

The Impact of Experimental Setup on Prepaid Churn Modeling: Data, Population and Outcome Definition

Dejan Radosavljevik¹, Peter van der Putten¹, Kim Kylesbech Larsen²

¹Leiden Institute of Advanced Computer Science,
Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

²International Network Economics, Deutsche Telecom AG,
Landgrabenweg 151, D-53227 Bonn, Germany
{dradosav, putten}@liacs.nl
kim.larsen@t-mobile.nl

Abstract. Prepaid customers in mobile telecommunications are not bound by a contract and can therefore change operators ('churn') at their convenience and without notification. This makes the task of predicting churn both challenging and financially rewarding. This paper presents an explorative, real world study of prepaid churn modeling by varying on three dimensions: data, outcome definition and population sample. Firstly, we add Customer Experience Management (CEM) data to data traditionally used for prepaid churn prediction. Secondly, we vary the outcome definition of prepaid churn. Thirdly, we alter the sample of the customers included, based on their inactivity period at the time of recording. While adding CEM parameters and varying the sample did not influence the predictability of churn, a particular change in the outcome definition had a substantial influence.

Keywords: Prepaid Churn, Data Mining, Customer Experience Management.

1 Introduction

The problem of churn, or loss of a client to a competitor, is a problem facing almost every company in any given industry. This phenomenon is a major source of financial loss, because it is generally much more expensive to attract new customers than it is to retain and sell to existing ones. Therefore, churn is important to manage, especially in industries characterized by strong competition and saturated markets, such as the mobile telecom industry. Prepaid mobile phone customers are not bound by a contract; therefore, they can churn at their convenience and without notification, which makes the task of predicting the likelihood and moment of churn very important.

The first step in managing churn is identifying the customers with high propensity to churn. There are not many published papers available on churn modeling, due to the commercially sensitive nature of the problem. Most of the available research relates to postpaid churn [1, 2, 3, 4]. Even fewer studies are available for prepaid churn [5, 6, 7]. The majority of these assume a fixed, single experimental setup in terms of outcome definition, population and data parameters. Their focus is mostly on

the data mining algorithm used or algorithmic tuning. In order to understand prepaid churn modeling better, we decided to take an end-to-end view and test different experimental setups by varying on three dimensions: data, population sample and outcome definition. We used standard data mining algorithms, decision trees and logistic regression [8], because it was not our objective to determine the impact of the algorithm; research on data mining algorithms is abundant.

Making new types of data available for modeling may improve model performance. We provide an overall framework for measuring customer experience, and in the experiments we investigate the added value of customer level service quality metrics (dropped calls, call setup duration, SMS delivery failure rate etc).

The definition of a population sample and outcome is primarily driven by the business objectives, i.e. how the model will be used. In general, operators consider customers to have churned if they have not had an activity (e.g. made a call, sent an SMS, etc.) for a period of six months. Usually, the customer has churned long before that time. From a business perspective, it is much better to use a shorter period of inactivity as an outcome definition of churn. Our variations in population sample are based on the customer's inactivity period at the moment of recording. The dilemma here is, if customers have been inactive already for one month at the time of recording, are they retainable, or have they simply thrown away the SIM card? What should be the maximum period of inactivity allowed at the time of recording?

The main contributions of this paper are the following. First, we provide a novel framework for measuring customer experience. Second, to our knowledge we are the first to investigate the added value of service quality centric customer experience data for predicting prepaid churn. Third, we explore the impact of changing the criterion for the population sample. Fourth, we measure the impact of the change in outcome definition of churn on the model performance. Finally, this research was conducted by one of the leading mobile telecom operators in Europe, driven by high priority business needs and using large amounts of customer data, which gives it a real world dimension. Given the highly competitive and strategic nature of wireless communications churn management, real world results are only scarcely available in research literature.

The remainder of this paper is structured as follows: In Section 2 and 3 we present the basic concepts of prepaid churn and customer experience management, respectively. In Section 4, the research setup and the experimental design are discussed, followed by results in Section 5. We end the paper with a discussion (Section 6) and a conclusion (Section 7).

