Revenue Optimization of Telecom Marketing Campaigns for Prepaid Customers

Maurus Riedweg

Consulteer AG, Switzerland
Email: maurus.riedweg@consulteer.com

Pavol Svaba and Gwendolin Wilke

Lucerne School of Information Technology, Switzerland
Email: {gwendolin.wilke, pavol.svaba}@hslu.ch

***Abstract*—The design and optimization of marketing campaigns today usually still includes a high level of manual expert involvement. This applies particularly to the prepaid mobile phone sector of the highly competitive telecommunication industry. Since prepaid telecom customers are characterized by highly volatile and sparse usage data their future behavior is hard to predict, and marketers often rely mainly on experience and gut feeling when designing marketing campaigns, using only simple data analysis tools. The project developed a methodology and software prototype that helps marketers in this area to exploit the full potential of real-time big data-driven analytics for microtargeting, allowing them to make fact-based and informed decisions. Specifically, it provides an interactive solution for the semi-automated visual support of the design and optimization of single-channel marketing campaigns. The developed solutions bring a huge step towards the automation of the whole process of optimizing marketing campaigns in the telecommunication business, keeping the possibility of interactive interventions of marketers to implement strategic management decisions or use their expert knowledge. The system provides enough information for marketer to comprehend the reasons for the decision and by retracing it provides precious insights for the design of new campaigns. The solution uses machine learning closed loop and intuitive visualization based on nomograms and was prototypically implemented on Apache Spark big data stack and evaluated on sample data from two real-world prepaid telecom use cases.**

Keywords*–*telecom; churn prediction; predictive analytics; Naïve Bayes; Nomogram

I. Introduction

Classical data mining for marketing campaigns is usually a time consuming task. Despite the availability and use of automated analytics algorithms, the overall process of designing, adapting, executing and evaluating a marketing campaign still includes a high level of expert involvement and manual decision-making. In the age of big data, the increased volume and velocity of available data allows for added business insights and helps companies to stay competitive. At the same time, it becomes increasingly difficult to fine-tune the design and management of analytics-based marketing campaigns manually. Therefore, to fully tap the potential of real-time big data marketing analytics, a high degree of automation is needed.

The success rate of a marketing campaign depends on targeting the right customer needs and preferences. Consequently, there is an increased need for highly targeted marketing campaigns that are tailored to the needs and preferences of specific, possibly small customer groups (micro-segmentation) or even individuals (direct marketing campaigns). This requires highly agile marketing campaign management facilitating fast reaction times to customer or market events despite continuously increasing data volumes. Additionally, structural adaptations that allow for reacting to changes in the market structure as a whole must be accounted for. There is a trend in marketing towards building long-lasting relationships with customers. Since the cost for customer acquisition is much greater than the cost of customer retention [1], marketers have become increasingly interested in the latter [2]. Customer retention models aim at furthering customer loyalty and preventing active customers from changing to other providers (churning) via marketing campaigns. To achieve this, tools that support the development, management and application of customer retention models and the corresponding marketing campaigns are required and are essential Business Intelligence (BI) applications. In telecom business, the customer churn term refers to the customer turnover, i.e., loss of a service subscriber. The common reasons for churn are dissatisfaction with an existing provider, the lure of a lower price for equivalent service or better service for the same price from a different provider. Churn rate, the proportion of churned clients during a given time period, is one of the key business metrics and an indicator for customer dissatisfaction.

