Performing Machine Learning Techniques in a Contextual Marketing Scenario

Francisco Capetillo, Telefonica Chile

Abstract

Although information for identifying high-potential customers is a key piece in a Customer Value Management strategy, it is hardly available in real-world databases. This paper describes an analytical framework developed to find the next best offer for each of the over five million customers in the mobile prepaid business of Telefonica Chile. This framework is based on the analytical life cycle methodology proposed by SAS® best practices, which promotes the business understanding as a crucial control point in the use of advanced analytics in the real world. A supervised learning approach using artificial neural networks in SAS® Enterprise Miner™ is used to predict purchase behaviors in a window of time based on the state of a set of variables. Then the best offer for each customer is assigned, based on the scoring propensity for every single product. The customer-product data set is used as an input for SAS® Real-Time Decision Manager to accelerate the consumption of the customer's available balance every time a customer made a top up or uses his handset. Finally, as the next step in this process, the generation of predictive models for other business objectives is being planned. In this sense, counting on SAS® Viya® benefits through SAS® Visual Statistics and SAS® Visual Data Mining and Machine Learning, we can improve the experience of use in a production industrial environment with more flexibility and higher performance of the analysis involved.

Introduction

Telefonica is a Spanish multinational telecommunications company headquartered in Madrid, Spain, located as the largest telecommunications company in Europe and the fifth in the world.

Telefonica Chile, called Movistar offers fixed telephony, mobile and long distance; Internet access and switched via broadband ADSL and VDSL; DTH satellite television; transport services satellite radio and TV. Optical fiber to the home and IPTV

Telefonica brands are Telefonica, Movistar, O_2 and Vivo.

Figure 1 – Telefonica brands logos

These brands represent their mobile, landline, Internet and television telecommunication services to 347 million customers from several countries.

Figure 2 – Telefónica Global Presence

In November 2015 Telefonica outlined a five-year plan to become a 'Digital Telco', under the new slogan 'We choose it all'.

Telefonica is a company that is aware of the new challenges posed by today's society. This is why it offers the means to facilitate communication between people, providing them with the most secure and state of the art technology in order for them to live better, and for them to achieve whatever they resolve.

An innovative and attentive spirit with an immense technological potential that multiplies the ability to choose of its more than 346 million clients in 21 countries and an average of 127,000 professionals.

In January-September 2017 Telefonica had consolidated revenues of 38,846 million euros and more than 347 million total accesses, 272.4 million mobile phones accesses, 37 million fixed telephony accesses, more than 21.6 million Internet and data accesses and more than 8.3 million pay TV accesses.

Telefonica is a 100% listed company with more than 1.3 million shareholders, traded in several of the most important stock markets around the world.

Business context: The prepaid challenge

In the Chilean telecommunication business from the year 2015 there is a downward trend in the Industry level, especially on the most valuable customers. This trend is due to several reasons such as government regulations and the entry of aggressive new competitors, who end up marking the margins.

This effect is being amplified particularly in the prepaid segment, whose market has been declined in average by 5% year-over-year from 2015 and one of the reasons is the migration to postpaid subscribers due to the fall in the prices in this segment.