2 Prepaid Churn Modeling

Prepaid churn modeling is typically done by performing analysis on aggregated Call Detail Records, but additional factors, such as subscription details (e.g. duration, subscription or discontinuation of services) and handset data, are often included.

For example, Archaux et al. [5] take into account the following data: invoicing data, such as the amounts refilled by clients or amounts withdrawn by companies for the subscribed services and options; data relating to usage, such as the total number of

calls, the share of local, national or international calls, consumption peaks and averages; data relating to the contract, such as date of beginning (age of the contract), the current tariff plan and the number of different plans the client used; data relating to subscriptions and cancellation of services; data related to the current profitability and the previous profitability of the clients. Alberts [6] selects data only related to usage and billing information: average duration of a single incoming and outgoing call, ratio between outgoing and incoming call durations, sum of incoming and outgoing revenues, the current and the maximum number of successive non-usage months, cumulative number of recharges, average duration of incoming and outgoing calls over all past months, the number of months since the last voicemail call, the number of months since the last recharge, the last recharge amount, the average recharge amount of all past months, etc. Both papers have the algorithm in their focus: Archaux [5] compares the performance of models built using neural networks and support vector machines, while Alberts [6] compares models built using survival analysis with models built using decision trees.

Additional factors, such as the influence of Social Networks, have also been addressed in previous research. Dasgupta et al. [7] strive to answer the question whether the decision of a subscriber to churn is dependent on the existing members of the community with whom the subscriber has a relationship. They show that diffusion models built on call graphs (used for identification of Social Networks) have superior performance to their baseline model. However, it is our opinion that these results are biased by the fact that their baseline model is constructed on CDR data alone, without including other important factors, such as the handset, prepaid account balance or inactivity period.

3 Customer Experience Management

To put our churn management activities in perspective, we see them as part of a wider ranging effort to manage and optimize the customer experience. Products and services (e.g. telecommunication services) tend to become commodities over time; therefore, managing the customers' experience can be a source of sustainable competitive advantage [9]. Pine and Gilmore [9] state that experiences are as distinct from services as services are from goods, thus arguing that experiences are the next evolutionary step in creating customer value. Smith and Wheeler claim that branded customer experience drives customer loyalty and profitability [10]. Customer Experience Management (CEM) is the process of strategically managing and optimizing these experiences across all customer touch points and channels, in interactions that are either customer initiated or company driven, direct or intermediated [11,12]. CEM is closely related to its predecessor Customer Relationship Management (CRM). The distinction is somewhat artificial but one could argue that CRM focuses more on providing a 360 degrees view of the current relationship whereas CEM provides tools and methodologies to improve and extend the relationship by optimizing the overall customer experience. Churn management can be a key part of this. Churn models, in particular, can be important components of

overall strategies that make these predictions actionable to drive intelligent interactions and experiences.

3.1 Measuring the Customer Experience

The focus of this paper is on building better prepaid churn models, both from a predictive power perspective, as well as from the point of view of predicting behavior that is actionable and makes business sense. For instance, predicting short term inactivity for a base that includes many already dormant accounts will result in a powerful, yet not very useful, model. Embedding the churn models in larger strategies to optimize experience in real time is more the scope of future research. From a CEM perspective, our topic of interest is measuring past customer experience as a potential input to churn models.