Churn prediction modeling techniques attempt to understand the precise customer behaviors and attributes, which signal the risk of customer churn. In predictive analytics, the typical approach to data-driven churn prediction is to use a sufficiently large historical data set of customer records, containing churning and non-churning customers. The records contain customer attributes such as age, gender, tariff, average call frequency, etc., together with the information if a customer has churned at some point in time or not. The data set is used as training set for supervised learning to construct a classifier, i.e., a predictive model that classifies a customer as a potential churner or non-churner based on the knowledge of the other attribute values (the predictor variables). In an actual prediction task, the model is applied to a new data record, where the values of the customers predictor variables are observed while the possible churn event lies in the future, so that value of the class variable churn is unknown. The classification done by the model is used as a prediction of churn or non-churn of the customer segment that corresponds to the respective predictor attribute valuations (the feature vector). In case the system has predicted a customer segment of likely churners, a suitable marketing action can be launched to prevent these customers from churning. To select such an appropriate action the underlying reasons for their churning must be analyzable, i.e., a proper insight on the reasons for churning is essential in order to design effective retention methods.

Predictive analytics models traditionally used for data-driven customer churn prediction are Decision Trees [3]–[8] and Regression analysis [9] and [10], which have been com-

 ISBN: 978-1-61208-614-9

plemented with Naïve Bayes classifiers and Artificial Neural Networks [11]–[13] in recent years [14] and [15]. In Naïve Bayes classifier (NBC) models, the attributes of a domain are interpreted as random variables and are represented as vertices in a probabilistic graphical model. Direct influences of the observed predictor variables on an unobservable target variable are represented as directed edges in the graph. The result is a simple tree graph structure. The strength of influence of input variables on the target variable is given by conditional probability tables, which are learned from historic data. In a prediction task, Bayes Rule [16] is used to infer from values of the observed variables the probability of possible values of the unobserved target variable. NBCs are thus map based classifications with the additional assumption of conditional independence of input variables. Examples of applications using NBCs for churn prediction are Nath and Behara [17] and and Kirui et. al. [18] or Shaaban et. al. [13]. Other predictive models that have been used for churn prediction are, e.g., rule based classifiers (Ripper, PART), Nearest Neighbor (KNN), Self Organizing Maps [19], Genetic Algorithms [20], Linear Discriminant Analysis [21], Support Vector Machines [13], Sequential Pattern Mining and Market Basket Analysis [22] or Rough Sets [23].

As a reaction to the increasing data volumes in the telecommunication sector, approaches towards churn prediction on massive data sets have been investigated. For example, Kamalraj and Malathi [24] discuss the use of data mining techniques on big data clusters in the telecommunication industry. Balle et. al. [25] describe a prototype for churn prediction using stream mining methods, which offer the additional promise of detecting new patterns of churn in real-time streams of high-speed data, and adapting quickly to a changing reality. Different scalable algorithms have been developed and implemented, e.g., Apaches open source libraries Mahout and MLlib include, besides scalable regression models, decision trees and customer segmentation models, also Naïve Bayes Classifier learning.

The research presented in this paper was funded by the Swiss Commission for Technology and Innovation CTI and is part of the project done together with two expert partners, one software developer with extended history in telecommunication business, and one large telecom service provider located in Switzerland.

The paper is organized as follows. After short introduction we describe the problem and its requirements in Section II. In Section III, we address the conceptual model. Section IV brings the insights of development of the predictive analytics and visualization core based on data analysis done in statistical environment R. Section IV shortly describes the implemented big data solution. Section V summarizes the project results and Section VI concludes its achievements.

II. Problem Description

The mission of this project is to develop a methodology and a software prototype for semi-automated predictive marketing analytics and campaign management. It will automate the life cycle of the data mining and the campaign management process as a whole, including feature selection, model learning, prediction and decision making; iterative dynamic model adaptation over time will allow for closing the analytics loop. In particular, our goal was to optimize campaign targeting in prepaid segment of a large Swiss telecom company. We aim on customer retention campaigns in general, or on churn prevention in a special case. The existing system, and its data, puts some critical constraints we have to deal with. For example, the provided historical data are aggregated and contain time series from the narrow time slots before and after the campaign offer. Within a campaign, the existing system chooses a fixed number of customers each day and offers them a particular bonus. The offer is valid for couple of days, and expires afterwards. Some particular campaigns with a bonus credit have been chosen as a proof of concept. The campaigns run for a defined period of days, their results are reported on a daily basis, and summarized at the end of the campaign.

