Analysis of Churn in Mobile Telecommunications: Predicting the Timing of Customer Churn

DOI: 10.26573/2019.13.2.3
Volume 13, Number 2
May 2019, pp. 127-141

Barry E. King
Jennifer Rice
Butler University
(king@butler.edu)
(jlrice@butler.edu)

Attrition, or churn, at a Venezuelan mobile telecommunications company was examined using survival analysis. Main contributors to predicting churn were total number of message calls, subscription tenure, and account tenure. Three measures were generated for each account holder: 1) mean residual lifetime, 2) risk of attrition, and 3) customer lifetime value. Customer lifetime value is particularly useful for management in deciding how much to spend to encourage the account holder to stay with the mobile telecommunications company. A company may decrease the probability of a customer leaving by encouraging the customer to enter a loyalty program.

Keywords: Customer Churn, Customer Lifetime Value, Telecommunications, Survival Analysis

1. Introduction

The churn is defined as the rate of attrition, is the percentage of subscribers to a service who discontinue their subscriptions to that service within a given time period. For a company to expand its clientele, its growth rate, as measured by the number of new customers, must exceed its churn rate (Investopedia, 2016).

The global telecommunications industry has become a highly mature and competitive industry over the last couple decades. Maintaining a sound customer base is a primary concern for all telecommunications industries, specifically the mobile telecommunications sector. Marketing efforts designed to sustain market share and increase profit levels have switched from focusing on acquiring new customers to retaining existing customers or reducing customer churn. The telecommunications industry experiences an average 30-35 percent annual churn rate and it costs 5-10 times more to recruit a new customer than to retain an existing one (Lu, J., 2002). Customer churn in the telecommunication industry is the process by which existing customers end their service with a provider. Although churn is an issue in any industry that offers a service or subscription, it is most prevalent in the mobile telecommunications industry. Churn prediction is the ability to detect customers who are likely to end their service or subscription prior to contract termination. Being able to determine which customers are most likely to churn allows a company to target their marketing efforts (incentives and discounts) toward a specific segment of the existing customer base in an attempt to improve customer retention and increase revenue (Owczarczuk, 2010). Churn prediction helps to

allocate current customers into different groups based on their probability of churning. Shapoval and Setzer (2015) discuss term of contract in predicting churn. It is critical to have a solid performing churn prediction model that accurately predicts each customer's probability of churning because it can be extremely costly and inefficient to market to all existing customers, especially loyal and long tenure customers who have very low probabilities of churning. Thus, a critical component to an efficient and successful marketing strategy that improves customer retention is accurately predicting the existing customers who are most likely to churn. To do this, companies must understand their customers churning process and the factors that increase the probability of customer churn.

In the mobile telecommunications industry, it is not only necessary to predict which customers will churn, it is also necessary to calculate when they will do so. Various statistical techniques have been used to develop churn prediction models. Logistic regression and data mining methods, such as neural networks, have often been used to score customers based on their predicted probability of churning. However, both techniques have associated shortcomings. Logistic regression accurately predicts probabilities of churning and explaining why customers will churn based on usage data and customer personal information, but it does not provide any information on when customers will churn. Likewise, neural networks also perform well in predicting probabilities of churning, but this method suffers from the "black box" syndrome, which refers to the challenge of interpreting the results. Neural networks, unlike logistic regression, does not possess the ability to explain why a customer will churn, but rather only provide the predicted probability of churn. These methods also do not offer any information about the timing of when a customer will churn. An alternative statistical technique is survival analysis. Survival analysis is commonly used when the objective of the model is to predict the timing of an event from its origin to its occurrence. This method lends well to churn prediction modeling, where the objective is to predict when a customer will churn from a time of origin. Survival analysis fills the void that logistic regression and data mining techniques leave by offering, in addition to predicting the probability of churning, the timing of when a customer will churn. Like logistic regression, survival analysis identifies the key drivers that cause customers to churn. Survival analysis is becoming the modeling technique of choice since traditional regression techniques identify churners but not predict when they will churn.