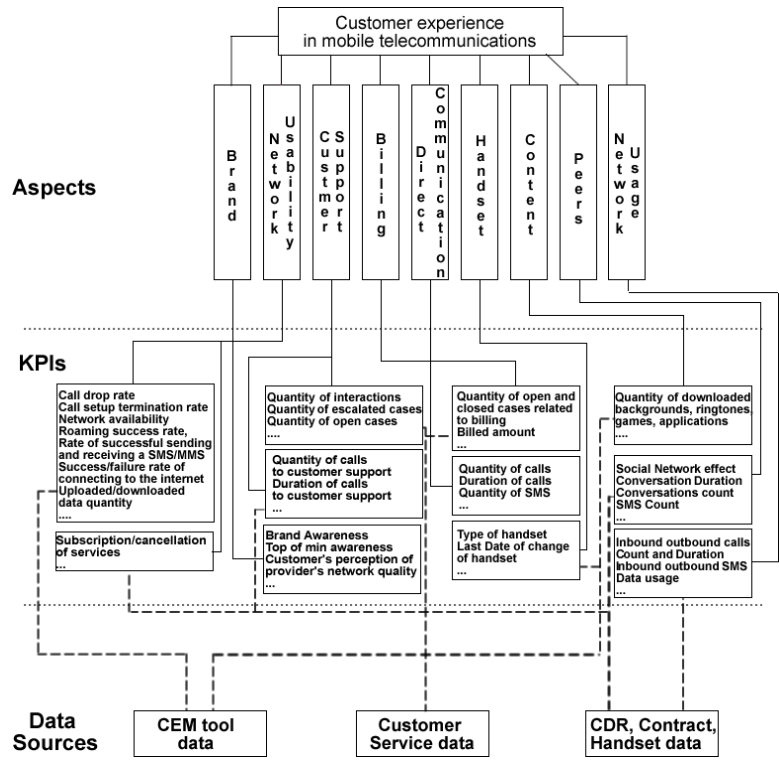


Fig. 1. Customer Experience Framework for Mobile Telecommunications

We have created a novel framework for measuring customer experience in mobile telecommunications in ideal circumstances (Figure 1). This framework provides a roadmap and direction for the company for the coming years. In the short term, it is very challenging to measure most of these dimensions reliably. The framework

contains three levels: Aspects, KPIs and Data Sources. The Aspects level represents the various aspects of customer experience in mobile telecommunications. The KPIs level defines possible measurements for the different aspects of customer experience. The Data Source level describes the potential location of various KPIs.

We are interested in the relative value of these CEM KPIs for general marketing use. Specifically for prepaid churn prediction, the following six aspects of the CEM framework presented above can represent this added value, as they have not been used before for prepaid churn prediction: Brand, Network Usability, Customer Support, Direct Communication and Content. Network usage, Billing and Handset have been addressed by [5,6], while Peers (Social Networks) is the topic of [7]. For this paper we want to particularly focus on the added value of the CEM tool data source. This tool measures service quality (dropped calls etc.) for operational purposes, but claims are made that this data is also of key importance for prepaid churn management and we want to validate whether these claims are warranted from a business case perspective, assuming that traditional usage and customer relationship data is already available.

4 Research Setup

In order to examine the influence of the CEM metrics, variations in population sample and outcome definitions, we constructed three separate experiments which are described in the subsections below.

A set of general rules applies to all three. The predictors were recorded at the end of March 2009 with one, three and six months of history. The random sample of around 10% prepaid customer base was divided into training, validation and test sets, using the 50:25:25 ratio. The validation set was purely used as an additional test set. It was not used for developing the models in any automated way (e.g. model parameter setting or pruning using automated cross validation). The next section will provide some more technical details on the end to end data mining process followed.

4.1 End to End Data Mining Process

The commercial data mining tool Chordiant Predictive Analytics Director [13] was used for automated data preparation (attribute binning, grouping, recoding, selection) and modeling (logistic regression, decision trees).

The objective of data analysis is to transform the various attributes, a.k.a. variables (columns in the flat table) of the research sample, and identify the most informative ones among them, at this stage, from a univariate perspective. A variable is considered informative if it has a certain level of influence over the outcome variable (which in this case is churn). Both statistical and graphical tools are used to establish this degree of influence. In this research, the chosen statistical criterion for evaluating the performance of the variables (both for a particular variable, and a set of variables contained in a model) with relation to churn is the Coefficient of Concordance (henceforth CoC), which is a rank correlation measure, related to Kendall's tau [14]. Histograms are the chosen graphic tool for data analysis. The data analysis process

begins by discretizing continuous variables into a large number of bins. Numeric bins and symbolic values without statistically significant difference in churn rate are then grouped together.

After the initial data analysis, for the purpose of predictor selection, the software groups the variables that are correlated independent of the target. A given predictor may have a high univariate performance, but also be correlated with other candidate predictors that are even stronger, hence not adding value to a model. The user has three options when selecting predictors to be used for modeling: all predictors, only the best of each group, or manual selection. In our case we used the best predictors of each group as a starting point, and then experimented with further manual selection, for instance by removing predictors from the bottom scoring groups altogether, or changing the parameters of the grouping process to force more or less groups. The full dataset consisted of more than 700 attributes, and the optimal number of predictors for final models typically ranges between 5 and 10 predictors.