The designed solution must possess the following properties:

(1.) Big data-driven knowledge discovery allows for detection of hidden patterns in the growing numbers of data records and customer attributes available;
(2.) A high degree of automation in analytics, prediction, decision-making and campaign execution allows for timely reaction to events, and allows to cope with big data sets;
(3.) Self-adaptability accounts for changes in the data sets and the domain structure, as well as for lessons learned of past campaigns, embedding them in a closed feedback loop for autonomous campaign optimization;
(4.) Scalability of the algorithms to a distributed big data processing environment provides the necessary computing power to analyze the increasing amount and velocity of (big) data sets in time;
(5.) A high degree of model accuracy allows for detailed or even personalized marketing decisions;
(6.) A high degree of interpretability (comprehensibility) of the domain model and derived predictions, decisions and automatic campaign executions ensure acceptance by marketers.

In the marketing field, acceptance is of pivotal importance. Campaign managers must be able to explore influence factors and variable dependencies in order to design appropriate marketing instruments (such as special offerings for young adults with high sms-frequency). They also need to understand the rationale for an automatically triggered marketing action in order to be able to justify it to potential business customers and to top management, e.g., by pinpointing relevant influence factors. Even though the high degree of automatization is desired, the choice of contents of a marketing campaign is usually not entirely reducible to structured information and often includes highly emotional contextual information (such as the choice of a protagonist in a TV spot), therefore the possibility of manual intervention of marketing campaign must be guaranteed by any decision support system in the marketing area. Particularly, it must allow the campaign manager to integrate his experience and real-world background knowledge, as well as to implement strategic managerial requirements.

The whole marketing effort serves the purpose of sustainable increase (or of interrupting decrease) of the usage of the provider's service, in extreme case, of preventing clients churn. In contrast to postpaid segment, ultimately fewer resources are used in prepaid segment on maintaining contacts with customers and marketing analytic activities in general. This lack of interest in prepaid segment is caused by its low market share in Switzerland. As a consequence, we rely here purely on data induced by the service usage. Further, due to its nature, the service usage by prepaid customers is usually not as homogeneous as by postpaid clients. Thus, the data available on prepaid users are much more inadequate. This adds to the complexity of modeling and causes a decrease in the model accuracy. The existing marketing solution rates the success of the campaign based on the acceptance ratio. This clearly does not bring the desired effect. Customers, which exploit the offered bonus but do not change their usage habit, are counted as profitable, whereby in real they generate a loss. Additional requirements have been set on the designed system and its models, such as simplicity and easy interpretability, possibility of intervention, capability of effective and manageable visualization, scalability, ease of use.

III. Conceptual model

Two use cases have been specified based on obtained campaign data:

(1.) the semi-automatic support of optimized campaign template instantiation for increased acceptance ratio, and
(2.) the semi-automatic support of campaign template design for increased revenue creation.

Specifications included existing situation, description, use case properties (actors, stakeholders, levels, triggers, frequency, preconditions, post-conditions), data characteristics (names, accessibility, type, volume, temporal resolution, heterogeneity, inconsistency, constraints, analysis needs) and possible use case extensions. They are the result of workshops with the telecom partner prepaid marketing team and a number of project internal workshops (HSLU-I and Consulteer AG).

The conceptual solution addresses the following issues:

(a.) *automatization of the whole marketing process (from data mining to decision making);*
Closed loop architecture with constantly trained and updated model has been specified and build for the developed prototype, containing modules for feature preselection, prediction, evaluation and adaptation. This was done in two stages. First, methods for modules have been specified and verified with small data in statistical environment and programming language R. Second, the modules have been ported to big data stack (see Section IV) and the loop has been closed. The automatization of the whole marketing process remains a challenge, the reasons are twofold: (a.) *technical* – the built prototype has to be incorporated and verified within the live system; (b.) *non-technical* – the existing marketing processes have to be rebuild towards the high level of automatization and utilizing the implemented solution.