The objective of this is to perform an analysis to estimate the probability and timing of customer churn in the mobile telecommunications industry in Venezuela. The main objectives of this analysis are to predict if and when a customer will churn, and what is the customer lifetime value, given the customer's churn prediction. Using data from the mobile telecommunications industry in Venezuela, the model incorporates customer demographics, type of subscription plan and characteristics, and plan usage information to predict whether a customer will churn and the timing of customer churn.

The paper is organized as follows: a comparison of existing churn models and statistical techniques are evaluated, the conceptual framework that drives the empirical model is then discussed, followed by the econometric model and description of data. The econometric methods are described in the next section with the last two sections presenting the results and a conclusion.

2. Literature Review

2.1 Comparison of Churn Prediction Models

Although there is common agreement that it is necessary to have an accurate prediction model to effectively predict customer churn, there is considerable discussion on which modeling techniques offer the best and most accurate predictions. Various types of statistical modeling techniques have been used to develop prediction models, ranging from very sophisticated data mining techniques to basic linear regression methods. Many researchers have used conventional statistical methods, such as logistic regression, ordinal regression, linear regression, and decision trees (Coussement & Poel, 2008; Burez & Poel, 2007; Buckinx & Poel, 2005; Kim & Yoon, 2004; Wei & Chiu, 2002). These methods have proved to be effective and offer stable and accurate predictions. Although many of these studies used data from non-telecommunications industries, such as pay-tv, internet services, retailing, financial services, and network banking; the models are all relevant to any subscription type service, including mobile phone service.

Several studies suggest that the logistic regression performs better than decision trees and linear regression when modeling customer churn in the cellular industry (Owczarчук, 2010; Ahn, Han, & Lee, 2006). Logistic regression offers an easier interpretation of the results and provides detail on the key drivers that cause customers to churn. However, one shortcoming of the logistic regression is that it does not provide any information on when a customer will churn or how long a customer will maintain their subscription (Lu, 2002).

More non-conventional statistical methods have also been used to model customer-churn, such as sophisticated data mining techniques. Oskarsdottir et al. (2016) and Oskarsdottir et al., (2017) use social network classifiers and analytics for predicting customer churn in the telecommunication industry. Many studies have used support vector machines, random forests, and neural networks (Coussement & Poel, 2008; Burez & Poel, 2007; Hung et al., 2006; Buckinx & Poel, 2005; Tsai & Lu, 2009; Zhang et al., 2011). These studies suggest that data mining techniques perform better throughout all percentiles of customers when the customer base is very large, whereas, predictions based on logistic regression begin to break down and become unstable when the number of customers is large. Data mining techniques are designed to deal with large data sets, and thus have more accurate predictions with many customers. However, most data mining techniques suffer from the stigma of the “black box” interpretation – the results are extremely difficult to interpret. Although these techniques provide stable and accurate predictions, they do not identify the key drivers of customer churn, which is important information when developing marketing strategies.

A third group of research focuses on using survival analysis to model customer churn (Van den Poel & Larivie're, 2004; Gopal & Meher, 2008; Lu, 2002; Lu, 2003; Oujezskyy, Horvath, & Skorpil, 2017). Survival analysis is a common statistical technique used when the objective is to model the time to the occurrence of an event (Allison 1995). Thus, it fits well with modeling customer churn, i.e., modeling the time to when a customer cancels their subscription. Using survival analysis, customers are observed beginning at some well-defined point in time and are followed until they churn. Various usage data and personal information is incorporated into the model to help provide more accurate predictions of churners, and to indicate why subscribers churn. Survival analysis, like logistic regression,

accurately predicts churners and indicates the underlying factors of why customers churn but may also become unstable when applied to large datasets. Unlike logistic regression and data mining techniques, survival analysis provides additional information, such as, when will a customer churn and how long will a customer maintain their subscription.