This creates a gap at industry level that challenges telcos to generate new products and new consumption dynamics and stimulation that allows maintaining the income levels. But it's in the prepaid market where the greatest challenges arise.
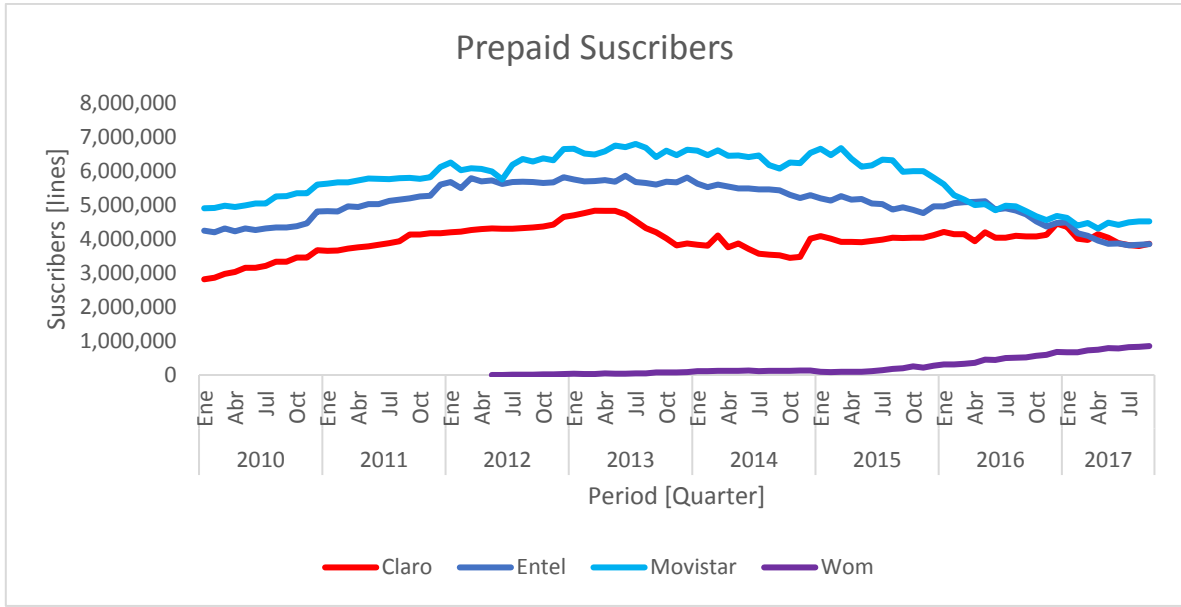
Figure 3 – Prepaid market trend. **Source**: Subtel, 2017.

The prepaid business is characterized by being a transactional business, where customers prepay the consumption of the products and then consume through the usage of minutes or through the acquisition of packs (data or minutes) (figure 4). In Chile, the prepaid market still represents close to 65% of the total mobile communications market.
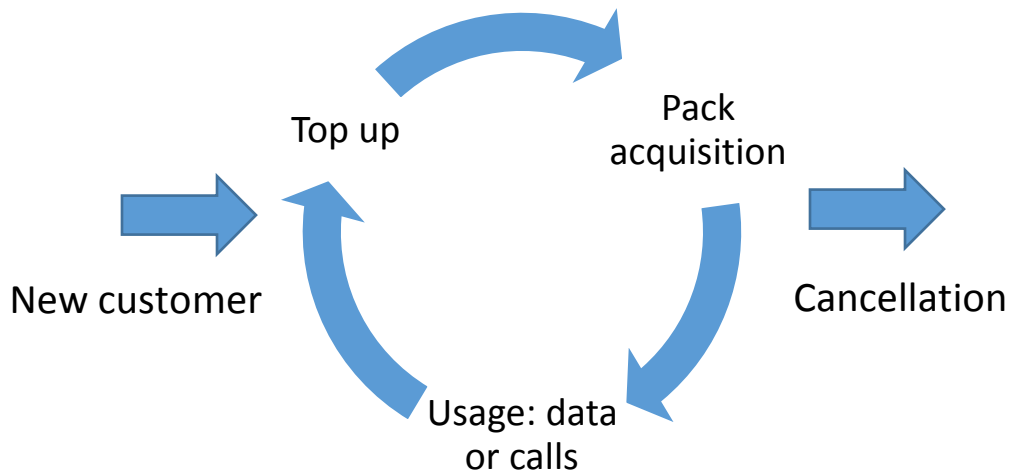

Figure 4 – Prepaid product lifecycle

The usage of mobile products is recorded in the operational systems; thus, it is possible to see in real time or near real time the consumption of minutes, data, use of bags (promotions), consumption of packs and top up.

PrePlan: A new prepaid product

Movistar Prepaid PrePlan service is a periodic bag that has minutes to talk, megas to surf the Internet and megas to use exclusively on social networks (Facebook, Facebook Messenger, WhatsApp and Modo Fútbol (Soccer)). It is automatically renewed according to the validity period it has and it is deducted from the balance of the client depending on the plan that he chooses.

PrePlan is a new category for mobile users that seeks to combine the freedom of prepaid with the benefits of postpaid plans.

