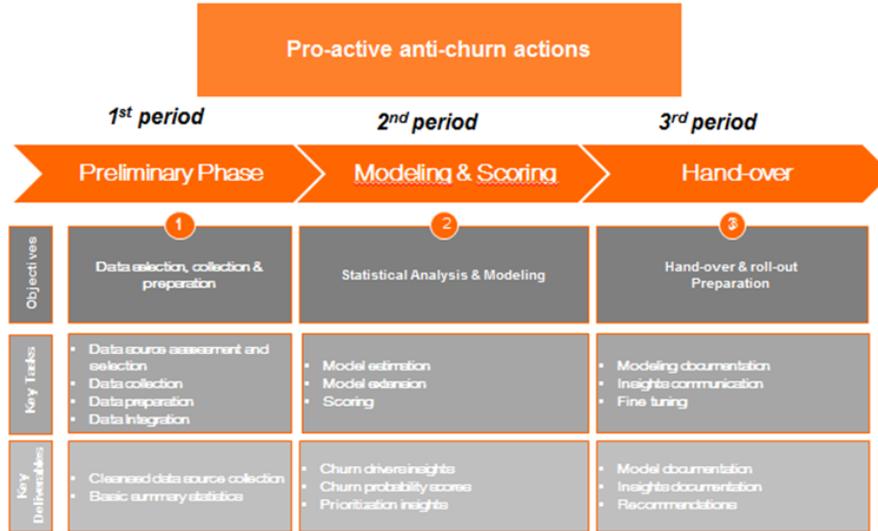


Figure 4: Methodology

In parallel of model building, we tried to get some insights from data. We analyzed different variables and tried to find correlations with churn. In the figure 5, we fetched for any relationship between churn and tenure, Voice, SMS or data usage As it can be seen, no obvious relationship was found.

3.1.3. Amount of explanatory variables used

During building models, we applied backward elimination variable approach [1] reducing the number of variables in iterative steps according to models efficiency. If the model performance was similar, then we would drop the variables as explained:

- Iteration 1: lagged variables out of model building history window scope.
- Iteration 2: all remained variables taken for modelling, and then we decrease the number of variables until getting to an optimal number. We removed variables with (1) small weights⁴ calculated by Khiops, (2) low levels calculated by Khiops.

⁴To simplify: the weights are the importance of the variables when they are together at the input of the scoring model and the levels are univariate importance of each variable when using only one of them, see [1] for more details.

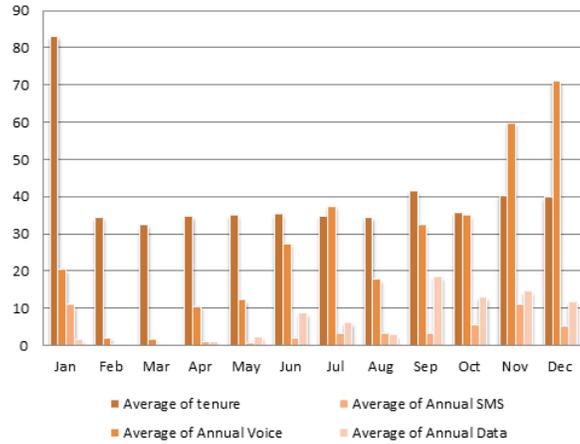


Figure 5: Example of an insight

- Iteration 3: intermediate variables : dropped if accuracy not proved on models built on it.

3.2. Experimental results: Models and scores delivery

The model aims to predict churners (according to the definition) at end-user level and not at account level. The principle of the approach is to classify the identifiers of the end-users, by decreasing probability to churn. We hope by that, to find in the first sorted end-users the highest potential churners in order to focus retention actions toward them. This is justified by operational reasons, as we can not afford to contact all customer data base each month to retain them.

We found after different iterations with the “classical” data mining process [3] that the optimal number of explanatory variables (see Section 3.1.2, with accurate models is obtained by four months of historical data (O1=2 months in the Figure 1), with a two months horizon forecasting (P1=2 months in the Figure 1) and no latency period (L1=0 in the Figure 1).