For logistic regression models the raw predictor values are replaced with a normalized rank score concordant with the churn rate for the group (for instance for a particular record and the predictor 'age', the churn rate associated with age 22-26). Logistic regression models are then built on the recoded data, to allow capturing of complex non-linear behavior, whilst using a stable low-variance modeler. Decision trees will only be able to split between groups. The environment also offers capabilities for building models through Genetic Programming but for the purpose of this paper we restricted ourselves to regression and decision trees as more generally known techniques.

The modeling process results in scoring models, where the instances are sorted by the probability of being a churner. Models were compared using gain charts. On the y-axis, these charts show the captured proportion of the desired class (i.e. churners) within a percentage of the population of increasing size (x-axis), ranked by the model (see Figure 3).

4.2 Definition of Prepaid Churn and Initial Sample

The working definition of activity within the company where this research was conducted is used as the *operational definition of activity*. Any initiated calls, top-up (recharge) of the account, sent SMS/MMS messages and received and answered calls are considered an activity. Unanswered received calls, received SMS/MMS messages and bonus credit top ups awarded by the company are NOT considered an activity.

Involuntary prepaid churn in this company occurs after six consecutive months of no activity. After this, the customer can no longer use the services of the company via that particular SIM card, and the phone number may be reissued to another client after a certain period. However, six months of no activity is too long of a period to investigate voluntary prepaid churn. Most of the internal projects investigating prepaid churn within this company consider customers to have churned if they had not been active between two and three consecutive months. Therefore, we propose the following *operational prepaid churn definition*: customers are marked to have churned if they are registered to have two consecutive months of no activity, or more.

It is also necessary to set certain boundaries for the population taken into account for the modeling process. Our population consisted of prepaid subscribers that met the following two conditions: 1) they had at least one registered activity in the last 15 days, and 2) their first activity date was at least four months before recording. The first condition enables avoiding subscribers who can already be classified as churners using the definition above. Arguably, to avoid this group, it would be sufficient to set the limit in condition 1 to 59 days, but this would make the prediction task trivial. The boundary was set to 15 days as a balance between excessive reduction of the sample and limitation of the information spillover (the subscribers inactive for 30 days have already begun to display churn behavior). The second condition is useful for avoiding frequent churners (it implies loyalty) and for avoiding bias when measuring communication with previous churners.

In summary, for the particular data set constructed, all predictors were measured in the first trimester of 2009; churn was measured two months later; inactivity at recording was limited to maximum 15 days and first activity date had to be minimum four months prior to recording. This served as baseline data set.

4.3 Experiment A: Addition of CEM Parameters

Capturing the CEM metrics was performed by using a CEM tool deployed in the company. This tool suffered from two limitations. First, it contained only 40 days of user history. In order to give the CEM parameters a fair competing chance, we only used traditional parameters with one month history. Second, this tool did not cover the entire customer experience as depicted on the CEM Framework on Figure 1. It is more of a network experience measurement tool due to the fact that it is focused on the more hygienic aspects of customer experience (aspects noticed by customers only if they are absent or have low quality). This database contains only the Network Usability aspect (e.g. Call Success Rate, SMS Success Rate, Call Setup Duration, etc.) and the Content aspect (e.g. count of accessing company's website for downloading ringtones, backgrounds, etc.) of our CEM framework. However, this tool did not capture any data related to Customer Support; therefore, Customer Support data was added from the CDRs. Other aspects of the CEM framework, such as Network Usage, Handset and Peers, which were recorded from the company's CDRs and customer "contract" data, were used for benchmarking purposes (to test the added value of parameters that are viewed exclusively as CEM KPIs). Therefore, they cannot be treated as potential contributions of CEM to the model's performance, but rather as traditional parameters as described in Section 2. The two remaining aspects, Brand and Direct Communication, could not have been analyzed because the company did not have data appropriate for data mining (Brand), or did not communicate proactively to its prepaid customers. Therefore, only three aspects were left for analysis: Network Usability, Content and Customer Support. Hence, the only added value stemming from using CEM KPIs on prepaid churn modeling in this research can exclusively originate from one of these three aspects of the CEM framework. In order to determine the added value of CEM, we compare models consisting only of "traditional" parameters, to models consisting of both "traditional" and CEM parameters.