This new approach becomes a challenge when determining which the best offer is for each of the customers in the prepaid segment. Currently the offer of PrePlans consists of five products and the solution presented allows each client to have the best offer according to their behavior and preferences. The table below shows the current offer for the PrePlan products.

Product	Voice capacity [MINUTES]	Free use data capacity [GB]	Social Networks use data capacity [GB]	Validity period [DAYS]	Price [CLP]
Preplan 2000	160	1	Unlimited	7	2,000
Preplan 3000	200	1.6	Unlimited	7	3,000
Preplan 5000	300	3	Unlimited	15	5,000
Preplan 7000	400	4	Unlimited	30	7,000
Preplan 10000	600	6	Unlimited	30	10,000

Table 1 – PrePlan offer.

Problem definition: "All models are wrong, but some are useful" [1]

Objective: Assign a PrePlan offer to the entire prepaid customer base.

This offer allocation problem was modeled as a linear optimization problem where it is intended to maximize a function that is the product between the probability for each customer to acquire a specific product (PrePlan) and a binary variable that indicates whether or not that product is assigned to the customer.

Below is shown the problem formulation as a graphic and the formal mathematical definition.

[1] Box, George E. P.; Norman R. Draper (1987). Empirical Model-Building and Response Surfaces, p. 424, Wiley. ISBN 0471810339
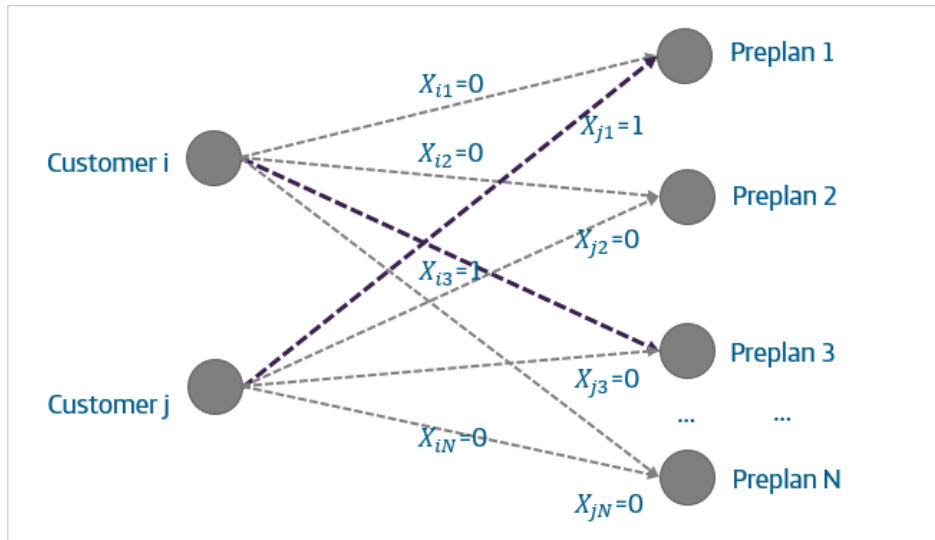
Figure 5 – Graphical formulation of the problem.

Objetive function:

$$maximize \quad \sum_{i \in C} \sum_{j \in P} p_{ij} \times X_{ij}$$

Subject to:

$$\sum_{j \in P} X_{ij} = 1 \quad ; \quad \forall\, i \in C$$

$$X_{ij} \quad \in \quad \{0,1\}$$

Where:

C: Prepaid customers base set.

P: Preplan offer set.

p_{ij}: Probability for customer **i** of buying the product **j**

$$X_{ij} = \begin{cases} 0, & when\ customer\ i\ is\ not\ asigned\ the\ offer\ j \\ 1, & when\ customer\ i\ is\ asigned\ the\ offer\ j \end{cases} \quad ; \quad \forall\, i \in \boldsymbol{C}\,, \ \forall\, j \in \boldsymbol{P}$$

The probability of buying every preplan of the offer for every prepaid customer was calculated through a propensity model for each product. There are currently five PrePlans, so five models were created. For the creation of the models, SAS® Enterprise Miner™ was used as a statistical engine. The input data was prepared in a Hortonworks Hadoop environment mounted in the Big Data Cluster of Telefónica Chile.

For each of the models, the same dataset was used in which the state variables that describe the behavior of each client are maintained and the target variable is changed accordingly.