The variables retained belong to the following classes:

- Active Mobile: MSISDN, internal identifier, Code of the account, offer, activation date, last status date, segment, section, activation type
- No active Mobile: MSISDN, internal identifier, Code of the account, churn type, segment, offer, activation date, deactivation date, reason
- Trouble Tickets: internal identifier, section, ticket number, service domain, status
- Yearly Customer Revenue: internal identifier, section, data revenue, fixed revenue, mobile revenue
- Usage: MSISDN, destination, service, charge

- Mobile characteristics: MSISDN, internal identifier, Code of the account, Sales Representative Code, name, account manager, handset type, handset subtype, handset date
- Total bill value: for each MSISDN

The lift curve results of the models obtained was one of the result criterion used to qualify the quality of our models. In this study the model obtained an AUC sufficient to be judge at "deployable".

Once the probability to churn is calculated at the end-user level, we proposed to the marketing team different ways of operational converting the results in retention campaigns:

1. taking the first A end-users and sorting them by most frequent companies to which they belong.
2. taking the first B end-users and sorting them by the highest revenue companies to which they belong
3. for all the database, calculate for each company the average probability of churning of all end-users of the company.

where the value of A , B are parameters to be discussed with the marketing team.

The model built for predicting churners detected 63% of the real churners in the top 100 of the potential churners and 37% in the top 2000, as shown in the Figure 6

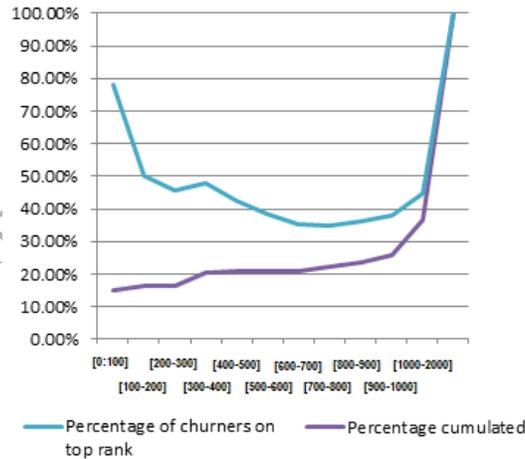


Figure 6: Percentage of churners in the ranked list

Moreover, thanks to Khiops interpretation, we could in addition to the probability of churning, get the most important variables to act on in order to decrease the churn probability. This allows giving for each end-user an operational lever to act on.

In the table 7, the Var_1 to Var_4 are the reinforcement variables on which operational marketing can build the retention speech.

msisdn	LaCode	NID	LaName	Section	AccountMa nager	ARPU previous 3 months (Aug- Oct14) all services	Total Revenue LACode previous 3 months (Aug-Oct14) all services	Total Revenue LACode previous 12 months (Nov13- Oct14) all services	probab to churn	Var_1	Var_2	Var_3	Var_4
	4.5715	200010914				4.65	714.03	4,036.37	87.29%	tenure	District_En	NotTtdosed	NID_yearly_Fixed
	4.137	915000				12.89	430.76	4,429.77	85.44%	tenure	District_En	NotTtdosed	Section
	3.128	200027946				17.86	1,322.04	4,696.25	82.10%	tenure	District_En	NotTtdosed	Sales_repCode
	4.2176	200010625				0.00	470.70	2,989.99	80.74%	tenure	District_En	NotTtdosed	Segment
	4.7914	2000105688				1.50	39.31	48.01	80.62%	tenure	District_En	NotTtdosed	NID_yearly_Fixed
	4.7914	2000105688				0.67	39.31	48.01	80.62%	tenure	District_En	NotTtdosed	NID_yearly_Fixed
	4.7965	200142175			NULL		3.67	3.67	78.97%	tenure	District_En	NotTtdosed	NID_yearly_Fixed
	4.7987	200142175			NULL		199.37	199.37	78.95%	tenure	District_En	NotTtdosed	NID_yearly_Fixed
	3.229	200000122				21.56	366.64	3,157.21	78.77%	tenure	District_En	NotTtdosed	Section
	3.229	200000122				0.87	366.64	3,157.21	78.77%	tenure	District_En	NotTtdosed	Section
	4.7988	200007084			NULL		411.76	411.76	72.54%	tenure	District_En	NotTtdosed	NID_yearly_Mobil