3. Conceptual Framework

To be successful in allocating marketing resources and maximizing profit, it is necessary for firms to attract a set of profitable customers who will maintain long-term relationships with the firm. Marketing efforts are designed to sustain market share and increase profit levels by focusing on retaining existing customers by reducing customer churn rather than acquiring new customers. The key focus of firms marketing strategies is to determine which customers will provide the greatest value from the firm-customer relationship. This concept, known as customer lifetime value, is measured as the present value of revenue arising from a customer's usage of the firm's service. Maximizing customer lifetime value, and ultimately firm profit, requires the ability to successfully identify which customers will provide the greatest customer lifetime value. In turn, a firm can then target their marketing efforts (incentives and discounts) toward the segment of customers that provide the largest customer lifetime value, as a means to improve customer retention and increase profit (Owczarczuk, 2010).

The key to maximizing customer lifetime value is to first understand the components that contribute to customer lifetime value. According to Blattberg, Malthouse, and Neslin (2009), customer lifetime value is determined by four key factors: (1) the length of time of the firm-customer relationship; (2) expected revenues from the customer during their relationship; (3) costs of marketing and servicing the customer throughout the relationship; and (4) the discount rate used to determine the present value of the customer revenue stream. The focus of this research is on the duration of the firm-customer relationship and how various factors, such as customer demographics, product quality, competitors' pricing and marketing, can affect the duration (Blattberg, Malthouse, and Neslin, 2009). Once, the duration of the relationship is determined, then potential customer future revenues can be determined and used to identify those customers who add the most value to the firm's customer base, and ultimately firm profit.

To market toward customers that add significant lifetime value, firms must possess an effective prediction model to distinguish between customers who have a high probability of ending their relationship (churning) and those who have a low probability of terminating their relationship (not-churning). Having a churn prediction model that accurately predicts customers' probability of churning and help direct marketing efforts can help reduce marketing costs since it can be costly to market to all existing customers. Thus, a critical component to an efficient and successful marketing strategy that improves customer retention is to accurately predict which existing customers are most likely to churn. To do this, companies must understand the factors that affects the probability that a customer will churn (terminate their relationship with the firm). In turn, firms can improve their efficiency of marketing campaigns, thus reducing marketing costs and ultimately increase firm profit.

Using the customer relationship framework of Blattberg, Malthouse, and Neslin (2009), this research incorporates customer demographics, usage and service characteristics, and payment and pricing characteristics to predict the probability of churning by distinguishing between the characteristics that increase the likelihood of customers terminating their relationship and those characteristics that decrease the likelihood of churning. Using this information, firms can calculate customer lifetime value and develop marketing activities that will result in valuable customer responses that will positively contribute to firm profit.

4. Data and Methodology

The data used is from the mobile telecommunications industry in Venezuela. We wished to use data from a non-United States type of government system which has highly regulated telecommunications. Venezuela is a federal presidential republic where the president is both the head of state and head of government. We further desired to study a period before Venezuela's inflation problem that began in 2012. The carrier that provided the data is a multinational company operating in Venezuela. It allowed us to use their 2007-2008 data because they felt more recent data may expose proprietary information.

The initial data set contained 151 variables and observations for 20,000 customers, however, many observations had missing values. Variables with more than 20 percent of values missing were eliminated. Missing values for the remaining interval variables were imputed using SAS's PROC MI, and missing values for the remaining categorical variables were imputed from the value with the highest frequency observed. Subscriptions included in the final data had to be active on January 23, 2007 and were followed for 14 months through April 30, 2008. Because the data used in this empirical analysis is proprietary, the name of the company is anonymous, and value codes are not described in detail.

Using these data, the model incorporates customer demographics, type of subscription plan and characteristics, and plan usage information to predict whether a customer will churn and the timing of customer churn.

4.1 Dependent Variable

In survival analysis, the origin of time is a point in time from which the event of interest, leaving the mobile telecommunications firm, is measured. The origin of time for this study is January 23, 2007 and extends through April 30, 2008. The dependent variable (DUR) is the duration of a customer's subscription in months from January 23, 2007 until the customer ends their mobile telecommunications subscription. It is the number of months from the origin of time to the close date or, for censored cases, the number of months, 16, from the origin of time to the end of the study.