4.4 Experiment B: Variations in Population Sample

In order to test the impact of variations in the population sample, we changed the inactivity restriction from maximum 15 days of inactivity to maximum 30 days and maximum zero days. The reasons for varying the sample on maximum period of allowed inactivity are threefold. First, our intention was to inspect the impact of these changes on the models' performance. Second, we wanted to test the contribution of the CEM parameters under different circumstances. Third, the typical period of users' inactivity, before they become unavailable for contact and retention, is disputable. Additional motivation for experiment B can also be found in the results of experiment A.

In addition, using zero days of inactivity can be seen as a way to avoid information spillover. We defined churn as uninterrupted inactivity for two months. It is clear that the best predictor of this is if a user has already been inactive for a certain period. In other words, users have already started to display churn behavior. This spillover cannot happen when the inactivity period at recording is limited to zero. Additionally, these subscribers are likely to still be available for contacting. We can expect that churn is harder to predict for this subgroup of subscribers.

4.5 Experiment C: Change in Outcome Definition

The motivation for changing the outcome definition was to investigate the impact of this change on the performance of the models, while keeping the same sample and dataset, and test the added value of the CEM parameters in this situation. For these reasons, we introduce a so-called "grace" period of 15 days. Customers are marked as churners if they are registered to have at least one activity in the first 15 days after recording, followed by two consecutive months of no activity, or more.

This outcome definition also resolves the information spillover issue discussed in Section 4.2. Namely, subscribers must have an activity in the first 15 days ("grace period") after recording, thus breaking out of their already displayed churn behavior. Subscribers inactive in this grace period, who continue to be inactive in the future, are labeled as non-churners, because they have already churned at the time of recording, and are of no interest. In addition, subscribers recognized as future churners using this definition are more likely to be available for contacting and retention, because we are aiming at subscribers who are first active for 15 days, and then inactive for two months or more.

5 Results

In this section we report the results of the experiments of adding CEM data, changing the population based on inactivity duration and changing the outcome definition, or experiments A, B and C, respectively (see Table 1 for highlights). As stated above, the main tool used for comparing model performance is the gain chart. It is worth mentioning that performance-wise, two models can be compared, only if they have

been created under the same experimental setup (e.g. Models A_Excl_CEM and A_Incl_CEM can be compared, Models A_Excl_CEM and C_Excl_CEM cannot).

Table 1. Sample size and churn rate in experiments A, B1a, B1b and C

Experiment	Inactivity Allowed	Sample size	Churn Rate (%)	Churners % in Top 20% Population Scores
A	15 days	62565	3.88	78
B1a	30 days	67986	6.09	78
B1b	0 days	32104	0.70	80.5
C (Definition change)	15 days	62565	2.40	50

The gain chart of two models created for experiment A is presented on Figure 3c. Model A_Excl_CEM contains only traditional parameters and is built on six predictors, containing only one month history. Adding parameters with longer history did not change the performance of the model substantially. Model A_Incl_CEM contains the same six variables, plus two CEM parameters, which were selected using a trial and error process based on their respective CoC, the predictor group they belonged to and the improvement in performance they brought when added to the baseline model. Both models were built using decision trees, but models built on the same variables using logistic regression performed just as well. There was barely any visible difference between the performances of the models. Their gain charts are identical up to 50% of the population. Both models were able to identify 78% of the churners within the top 20% of the population scores. The results were consistent throughout the training, validation and testing sets. The strongest predictor for churn was the inactivity period, followed by the remaining credit on the user's prepaid account.

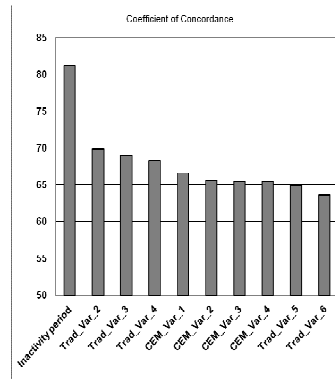


Fig. 2. Coefficient of Concordance of predictors grouped in group 1 for experiment A.