(b.) *allow interactive intervention of marketers (e.g., to implement strategic management decisions or experience based knowledge);*
The selected approach heads towards micro-segmentation by categorization, i.e., filtering out clients with particular combination of input feature values, which evidence the potential of revenue gain if those clients are presented with the bonus. The decision making is presented to marketer in a form of recommendation to select particular small set of clients for the campaign (microtargeting). The implemented solution utilizes the intuitive visualization based on nomogram paradigm (see Section IV). It allows marketer to manually override the automatically chosen (optimal) feature combinations with preferred values and presents the gain/risk predictions for the selected choice.

(c.) *provide interpretability of the automatically determined customer segments needed to support campaign design;*
The solution based on NBC and nomogram visualization satisfies stringent conditions set on the designed system, such as scalability, easy interpretability, possibility of intervention, etc. As opposed to black box models such as neural networks (NN) or random forests, the Naïve Bayes Model provides enough necessary information for marketer to comprehend the reasons for the selection of a client for a campaign. By retracing the decision up to the input features, and identifying those mainly responsible for the classifiers choice, it provides precious insights for the design of new campaigns and allows marketing managers to justify and explain the algorithm's results to their management. This is in contrast to black-box approaches such as deep NN. These have demonstrated strong potential on finding hidden patterns in big data collections, but do not comply with the requirement of interpretability by marketing managers.

(d.) *prediction algorithm needs to be able to deal with volatile and sparse prepaid data;*
Based on data analysis done with provided campaign data and consistent with the campaign management restrictions, the methods have been chosen for data cleaning and transformation, and feature selection, as well as suitable prediction algorithm.

(e.) *usage of behavior change instead of acceptance rate as a success measure.*
Through redesigning the metric for campaign success towards more reliable and long-term change indicator (see Section IV).

The prediction model comprises six parts:

(i.) data transformation methods;
(ii.) input feature construction method (based on data exploration results);
(iii.) feature pre-selection method with dynamic binning;
(iv.) method for target variable construction (implementing a new campaign success measure for behavior change);
(v.) Naïve Bayes prediction (classification);
(vi.) automated customer selection according to class predictions.

Closed loop learning enables a predictive analytics solution to automatically adapt to changing conditions. To achieve this, feedback loops must be integrated in the model architecture. For the given use cases, four levels of feedback have been identified: (1.) Data Level Feedback, (2.) Campaign Level Feedback, (3.) Causal Loops, and (4.) Model Level Feedback. Two of the four feedback levels have been selected for implementation (Data Level and Campaign Level Feedback).

IV. Data Analytics

In telecommunications, the usage data typically consists of time series for different variables such as number of calls, number of SMS sent, date of credit charge and charged value, revenue, etc. The provided data contain for each customer and each variable (called here accumulator) aggregated data for at least three months before and one month after the offer. In

 ISBN: 978-1-61208-614-9

the Figure 1, we show single accumulator data of two chosen customers. The *UserX* accepts the offer (at time denoted by dashed line A) but does not change its behavior remarkably (offer causes a small boost in his usage but it continues to sink which is observable by the slope of the blue line); the *UserY* shows the evident change of the behavior and generates notable growth in the revenue compared to the predicted values (green triangles). It is also possible that we witness the churn prevention in the case of *UserY*.