4.2 Independent Variables

Several independent variables are used in the study to determine the key factors that affect when a subscriber will churn. Variables include both categorical and continuous variables. Variables are shown in Tables 1 and 2.

The data does not provide many subscriber characteristics. Tables 1 and 2 show that the only subscriber demographics included in the analysis are subscriber gender and age. Gender is a categorical variable in which subscribers are classified as

female, male, or unknown (gender status was not identified in the data). Customer age is a continuous variable that includes all customers 18-71 years old and has an average subscriber in the sample being around 34 years old.

Table 1 *Categorical Variables*

Variable	Number of Levels	Level Codes
Customer gender	3	F, M, U
Customer type	2	B, C
Loyalty program member category	13	1, 3, 4, 8, 9, 11, 14, 18, 20,21,23, 25, 26
Payment collection status	2	LEG, OK
Price plan type	5	CAR, DAT, FLE, MAX, STA
Primary payment method	5	1, 2, 3, 4, 5
Subscription type	8	100, CFX, CIN, CMX, PRE, RFX, RMX, UND

Table 2 *Continuous Variables*

Variable	Statistics	Transformation	Filter
Account tenure (number of months account has been open)	mean: 17.05 , sd: 7.66		
Customer age (years)	mean: 34.35, sd: 11.29		≥ 18 , 71
Last purchase date (last date of a purchase against the account)	range: 25FEB2005 -- 30APR2008		
Total number of data calls	mean: 0.013, sd: 2.90		
Total number of message calls	mean: 37.34, sd: 116.61		
Total revenue for message calls		Log	
Total minutes of usage for voice calls	mean: 1.16, sd: 4.81		
Total number of voice calls	mean: 112.45, sd: 235.49		
Total revenue for voice calls		Log	

4.2.1 Categorical Variables

Table 1 provides a list of the categorical variables that are used in the empirical analysis. In addition to customer gender, the data includes information on type of customer and customer loyalty program membership. Customer type is divided into two different categories, while customer loyalty status has thirteen categories. Various variables are included to capture subscriber payment methods, pricing plan, and subscription type. To measure how payment methods affect a subscriber's churn rate, five different payment methods are used in the analysis and two payment collection statuses are incorporated into the model. In addition to payment methods, a subscriber can be grouped into one of five pricing plans and one of eight subscriptions plans.

4.2.2 Continuous Variables

Table 2 shows the continuous variables that are used in the model. The continuous

variables are designed to capture service usages and revenue. To determine the effect of usage on subscriber churn, patterns of service usage enters the model as data usage, message usage, and voice usage. To capture data and message, the model uses the total number of monthly data calls and message calls. Table 2 indicates that customers, on average, completed very few if any data calls and a significant number of message calls (37.34 per month). Usage of voice calls is measured as the total number of monthly voice calls as well as the average number of minutes for a voice call. Subscribers made an average number of 112.45 voice calls with an average time of 1.16 minutes per month (Table 2). To control for firm revenue, two variables are used in the model. Total revenue for voice calls and total revenue for message calls. Both variables enter the model as log transformations. Another continuous variable is account tenure, which measures the length of time a subscriber in the data has had a subscription. Table 2 shows that on average, subscribers in this data have had a subscription for 17 months.

5. Model Estimation and Results

5.1 The Model

The 20,000 observations in the dataset were split into three datasets: a training dataset consisting of approximately 80 percent of the observations (15,987 observations), a scoring dataset of non-censored observations capturing about ten percent of the observations (1955 observations), and a holdout dataset of censored data of about ten percent of the observations (2058 observations). Using the training dataset, we applied survival analysis to estimate the churn prediction model and determine the factors that affect the customer churn rate. We used the results from the training data to score the holdout data to find predicted probabilities for churning for the customers in our database.

Survival analysis was selected as the method to further analyze the data since, unlike log it analysis which only predicts if a customer will churn, survival analysis provides a probability of churning along with a prediction of when the customer will churn. The period of churn is January 23, 2007 to April 30, 2008, a time of 16 months. For the 16-month time, there were 9,774 subscribers who churned and 10,226 who did not churn. This indicates that 49 percent of the customers in the data who had a subscription on January 23, 2007 ended their subscription by the end of the time period of April 30, 2008.