Figure 7: "How" indication [10]

Using the scores and the "why" and "how" indications given respectively by Khiops and Khiops Interpretation the first proactive retention campaign targeted 1007 accounts, divided as the following: (i) Account Level - 27 accounts: Accounts with over 60% of the account identified as risky, (ii) User Level - 980 accounts: Accounts having 59% or less of the account's MSISDNs identified as risky.

The results were: (i) At MSISDN level, the churn dropped at 6% vs 34% (control group), (ii) At account level, the churn dropped at 4.5% vs 80% in the control group.

Note: To maintain the accuracy of the model at a convenient level, a periodic refresh of the model is required. It can be done each semester or after any important evolution in the offer or the market. As about recalculating the predictive scores, input variables refresh is a requirement. The predictive scores should be calculated at least every month (if not, bi-monthly). The model is a compilation of different sources as shown in the Figure 2. Refreshing the model is key for efficiency.

4. Conclusion

Following the CRISP-like methodology (CRISP stands for Cross Industry Standard Process for Data Mining) we used a scoring approach to improve loyalty in B2B. Different scenarios were tried to optimize the lift curve with as less as possible explanatory variables. We built different models playing with the historical window used for train, and the window used for test (using accuracy and AUC tests to choose the best model). We then deployed the model, in order to provide operational marketing team with high probability churners. The results obtained, according that models have to be refreshed regularly, are very interesting in terms of accuracy and efficiency of the marketing campaign. More with help of variable importance and interpretation on the scores obtained "customer per customer", we obtained not only a ranking from a black box but actionable actions.

References

- [1] Boullé, M. (2007). Compression-based averaging of selective naive Bayes classifiers. *Journal of Machine Learning Research*, 8, 1659–1685.
- [2] Boullé, M. (2014). Towards automatic feature construction for supervised classification. In: *ECML/PKDD 2014*, pp. 181–196.
- [3] Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Advances in knowledge discovery and data mining. chap. From Data Mining to Knowledge Discovery: An Overview, pp. 1–34, American Association for Artificial Intelligence.
- [4] Féraud, R., Boullé, M., Clérot, F., Fessant, F., & Lemaire, V. (2010). The orange customer analysis platform. In: Perner, P. (ed.), *Industrial Conference on Data Mining*, vol. 6171 of *Lecture Notes in Computer Science*, pp. 584–594, Springer.
- [5] Gupta, Sunil, D. R. L. & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.
- [6] Keith, R. J. (1960). The marketing revolution. *Journal of Marketing*, 24(1).
- [7] Kirca, A. e. a. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of Marketing*, 69(2), 24–41.
- [8] Kohli, A. K. & Jaworski, B. J. (1960). Market orientation: The construct, research propositions, and managerial implications. *Journal of Marketing*, 54(2), 1–18.
- [9] Kumar, V. & Shah, D. (2004). Building & sustaining profitable customer loyalty for the 21st century. *Journal of Retailing*, 80(4), 317–330.
- [10] Lemaire, V., Hue, C., & Bernier, O. (2009). Correlation explorations in a classification model. In: *Workshop Data Mining Case Studies and Practice Prize KDD 2009*.
- [11] Meyer-Waarden, L. (2006). The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing*.
- [12] Narver, J. C. & Slater, S. F. (1990). The effect of a market orientation on business profitability. *Journal of Marketing*, 54(4), 20–34.
- [13] Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing*, 63(4), 33–45.
- [14] Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 67(4), 30–45.