The inability of the CEM variables to substantially improve the base model is due to two reasons. First, a large number of these variables have a very low CoC (univariate performance). Second, even when the CoC is relatively high, in most groups there are traditional (Non-CEM) predictors with CoC values higher than the

CEM variables in the same group. Such is the case in the highest ranked group 1, which is depicted on figure 2.

Table 2. Grouping of variables of Model A_Incl_CEM

Predictors	CEM Framework Aspect	Group	Performance (CoC)
Inactivity Period	Network Usage	Group 1	81.14
Trad_var_x	Network Usage	Group 2	70.82
Trad_var_y	Network Usage	Group 2	70.63
Trad_var_z	Network Usage	Group 2	69.83
CEM_var_x	Network Usability	Group 2	68.87
Remaining credit	Billing	Group 3	67.39
CEM_var_y	Network Usability	Group 4	63.19
Handset type	Handset	Group 5	57.22

In order to illustrate the reasons for the weak effect on model performance improvement of the CEM variables, we isolated only the eight variables from model A_Incl_CEM, and ran the automatic grouping operation, with more strict grouping parameter settings than used previously, to force further grouping. One of the CEM parameters, CEM_var_x, has a reasonably high performance, but is in the same group as three other traditional variables, which means it has a degree of correlation with them, and has the lowest CoC in that group. This explains why CEM_var_x does not improve the model performance substantially. The second CEM variable in this model, CEM_var_y is in a class of its own, but it does not have a very high CoC; therefore, it cannot add substantial value to model performance.

Due to the strong influence of the inactivity period in experiment A, we decided to vary the population sample by changing the inactivity limit at recording to 30 and zero days (experiments B1a and B1b, respectively). Model B1a_Excl_CEM was built on seven non-CEM variables, using decision trees. B1a_Incl_CEM was built on the same seven variables plus two CEM parameters, selected based on the improvement they caused. Once again, there was no visible performance difference (see Figure 3a). The influence of the inactivity period was even stronger, once again followed by the remaining credit. The models' performance was stable over the three datasets. The gain chart on Figure 3a shows that both models identified about 78% of churners in the top 20% of the population. B1b_Excl_CEM is built on five non-CEM parameters. B1b_Incl_CEM is built on the same five variables, plus two CEM parameters. This time, both models were built using logistic regression. In this situation, models built on decision trees were less stable across the three datasets, and their performance on the test set was lower, even though the performance on the training set was similar. Nevertheless, even when using logistic regression, the model was not as stable as it had been in the previous cases. However, in this experiment, the model B1b_Excl_CEM identified 79% of the churners in the top 20% of the population, while B1b_Incl_CEM identified 1.5% more, which is shown on Figure 3b. However, a similar result can be achieved by adding two non-CEM parameters. In these models, the strongest predictors were the remaining credit and the phone type.

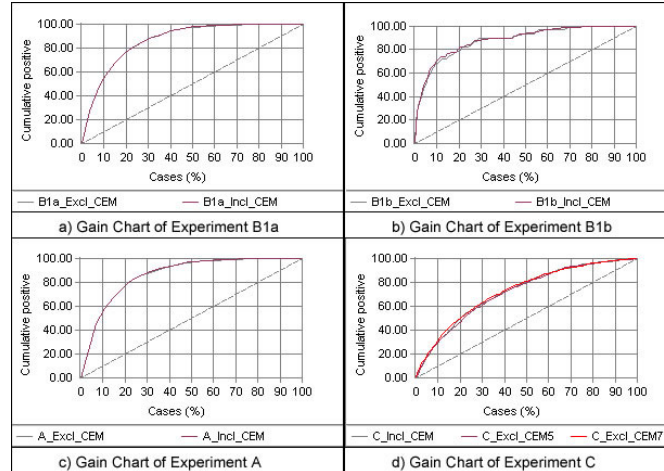


Fig. 3. Gain chart of models for experiment A, B and C.