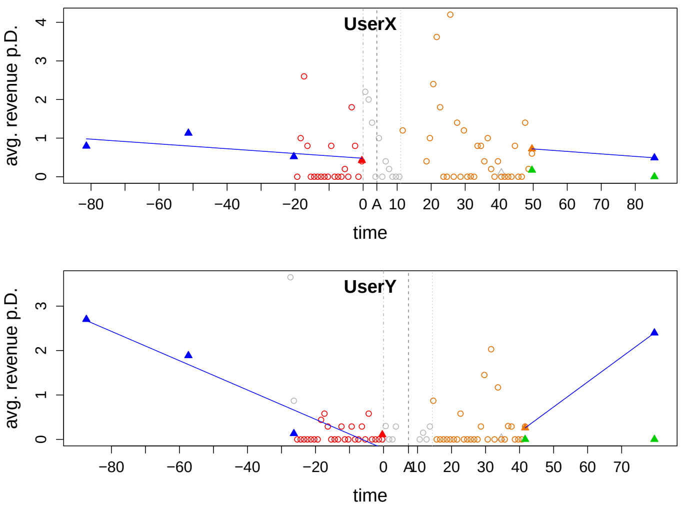

Figure 1. Customer Data Examples

The goal of the campaign is to motivate customers to increase their usage of the service. This is even the weaker condition on the selection of customers and a more general case as the churn prevention. For this purpose, it is important to classify which of the possible customers are likely to generate more usage and thus more profit for the company. Blind assignment of the offers can lead to considerable money loss. E.g., in the case, when a customer accepts the offer but does not change its usage habit (e.g., only exploits the bonus credit). Further, in the opposite case, which also has to be taken into account, the customer feels harassed by the offers he is not interested in, may even decide to churn. We utilize the classical data-driven machine learning pipeline with preprocessing phase (data cleaning and transformation), model building phase (feature extraction, model training and testing loop) and deployment phase. The trained classifier chooses the client for a campaign when it assumes that by doing so, he will increase his usage of the service and thus revenue of the company.

The current solution uses acceptance ratio as measure for the campaign success. We propose conversion shift towards more reliable and long-term behavior change indicator. We regress predictions of the accumulator values after the offer (green triangles, see Figure 1) and train the classifier with the new target variable set as the difference between predicted and real (measured) value. The predictors are variables extracted from the different accumulators, such as average number of national calls, average revenue, etc. The correlation coefficient is used to filter-out the predictors with non-significant influence on the target. In order to provide the "possibility of intervention", "capability of visualization", as well as the high degree of "comprehensibility and interpretability" that is needed for acceptance of an automated solution in the marketing daily business, the Naïve Bayes approach has been chosen.

NBC is simple and effective technique based on the Bayesian theorem. Even though, the Naïve Bayes is not the preferred classification method, its performance is often underestimated. It is fast and space effective, not sensitive to irrelevant features, and can handle streaming data well. For more information, the reader is referred to [26] and [27]. In particular, the binary NBC, where the target class can take only two possible outcomes, allows very elegant visualization with Nomograms [28]. (In case of multi-class classification problem, we create nomogram for one particular outcome class and the union of remaining outcomes represent the complement class.) The NBC nomogram provides a way to visualize the strength of influence of each input feature to a Naïve Bayes classification result, depending on its value or category. It assigns point scores to every predictor variable depending on the chosen feature value. The higher the range of the feature score is, the higher is its influence to the classification result. It thus provides marketers with the possibility to graphically explore how the choice of different feature values (i.e., attributes of a customer) influence the overall class probability (e.g., the probability of accepting a marketing offer or probability of revenue increase). The point scores of all features are easily summed up together and translated into the resulting overall probability. By fixing some attributes to particular values, we effectively filter the target population to microtargets. This allows us to optimize the target with respect to its resulting class probability and size.

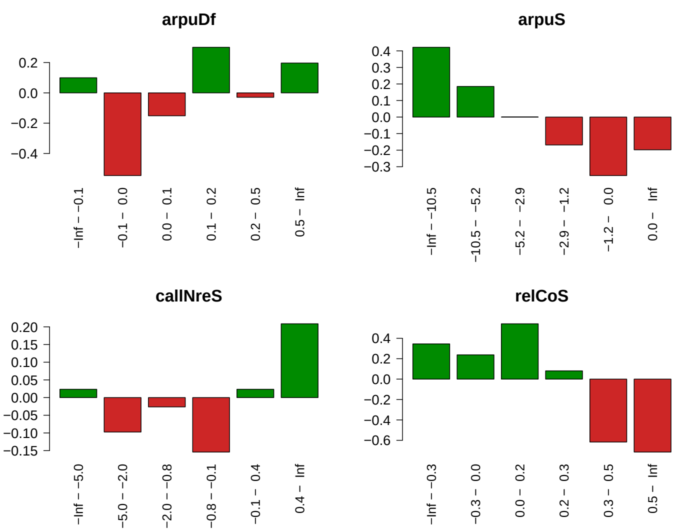

Figure 2. NBC Nomogram with 4 accumulator attributes.