Inputs for the model

- **Top ups statistics** (3 months): min, max, mode, average, periodicity, days since the last recharge

- **Bags**: Historical purchases of each type of product (PrePlans). Days since last purchase.

- **Traffic**: Voice and data traffic of the previous month

- **Others**: terminal (brand, model, technology), line aging.

The data set for modeling was built by taking transaction data for the first three weeks after the product was launched (trial period), and then it has have been feeding back week after week with significant samples of the purchase transactions and without purchase. In the table 2 it is showed the targets distribution in the training dataset after a long period of cumulative data, about 12 weeks.

PRODUCT	Q	%
PP2K	74,163	6%
PP3K	136,729	11%
PP4K	49,979	4%
PP5K	9,077	1%
PP7K	42,524	4%
PP10K	18,688	2%
WITHOUT PURCHASE	870,000	72%
TOTAL	1,201,160	100%

Table 2 – Target distribution in the training dataset.

The training dataset was divided in two subsets: 70% for training purposes and 30% for validation. Different models were run and compared by the *model comparison* tool included in SAS® Enterprise Miner™ then the model selected for each product was a neural network, which was the best performing model.

In figures 6 and 7 are shown two of the principal performance indicators for the PrePlan 2000 CLP model case and in the table 3 are shown the AUC (area under ROC curve), Lift in the first decile, and the Kolmogorov-Smirnov statistic for each of the five models.
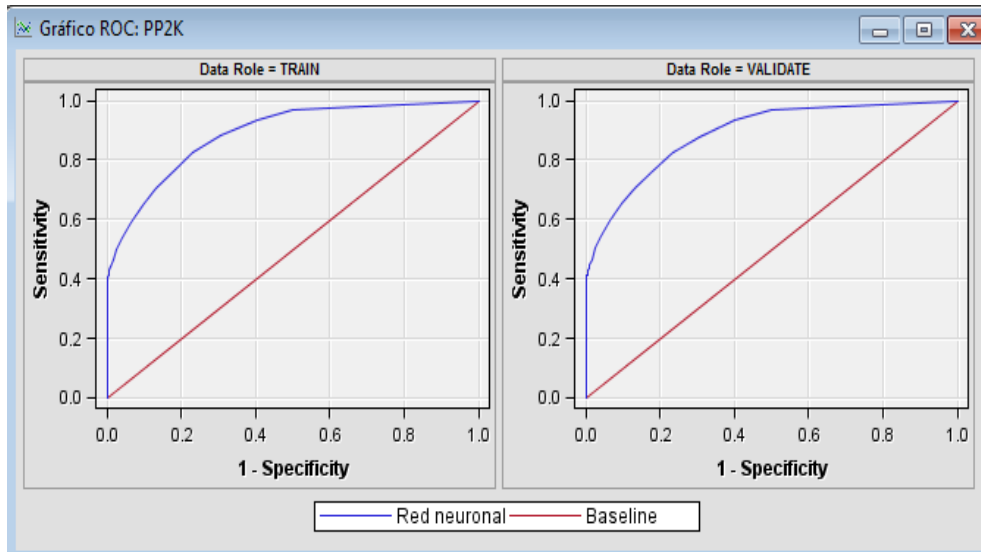
Figure 6 – ROC Graphic for propensity model to purchase PrePlan 2000 CLP, AUC = 0.89
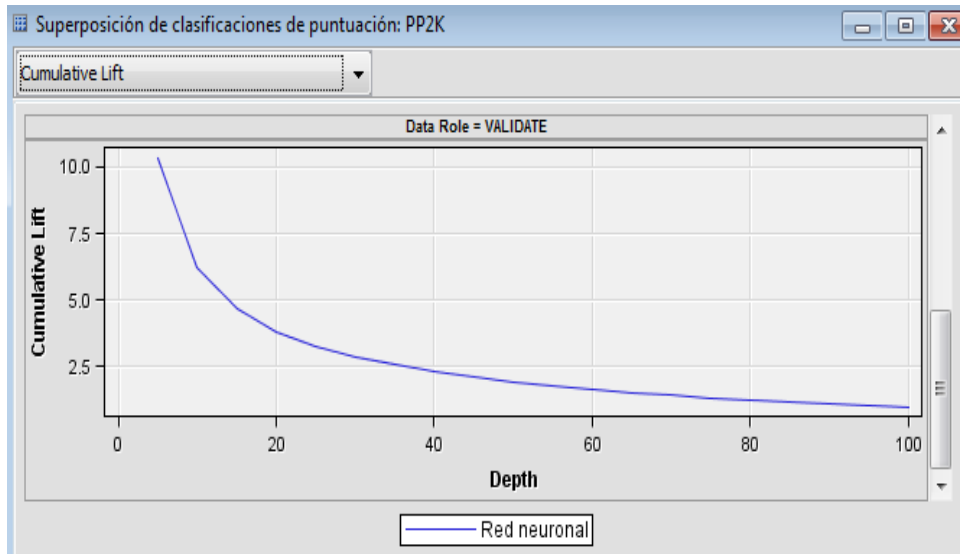