Finally, we present results for experiment C. The altered churn definition included the requirement of activity in the first two weeks of the outcome period, followed by a period of no activity of two months. Model C_Excl_CEM5, presented on Figure 3d contained only traditional parameters and was built on five predictors. Model C_Incl_CEM contained the same five variables, plus two CEM parameters. Both models were built using logistic regression, because they had a better performance on the test set when compared to models built on decision trees. There is a difference between the performances of the models higher than those in experiments A and B. Model C_Incl_CEM identified 50% of the churners within the top 20% of the population scores, compared to the 46.5% identified by C_Excl_CEM5. However, when two non-CEM parameters were added (model C_Excl_CEM7), the same performance was achieved. Nevertheless, this is 28% identified churners less (in the top 20% of population scores) than what was achieved in experiments A, B1a and B1b.

The results of experiment C were less consistent throughout the three datasets when compared to experiments A and B1a, but more consistent when compared to B1b. It is interesting that the inactivity period was *not* a factor in these models, even though 15 days of inactivity were allowed at recording. The most powerful predictors were the remaining credit, the handset and the call count.

Similarly to experiment A, the inability of CEM variables to improve the model performance in both experiments B and C is due to either low CoC or correlation with stronger traditional predictors.

6 Discussion and Future Research

We conducted three experiments in order to compare the influence of new CEM parameters, as well as changes in sample population and outcome definition, on prepaid churn model's performance.

CEM is advertised to have an added value in predicting churn, but this was not the case in our experiments. Models without CEM parameters performed almost the same as models which included CEM parameters in all three experiments. However, we only tested the CEM parameters with "hygienic" nature, which are only noticed when absent or unsatisfactory. The softer aspects of CEM remain untested. However, at this point we expect that it will be hard to find new non behavioral predictors with sufficient predictive power compared to the behavioral data. Even though the CEM data we had available contained only 40 days of history, it is unlikely that longer history would change the outcome, because the non-CEM parameters used by the models in most cases also had only one month history.

The rationale behind the low added value of CEM parameters for prepaid churn modeling may be found in several factors. The prepaid customers themselves are the first factor. Prepaid customers that were subject to this research had average call duration of around one minute. A very low percentage of prepaid users used data services or have called the Customer Service in the research period. Next, prepaid users are mostly interested in controlling (reducing) their mobile phone expenses; otherwise, they could switch to using post-paid services that offer less expensive calling tariffs. The interest of prepaid users to control their mobile phone expenses may have been enhanced by the Global Financial Crisis of 2007/2008. To summarize, prepaid customers are more concerned with the quantity of experiences, which is measured by traditional predictors (Section 2) rather than the quality of their experiences, measured by the CEM parameters. The only exception is the handset which is a parameter that deeply influences the quality of the experiences, and is also regarded as a lifestyle product. The second factor that could explain the low added value of CEM parameters was the high network quality. The percentage of customers experiencing network problems (negative experiences) is very small. However, the network quality cannot be seriously tested on average call duration of one minute. The third explanation for the low added value of CEM parameters is that the quality parameters are correlated (to a degree) to their quantitative counterparts (e.g. number of dropped calls is correlated to a degree with number of calls).

Changing the population sample by varying the inactivity limit at the time of recording also did not contribute to a substantial change in model performance. Models in experiments A, B1a and B1b, that allowed 15, 30 and zero days of inactivity at recording, identified between 78% and 80.5% of churners in the top 20% of the population. However, sample size and especially the churn rate did change drastically, as presented in Table 1. Furthermore, the change of allowed inactivity period to zero also influenced the model stability. Once again it is important to emphasize that customers with zero days of inactivity at the time of deployment are more likely to be available for retention than customers with 15 or 30 days of inactivity.

The most dramatic change in model performance was shown when the outcome definition was changed and a so-called grace period of 15 days was included to mirror

the inactivity period allowed at the time of recording. The model performance dropped from 80% to 50% of identified churners in the top 20 % of customers. This does not mean that models built under experimental setup C are worse than the others. It merely implies the expected performance under such conditions. This steep decline is due to the complicated churn definition we deployed (we targeted an inactivity-activity-inactivity pattern). In this case, the inactivity period, that was a dominating variable in experiments A and B1a, was not a factor at all. The benefit of using such a definition is that upon deployment, the identified churners are likely to have an activity in the next 15 days, which makes them available for retention.