As an underlying infrastructure, big data technology stack has been built based on the Apache SPARK framework for cluster computing. The main components used are: MLlib with Elasticsearch has been used to implement the Naïve Bayes prediction module, Cassandra for scalable storage and high performance data access in closed loop learning, Kafka for real-time streaming capability, and Kibana as a visualization component.

V. RESULTS

A conceptual model of a semi-automated visual campaign design support module has been developed (based on nomograms) that directly shows predicted economic gain/loss in dependency of the selected input variables. More precisely, the model is capable of (1.) visualizing the influence of individual buckets and features on predicted economic gain/loss (2.) visualizing the impact of manual bucket/feature choices interactively, (3.) recommending optimal buckets for a given feature set, and (4.) recommending optimal features for a given campaign. The presumed adaptiveness provided by the closed loop learning module could not be evaluated, since no appropriate data has been provided.

The prediction module has been developed and evaluated on small data using software environment and programming language R (because of data availability). Yet, the used algorithms are scalable and have been also implemented with the equivalent Spark algorithms to provide a proof of concept (1.) of the implemented big data stack itself (see Section IV) and (2.) to validate the R test results for Spark.

The evaluation results of the developed approach show that (1.) the accuracy of the developed prediction module exceeds the accuracy of the two benchmark models used (telecom expert decisions and classification by a decision tree) and (2.) intuitively and understandability of Visualization and Decision Model are satisfactory and have been accepted by the telecom company. Based on the NBC nomogram visualization paradigm, a decision model and an interactive GUI prototype has been developed that allow for semi-automated campaign design support. While the prediction module provides recommendations for customer segmentation w.r.t. an existing campaign, the design of a new campaign (or the optimization of an existing one) additionally requires interpretability of these recommendations in terms of strength of causal influence of input variables and their multidimensional interaction. The NBC nomogram paradigm provides this interpretability in an intuitive way. With the GUI, the marketer is provided with the opportunity (1.) to set manual constraints for an automated optimizer (e.g., to comply with given management strategies), (2.) to choose these constraints based on intuitive multidimensional visualization of feature nomograms, and (3.) to graphically inspect the predicted (simulated) effects of his decisions in terms of expected revenue change.

Even though the performance of the utilized prediction model is not astonishing, it meets the stringent requirements specified on the designed system and still achieves satisfying results if compared with the existing solution. We measured the average gain per person for customers chosen by the trained model and compare it with the existing campaign results. The decreased performance of the prediction model is caused by the low quality of provided data and lower stability of service usage in prepaid segment in general as mentioned in Section II. In Figure 3, we show verification results of 100 rounds performed by the trained classifier. The left three boxplots show the statistic of the data from the chosen marketing campaign if present targeting approach is used. The targeted users have been split into three groups: those which accepted the marketing offer, those which did not accept it (expire) and the control group (baseline). The red boxplot shows average arpu per person and day if microtargeting with NBC model is used, the blue box on the right summarizes the advantage of the new approach compared to the actual targeting method.

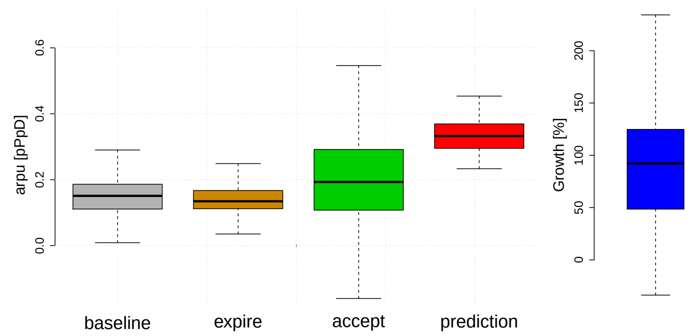

Figure 3. Testing results of 100 rounds.