Figure 7 – Lift curve for propensity model to purchase PrePlan 2000 CLP, Lift = 6.42

Model	ROC	LIFT	K-S
PrePlan 2000	0.89	6.42	0.6
PrePlan 3000	0.93	7.86	0.77
PrePlan 5000	0.87	6.66	0.54
PrePlan 7000	0.91	8.46	0.81
PrePlan 10000	0.93	9.21	0.85

Table 3 – Principal performance indicators for each of the five models developed

Once each model was built, tested and validated, then it proceed to the scoring process which implies to score all the prepaid customer base, more than 5 million of lines, for the campaign management in SAS® Real-Time Decision Manager.

When the scores distribution for each model is reviewed it can see that the scores are concentrated in the extremes of the domain range (near to zero and one) which it means that the models are predicting low uncertainty (scores neighboring to 0.5) which is an expected characteristic of a binary classification model. In figure 8 is shown the scores distribution for the PrePlan 2000 CLP.
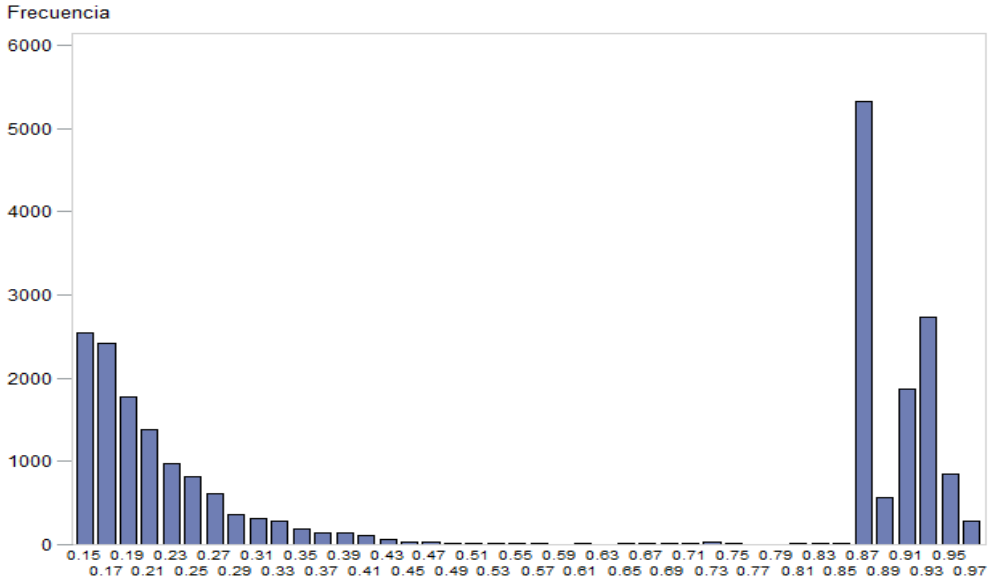

Figure 8 – Scores distribution for PrePlan 2000 model.

Practical Results

The Real Time Decision Business Model consists in that every time a customer made a top up, it receives a message (SMS, SAT Push, USSD or other available communication channel) suggesting the usage of the balance amount in a more convenient promotion than the pay per usage in minutes, SMS or megas for surfing the internet. Obviously the customer can accept or reject this suggestion. More than 130,000 messages were sent daily with an average effectiveness ratio of 12%.