The results of the survival analysis as performed by SAS's LIFETEST procedure, are reported in Table 3 and Figure 1. The total number of subscribers that churned during the 16-month period was 9,774. Table 4 shows that the largest number of subscribers churned between 10 and 12 months, accounting for 58% of all subscribers who churned during the 16-month study period. Between 10 and 11 months, 31.5% of subscribers churned and between 11 and 12 months 26.5% churned. The probability for churning between 10 and 11 months is 19.1% and 19.7% for 11 to 12 months. Overall, Table 3 indicates that 96.6% of all churning subscribers churned between 8 and 12 months. After having a subscription for 12 months, the probability of churning drops considerably to 2.8% in the 13th month of a subscription and continues to drop to 0.49% at month 16. The results in Table 3 suggest that if customers are going to end their subscription that they will do so within a year of having the subscriptions. Once a subscription hits a tenure of 12

months, the likelihood that they will churn in the next couple months is extremely low (less than one percent).

Table 3 Results of Survival Analysis

Interval		Number Failed	Conditional Probability of Failure	Survival
[Lower,	Upper)			
0	1	0	0	1.0000
1	2	0	0	1.0000
2	3	0	0	1.0000
3	4	0	0	1.0000
4	5	0	0	1.0000
5	6	0	0	1.0000
6	7	0	0	1.0000
7	8	0	0	1.0000
8	9	2394	0.119700	1.0000
9	10	1374	0.078000	0.8803
10	11	3085	0.190100	0.8116
11	12	2592	0.197200	0.6573
12	13	292	0.027700	0.5278
13	14	7	0.000682	0.5132
14	15	5	0.000488	0.5128
15	16	25	0.004870	0.5126
16	.	0	0	0.5101

Figure 1 Survival Probabilities

These results are further supported by Figure 1, which shows the probability of survival curve (the Survival column of Table 3). Subscribers are very unlikely to churn during the first eight months of a subscription, as shown by the horizontal line at the survival probability of 1.0. After hitting the ninth month of a subscription, the probability of keeping the subscription steadily declines until about the twelfth

month, as noted by the negatively sloped portion of the survival curve. This corresponds to the increasing probability of churning as given in Table 3. Once a subscription hits a low around twelve months, the probability of surviving levels off, which suggests that subscribers are likely to keep their subscriptions and not churn.

The results of the survival analysis are used to generate three important aspects of churning for firms: mean residual lifetime, risk of attrition, and customer lifetime value. All three of these measures are calculated for each customer using the holdout data set. These measures provide extremely important information for companies regarding their customer's behavior and churning process. The ability of a company to identify accounts that are of high-risk of churning and accounts that offer high lifetime value allows them to more narrowly focus their marketing efforts and incentive programs in hopes of increasing revenue.

Tables 4, 5, and 6 provide the results for the following three measures: 1) mean residual lifetime, 2) risk of attrition, and 3) customer lifetime value. All these measures are calculated using the results of the survival analysis which is based on the set of independent variables that were previously discussed, such as, customer demographics, payment structure, pricing plan, and service usage (see Tables 1 and 2). The three measures are calculated for each customer in the holdout data. Tables 4, 5, and 6 provide values for a few customers in the data.

Table 4 *Mean Residual Lifetime*

Observation	Account ID	Mean Residual Lifetime
		(months)
1	12449	6.630
2	14466	8.440
3	14157	8.452
4	13601	8.455
5	14825	8.461
6	14958	8.484
7	14369	8.484
8	13641	8.502
9	15536	8.515
10	14145	8.516
...		