A relatively new insight in telecom marketing has been gained, that the customer response rate of a digital marketing campaign usually does not appropriately reflect the business value generated by the campaign, and therefore is not an appropriate marketing success measure. More specifically, customer response measures the short-term campaign success, but disregards long term effects on return on investment induced by the campaign. Based on the data-driven marketing paradigm of customer retention beats customer poaching, a new model for measuring the marketing success of digital campaigns for volatile and sparse prepaid telecom data was developed. The metric is a function of the campaign-induced long-term behavior change of customers in the targeted segment and returns an assessment of the overall revenue increase or decline n months after the offer has been accepted. It is applicable to different intended behavior modifications, e.g., churn prevention, usage increase per channel, and can handle low quality, highly volatile and sparse input data.

VI. CONCLUSIONS

The goal of this research is twofold. To develop a methodology and a software prototype for semi-automated predictive analytics and campaign management, and as a proof of concept, to verify the designed system on a chosen use case provided by a large telecom company in Switzerland.

The implemented prototype shows that machine learning can be used to support decision makers in the telecommunication business to optimize marketing campaigns through microtargeting. Moreover, it shows that Naïve Bayes is suitable model if full control need to be granted over the decision support system. A nomogram based decision module has been developed that optimizes the size of micro segments w.r.t. predicted long-term revenue increase caused by stimulus-induced behavior change. The use of long-term revenue increase as a campaign success measure is in contrast to commonly used customer response success measures, which do not take long term effects into account and often disregard the intended effect on return on investment. Besides fully automated optimization over all input features, the decision module also allows the marketer to manually set constraints to the optimizer, e.g., to implement higher-order management strategies or to explore the customer structure in terms of effects on revenue in a simulation run. In order to permit the decision model to adapt to changing customer structures, a closed-loop

 ISBN: 978-1-61208-614-9

learning approach has been used that continuously updates the underlying prediction model and corresponding nomogram values. The solution has been prototypically implemented based on an Apache big data stack and tested on two real-world use cases with prepaid customer data sets provided by the telecom project partner. The prediction accuracy of the Naïve Bayes approach has shown to exceed a decision trees approach as well as the currently used benchmark with mostly manual and experience-based campaign definition.

The detailed description of the prototype developed in this research is omitted in order to honor the non-disclose agreement with the involved industry partners.

Acknowledgment

The research presented in this paper was funded by the Swiss Commission for Technology and Innovation CTI under grand number 18714.1PFES-ES.

References

[1] CIM, "The cost of customer acquisition versus customer retention." 2010, URL: http://www.camfoundation.com/PDF/Cost-of-customer-acquisition-vs-customer-retention.pdf [retrieved: April, 2016].

[2] K. N. Lemon, T. B. White, and R. Winer, "Dynamic customer relationship management: incorporating future considerations into the service retention decision," Journal of Marketing, vol. 66, 2002, pp. 1–14.

[3] J. R. Quinlan, C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[4] M. Mehta, R. Agrawal, and J. Rissanen, "Sliq: A fast scalable classifier for data mining," in Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, ser. EDBT '96, 1996, pp. 18–32.

[5] J. C. Shafer, R. Agrawal, and M. Mehta, "Sprint: A scalable parallel classifier for data mining," in Proc. 22nd Intl. Conf. on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 1996, pp. 544–555.

[6] J. Gehrke, R. Ramakrishnan, and V. Ganti, "Rainforest - a framework for fast decision tree construction of large datasets," in Proc. 24th Intl. Conf. on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 1998, pp. 416–427.