The focus of the research was on the impact of the experimental setup, rather than the algorithm used to create the model. Therefore, we used standard data mining algorithms, such as decision trees and logistic regression. Having said that, we would like to emphasize that in experiments A and B1a, there was barely any difference on model performance that could be attributed to the usage of the particular algorithms. In experiments B1b and C, there was a small difference in performance of algorithms, but it was not as substantial as changing the outcome definition. In these two experiments, logistic regression had a more stable performance on the training, validation and test sets, compared to decision trees.

In terms of directions for future research, we would like to investigate a richer set of customer experience data, particularly around proactive communications and brand. Additionally, it would be worthwhile to investigate the relations between duration of inactivity and availability of subscribers for retention, by inspecting their presence on the network (regardless of making calls). Last but not least, we would like to take into account the feedback from retention campaigns, in order to focus on “retainable churners.” After all, the end target of churn prediction is retention.

7 Conclusion

In this paper, we presented how performance of prepaid churn models changes when varying the conditions in three different dimensions: data- by adding CEM parameters; population sample- by limiting the inactivity period at the time of recording to 15, 30 and zero days, respectively; and outcome definition- by introducing a so-called grace period of 15 days after the time of recording, in which customers must make an activity in order to be classified as churners.

Adding the CEM parameters into the models did not add substantial value in model performance under any of these conditions. Similarly, switching the population sample on the period of inactivity at the time of recording did not influence model performance, only model stability and churn rate. On the contrary, a dramatic performance reduction of nearly 30% less identified churners within the top 20% of the population was noted as a consequence of changing the outcome definition by setting a grace period, thus making the behavior to be predicted more complex. This change obviously influenced the churn rate as well.

References

1. Hung, S., Yen, D. C., & Wang, H.: Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515-524 (2006)
2. Mozer, M., Wolniewicz, R., Johnson, E., & Kaushansky, H.: Churn reduction in the wireless industry. *Proceedings of the Neural Information Systems Conference* (1999)
3. Kim, H. & Yoon, C.: Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9-10), 751--765 (2004)
4. Ferreira, J., Vellasco, M., Pachecco, M., & Barbosa, C.: Data mining techniques on the evaluation of wireless churn. *ESANN' 2004 proceedings - European Symposium on Artificial Neural Networks*, pp. 483—488 (2004)
5. Archaux, C., Martin, A., & Khenchaf, A.: An SVM based churn detector in prepaid mobile telephony. *International Conference on Information & Communication Technologies (ICTTA)* pp. 19--23 (2004)
6. Alberts, L. J. S. M.: Churn Prediction in the Mobile Telecommunications Industry, MSc Thesis, Department of General Sciences, Maastricht University (2006)
7. Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., & Nanavati, A. A.: Social Ties and their Relevance to Churn in Mobile Telecom Networks. *Proceedings of the 11th international conference on Extending database technology* pp. 668--677 (2008)
8. Witten, I. H. & Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*. (2nd ed.) San Francisco: Morgan Kaufmann (2005)
9. Pine, B. J. I. & Gilmore, J. H.: *The Experience Economy*. Boston, MA: Harvard Business School Press (1999)
10. Smith, S. & Wheeler, J.: *Managing the Customer Experience: Turning Customers into Advocates*. Harlow, USA: FT Prentice Hall (2002)
11. Meyer, C. & Schwager, A.: Understanding Customer Experience. *Harvard Business Review*, 85, Vol. 2, pp. 116--126 (2007)
12. Schmitt, B. H.: *Customer experience management: A revolutionary approach to connecting with your customers*. Hoboken, NJ: John Wiley & Sons (2003)
13. Chordiant Software: Chordiant Predictive Analytics Director [Computer software]. Chordiant Software, now part of Pegasystems (www.pega.com) (2008)
14. Kendall, M.: A New Measure of Rank Correlation. *Biometrika* 30 (1-2): 81--89 (1938)