Table 5 *Risk of Attrition*

Observation	Account ID	Churn Probability
1	1154	0.9999
2	1155	0.9998
3	1162	0.9826
4	17180	0.9472
5	17188	0.9472
...		
2054	15791	0.0376
2055	15795	0.0345
2056	15781	0.0338
2057	15787	0.0332
2058	7987	0.0004

Table 6 Customer Lifetime Value Discounted at 20 Percent per Year

Observation	Account ID	Monthly Revenue (USD)	Expected Number of Months Before Churn	Current Month	Present Value Until Churn (USD)	Customer Lifetime Value (USD)
1	3800	316	4	Nov-07	716	4082
2	7092	160	8	Nov-07	721	2067
3	1540	359	3	Nov-07	619	4636
4	5820	262	4	Nov-07	592	3379
5	9682	260	4	Nov-07	589	3359
...						
2054	15880	-31	2	Nov-07	-37	-400
2055	114	-20	6	Nov-07	-68	-259
2056	461	-29	6	Nov-07	-97	-371
2057	7508	-38	5	Nov-07	-106	-488
2058	1778	-64	4	Nov-07	-146	-831

The mean residual lifetime is the number of months that a subscriber is expected to maintain their subscription before churning (canceling subscription). This predicted (scored) value is provided by the SAS Enterprise Miner Survival Node. Table 4 shows the mean residual lifetime in months for a sample of 10 observations in the data. Overall the residual lifetime ranges from one month to sixteen months for the 2,058 customers included in the holdout data.

Table 5 provides a churn probability that can be used to identify those accounts that are considered high-risk for churning. The churn probability ranges from less than 1 percent to a high of 99.9 percent across the 2,058 subscribers in the holdout dataset. The probability of churning is influenced by a subscriber's demographics, pricing plan, membership in loyalty plans, and service usage as determined by SAS's LIFEREG procedure.

Present value until churn and customer lifetime value are reported in Table 6. To calculate the present value until churn for a subscriber we use the monthly revenue for an account, based on current subscription and pricing plans, then we use the number of months before a customer will churn, based on their churn probability, discounting the predicted value at 20 percent per year (a typical value for a discount rate), we obtain the customer's present value until churn to the firm.

$$PVUC = MR \sum_{i=1}^T (p^i / \left(1 - \frac{r}{12}\right)^{i-1})$$

PVUC is the present value until churn beginning with the current month, November 2007.

MR is the monthly revenue.

T is the number of months until churn.

p^i is the probability of churn for period i. See the Survival column of Table 3.

r is the annual discount rate. Here the annual discount rate is 20 percent leading to $(1 - 0.2 / 12) = 0.9833$

For example, the present value until churn for Observation 3, Account ID 3800, that is expected to churn four months after the current month, November 2007, is calculated

$$316 (0.6573/0.9833^0 + 0.5278/0.9833^1 + 0.5132/0.9833^2 + 0.5128/0.9833^3) = 716.$$

Customer lifetime value is defined as the present value of average monthly revenue discounted for the number of months the customer is expected to maintain their subscription (Lu, J., 2003). Here we are using a planning horizon of 20 months. The equation is the same as that for present value until churn except T is now 20. The customer lifetime value for Observation 3, Account ID 3800, is 316 $(0.6573/0.9833^0 + 0.5278/0.9833^1 + \dots + 0.5101/0.9833^{19}) = 4082$.

The telecommunications firm may consider spending up to the present value to churn on incenting a customer to maintain their subscription with hopes of receiving the customer lifetime value.

5.2 Variable Importance

Figure 2 shows those variables that are important to predicting telecommunication churn.

Figure 2 Variables that are Important to Predicting Churn

6. Conclusion

This study investigated the application of survival analysis to customer churn to develop a prediction model using data from a non-United States mobile telecommunications company. We applied survival analysis to a sample of 20,000 customers. The data set included customer demographics, pricing and payment details, subscription plan and loyalty membership, as well as service usage from January 23, 2007 through April 30, 2008.

We began with dividing the data into three subsets of data: (1) the training data: 80% of observations; (2) holdout data: 10% of observations; and (3) non-censored data: 10% of observations. Using the training data, we applied survival analysis to estimate the churn prediction model and determine the factors that affect the customer churn rate. We used the results from the training data to score the holdout data to find predicted probabilities of churning for customers in our database.