[7] R. Rastogi and K. Shim, "Public: A decision tree classifier that integrates building and pruning," in Proc. 24th Intl. Conf. on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 1998, pp. 404–415.

[8] J. Gehrke, V. Ganti, R. Ramakrishnan, and W.-Y. Loh, "Boat – optimistic decision tree construction," in Proc. 1999 ACM SIGMOD Intl. Conf. on Management of Data. ACM, 1999, pp. 169–180.

[9] G. Kraljević and S. Gotovac, "Modeling data mining applications for prediction of prepaid churn in telecommunication services," Automatika, vol. 51, no. 3, 2010, pp. 275–283.

[10] T. Simsek Gursoy, "Customer churn analysis in telecommunication sector," Istanbul University Journal of The School of Business Administration, vol. 39, no. 1, 2010, pp. 35–49.

[11] C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," Expert Systems with Applications, vol. 23, no. 2, 2002, pp. 103–112.

[12] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn Prediction: Does Technology Matter?" Int. J. Intell. Technol., vol. 1, no. 2, 2006, pp. 104–110.

[13] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," J. Eng. Res. Appl., vol. 2, no. 4, 2012, pp. 694–697.

[14] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer Assisted Customer Churn Management: State-of-the-art and Future Trends," Comput. Oper. Res., vol. 34, no. 10, 2007, pp. 2902–2917.

[15] N. Kamalraj and A. Malathi, "A survey on churn prediction techniques in communication sector," International Journal of Computer Applications, vol. 64, 2013, pp. 39–42.

[16] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2nd ed. Pearson Education, 2003.

[17] S. V Nath, D. Ravi, and R. Behara, "Customer churn analysis in the wireless industry: A data mining approach," 2003, URL: http://download.oracle.com/owsf_2003/40332.pdf [retrieved: April, 2016].

[18] C. K. Kirui, L. Hong, W. K. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," in Int. J. Comput. Sci. Iss. (IJCSI), vol. 10, no. 2, 2013, pp. 165–172.

[19] A. Ultsch, "Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series," in Kohonen Maps. Elsevier, 1999, pp. 33–46.

[20] W.-H. Au, K. C. Chan, and X. Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," Trans. Evol. Comp, vol. 7, no. 6, 2003, pp. 532–545.

[21] Y. Xie and X. Li, "Churn prediction with linear discriminant boosting algorithm," in Proc. 7th Intl. Conf. on Machine Learning and Cybernetics, ICMLC, vol. 1, 2008, pp. 228–233.

[22] H. Farquad, R. Vadlamani, and R. Bapi, "Churn Prediction using Comprehensible Support Vector Machine: an Analytical CRM Application," Applied Soft Computing, vol. 19, 2014, pp. 31–40.

[23] A. Amin, S. Shehzad, C. Khan, I. Ali, and S. Anwar, "Churn prediction in telecommunication industry using rough set approach," Studies in Computational Intelligence, vol. 572, 2015, pp. 83–95.

[24] N. Kamalraj and A. Malathi, "A survey on churn prediction techniques in communication sector," in Int. J. Comput. App. (IJCA), vol. 64, no. 5, 2013, pp. 39–42.

[25] B. Balle, B. Casas, A. Catarineu, R. Gavaldà, and D. Manzano-Macho, "The architecture of a churn prediction system based on stream mining," in Frontiers in Artificial Intelligence and Applications, vol. 256, 2013, pp. 157–166, URL: http://www.lsi.upc.edu/ gavalda/papers/ccia2013Churn.pdf [retrieved: November, 2017].

[26] C. M. Bishop, Ed., Pattern Recognition and Machine Learning. Springer, 2006.

[27] T. M. Mitchell, Machine Learning. WCB McGraw-Hill, 1997.

[28] M. Možina, J. Demšar, M. Kattan, and B. Zupan, "Nomograms for Visualization of Naive Bayesian Classifier," in Proc. PKDD 8th Eur. Conf. Principles and Practice of Knowledge Discovery in Databases, 2004, pp. 337–348.