The success of the models are measured by counting the acceptance rates for each one of the five products in every decile of the scores distribution. The figure 9 is showing the sales for every decile of the PrePlan 2000 model in a specific week.
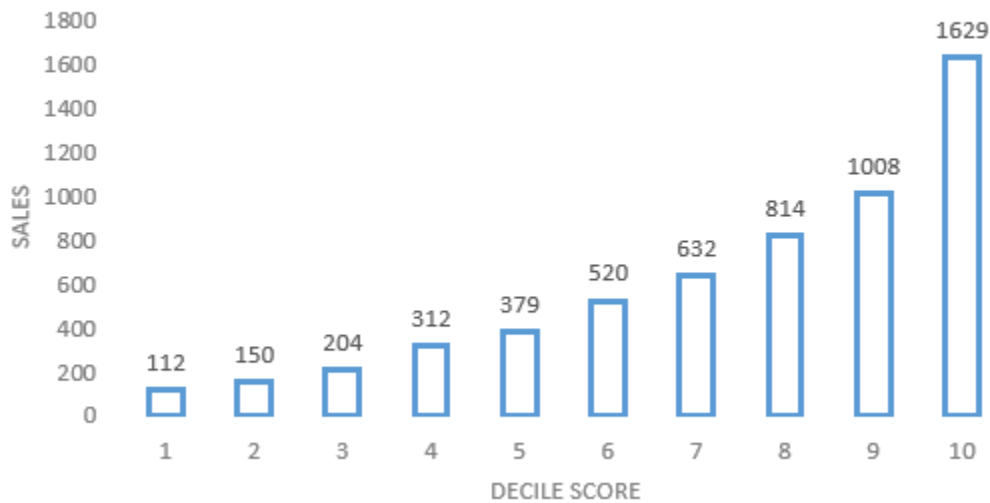
Figure 9 – Sales of PrePlan 2000 CLP

The sales rates in the best scoring deciles are considerably higher than in the first deciles and this demonstrates that the model is correctly picking out between good and bad records.

The Real Time Decision Business Model works with control groups that are selected randomly. They consist of the 5% of the customer base. In this case the effectiveness in the managed group over the control group are more than 2.9 times and that implies the generation of more than CLP$ 260,000,000 / USD$ 430,000 extra income monthly. Also the average revenue per user (ARPU) increase in more than 20%.

Next steps

For the complete integration of the models developed and the SAS suite (SAS® Real-Time Decision Manager, SAS® Marketing Automation, SAS® Event Stream Processing, SAS® Enterprise Miner™, SAS® Enterprise Guide, SAS® Visual Analytics), it is needed to develop an appropriate analytical environment to manage all the prepaid customer base and a new strategy and offer for the customers that have lower scores in all the five models.

It is also important to encourage the top up in the customers that not have sufficient balance amount to buy any of the PrePlans offer and it is important to stimulate the data usage in an important customers segment that only consumes voice and no data

Conclusion

Today, companies save millions of data records daily, many do not take advantage of them, but those that achieve competitive advantages are those that know how to use these assets to improve the relationship and services provided to their customers.

Telefonica is currently undergoing a transformational process in which the use of data is essential for the management of its clients, safeguarding their privacy, of course, and trying to improve the experience they live daily as users.

The framework presented in this paper is one of the methodologies used by the company to manage its client portfolio in marketing campaigns in real time as well as in batch campaigns. This methodology allows a fast development of models that can be produced in a short time and that can be quickly adapted to changes in the definitions of marketing areas, which occur very often in a telecommunication market such as the Chilean one, which is one of the most dynamics in the world.

For this, it is essential to have an advanced analytics team that is able to understand business problems and also has a strong background in the use of data management tools, computer sciences and mathematical modeling.

Bibliography

Box, George E. P.; Norman R. Draper (1987). *Empirical Model-Building and Response Surfaces*, Wiley.

Casella, George; Berger Roger L. (1990). *Statistical Inference*, Duxbury Advanced Series.

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

> Francisco Capetillo Cortés
> Data Scientist
> Telefónica Chile
> francisco.capetillo@telefonica.com

Trademarks of Telefónica