The results of the modeling are used to generate three important aspects of churning: mean residual lifetime, risk of attrition, and customer lifetime value. All three of these measures are calculated for each customer in the holdout data set. These measures are important variables for a company to completely understand their customer's behavior and churning process.

The survival analysis results successfully answer the question of if and when a subscriber will churn (cancel subscription). Using the results from the survival analysis, we use present value methodology to determine customer lifetime value to the firm.

Prior to eight months and after twelve months, the probability of churning is very near zero. Given this result, firms should focus their incentive plans and marketing efforts toward subscribers who are nearing the end of their first year of their subscription. Table 7 suggests offering a loyalty program scheme may be one incentive if the cost of offering the loyalty program does not exceed the customer's lifetime value.

Table 7 *Loyalty Program Scheme and Account Status*

Account status	Enrolled in a Loyalty Program?	
	No	Yes
Closed	9642	0
Open	10	348

The results of this study provide critical insight to mobile telecommunication firms. The survival analysis provides an accurate way to identify and predict subscribers who are at high risk for churning. This is essential for a firm to know since they want to focus their marketing efforts (incentives and discounts) toward existing customers who are likely to churn as a way to improve customer retention and increase revenue. Churn prediction helps to allocate current customers into different groups based on their probability of churning. It is critical to have a solid performing churn prediction model that accurately predicts each customer's probability of churning because it can be extremely costly and inefficient to market to all existing customers, especially loyal and long tenure customers who have very low probabilities of churning. The model presented in this paper provides a way for firms to calculate predicted churn probabilities and segment the customer population into high and low risk accounts. In the mobile telecommunications industry, it is not only necessary to predict which customers will churn, it is also necessary to predict when they will churn. The results of this analysis suggest that customers have the highest probability of churning around the tenth to twelfth month of their subscription. Prior to eight months and after twelve months, the probability of churning is very near zero. Given this result, firms should focus their incentive plans and marketing efforts toward subscribers who are nearing the end of the first year of their subscription. Offering incentive programs and marketing toward customers who have had a subscription longer than twelve months will not provide the highest return to the firm's investment in incentive and marketing programs.

Customer lifetime value is also an important statistic for firms to consider when choosing which subscribers to target with incentive and loyalty programs.

Subscribers who will offer higher customer lifetime value should be the chosen target population for incentive and marketing programs. Focusing efforts on retaining subscribers who are likely to churn soon and offer lower lifetime value will be costly and offer little in terms of increased revenue.

Although the research presented in this project provides important insights and helpful information for mobile telecommunications firms, especially in the area of marketing efforts and customer lifetime value, the research is not without its limitations. The data used is from January 2007 to April 2008 and, while the data is older, it still provides a description of the key factors that affect the duration of the customer-firm relationship in the mobile telecommunications industry. Furthermore, we do not have data on competitor pricing and marketing efforts or on prior customer service experience with the existing company. It may be possible that these factors affect the probability of customer churn, however, if customers have maintained a long relationship with their current provider, they may be less likely to be influenced and persuaded by competition pricing and promotions.

7. References

1. Ahn, J., Han, S., and Yung-Seop, L., 'Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry', *Telecommunications Policy*, 30, 2016, 552-568.
2. Allison, P., *Survival analysis using the SAS system: A practical guide*. SAS Institute Inc., Cary NC, 1995.
3. Blattberg, R., Malthouse, E., and Neslin, S., 'Customer Lifetime Value: Empirical Generalizations and Some Conceptual Questions', *Journal of Interactive Marketing*, 23, 2009, 157-168.
4. Buckinx, W. and Van den Poel, D., 'Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting', *European Journal of Operational Research*, 164, 2005, 252-268.
5. Burez, J. and Van den Poel, D., 'CRM at a pay-tv company: Using analytical models to reduce customer attrition by targeted marketing for subscription services', *Expert Systems with Applications*, 32, 2007, 277-288.
6. Coussement, K. and Van den Poel, D., 'Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques', *Expert Systems with Applications*, 34, 2008, 313-327.
7. Gopal, R. K., & Meher, S. K., 'Customer churn time prediction in mobile telecommunication industry using ordinal regression', In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 884-889). 2008, Springer Berlin Heidelberg.
8. Hung, S., Yen, E., and Wang, H., 'Segmentation based on hierarchical self-organizing map for markets of multimedia on demand', *Expert Systems with Applications*, 31, 2006, 515-524.
9. Investopedia, Retrieved from <http://www.investopedia.com/terms/c/churnrate.asp>, 2016.
10. Keaveney, S., 'Customer switching behavior in service industries: An exploratory study', *Journal of Marketing*, 59, 1995, 71-82.

11. Kim, H. and Yoon, C., 'Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market', *Telecommunications Policy*, 28, 2004, 751-769.
12. Lu, J., 'Predicting customer churn in the telecommunications industry – an application of survival analysis modeling using SAS', *Proceedings of SAS Users Group International* 27, Orlando, FL, 2002, paper 114-27.
13. Lu, J., 'Modeling customer lifetime value using survival analysis – an application in the telecommunications industry', *Proceedings of SAS Users Group International* 28, Seattle, WA, 2003, paper 120-28.
14. Oskarsdottir, M., Bravo, C., Verbeka, W., Sarrute, C., Baesens, B., and Vanthiensen, J., 'A comparative study of social network classifiers for predicting churn in the telecommunication industry', *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Davis, CA, 2016, 1151-1158.
15. Oskarsdottir, M., Bravo, C., Verbeka, W., Sarrute, C., Baesens, B., and Vanthiensen, J., 'Social network analytics for churn prediction in Telco', *Expert Systems with Applications: An International Journal*, 85 Issue C, 2017 204-220.
16. Oujezsky, V., Horvath, T., and Skorpil, V., 'Botnet C&C traffic and flow using survival analysis', *International Journal of Advances in Telecommunications Electro techniques, Signals and Systems*, 6(1), 2017.
17. Owczarczuk, M., 'Churn Models for Prepaid Customers in the Cellular Telecommunication Industry using Large Data Marts', *Expert Systems with Applications*, 37, 2010, 4710-4712.
18. Shapoval, K. and Setzer, T., 'Term of Contract and Portfolio Aware Churn Modeling in Telecommunication Campaigns', *2015 IEEE 17th Conference on Business Informatics*, Lisbon, 2015, 26-32.
19. Tsai, C. and Lu, Y., 'Customer churn prediction by hybrid neural networks', *Expert Systems with Applications*, 36, 2009, 12547-12553.
20. Van den Poel, D. and Larivie're, B., 'Customer attrition analysis for financial services using proportional hazard models' *European Journal of Operational Research*, 157, 2004, 196-217.
21. Wei, C. and Chiu, I., 'Turning telecommunications call details to churn prediction: A data mining approach', *Expert Systems with Applications*, 23, 2002, 103-112.
22. Zhang, Y., Liang, R., Li, Y., Zhang, Y., and Berry, M., 'Behavior-based telecommunications churn prediction with neural network approach', *2011 International Symposium on Computer Science and Society*. DOI: 10.1109/ISCCS, 2011, 89. 2011.

About Our Authors

Barry E. King is an associate professor of information systems and operations management at the Lacy School of Business, Butler University, Indianapolis, Indiana, USA. His research has appeared in *AIMS International Journal of Management*, *Management Science*, *Harvard Business Review*, and elsewhere. He is an active member of the Decision Sciences Institute and consults regularly with American firms. His current interests are in machine learning and operations research.

Jennifer L. Rice is an assistant professor of Economics at the Lacy School of Business, Butler University, and Indianapolis, Indiana, USA. She received her Ph.D. in Economics from Indiana University, Bloomington, Indiana, USA. Her interests include applied econometrics, machine learning, and health economics. Her current research focuses on the application of econometric analysis and machine learning to a variety of industry settings, such as Health Care and professional sports. She has published in AIMS International Journal of Management, Applied Economics, Applied Economics and Health Policy, International Journal of Pharmaceutical and Healthcare Marketing, and elsewhere.