

WHITE PAPER

Using the Profitability Factor and Big Data to Combat Customer Churn



Succeed. Transform. Compute. Perform.

Using the Profitability Factor and Big Data to Combat Customer Churn

EXECUTIVE SUMMARY

In the increasingly competitive mobile telecommunications industry, retaining customers is a challenge that Tier 1 carriers are aggressively rising to meet. While many marketing efforts to combat churn are still reactive, the Alacer Group recently developed a model for an American telecommunications company that outlined proactive approaches for contacting customers before they decide to change service providers. Looking at more than 70 different data points on over 70,000 customers, we created a binary logistic regression that could predict with 60% accuracy whether or not a customer will churn. We then considered the Probability Factor, and calculated each customer's value to the carrier to create individual Profit-Churn scores. The results determined where each customer fit into a four-quadrant matrix, which gave the service provider a road map for determining the best actions for each audience.

A PERVASIVE PROBLEM: CUSTOMER CHURN

The American mobile telecommunications industry has matured from its early days of explosive growth to a saturated and fiercely competitive market where carriers are battling to simply retain existing customers. The current buzzword within the industry is **customer churn**, a term used to describe customers who voluntarily unsubscribe to join a competitive carrier. It's become so pervasive that the industry now has one of the highest customer churn rates in business; as an example, in August 2012, T-Mobile's churn rate was 2.10 percent, which may seem miniscule until you factor in its millions of subscribers. As a result, many companies are redirecting vast sums of their marketing dollars from developing customer acquisition programs to devising churn management strategies. In fact, the analysts at Gartner Research recently reported that the marketing budget allocated to retaining customers and increasing loyalty through customer service will more than double by 2015.

Most often, service providers are reactionary in their efforts to address voluntary churn. But what if unstructured data could be used to develop a model that could identify the most lucrative customers and help develop a proactive program for ensuring their loyalty? Big data offers just this opportunity.

With this in mind, the Alacer Group recently worked with a Tier 1 service provider to leverage its customer data, which could then be used to create a proactive churn management program. Using 75 relevant data

A close-up photograph of a person's hand holding a black smartphone, with the screen facing away from the camera.

Highlights

To combat customer churn within the mobile telecommunications industry, the Alacer Group used big data combined with an understanding of profitability to help a Tier 1 carrier better target its marketing efforts.

points for each of approximately 70,000 customers, 28.8% of which had churned (see Figure 1), we could design a useful churn prediction model for restructuring marketing operations.

Fig 1–Customer Churn Status, 31-60 days after obs_date				
	Frequency	%	Valid %	Cumulative %
Did Not Churn (0)	50138	71.2	71.2	71.2
Churned (1)	20299	28.8	28.8	100.0
Total	60437	100.0	100.0	

DEVELOPING A CHURN SCORING MODEL

In order to predict the likelihood of a customer churning, it is important to understand what variables may cause a customer to leave in the first place. To do this, we created a Churn Scoring Model using binary logistic regression.

We began by dividing our data set into estimation and validation samples; the first was to be used to create a predictive model, with the validation sample as a test of the model's performance. Next, a logistic regression was created using almost all of the independent variables as inputs and the binary churn variable as the output. If done manually, it would take a long time to test and eliminate the many independent variables that are either not statistically significant or that have high correlations with other independent variables used in the model. Fortunately, IBM's SPSS statistical software has a feature that automates much of the variable selection process.

With most of the 75 independent variables inputted into a logistic regression, we ran an analysis using the *Backward: Conditional* method. This analyzes all variables and then eliminates them one-by-one until a desired model is achieved.

With this regression model, we identified 36 of the independent variables that are statistically significant for predicting customer churn (see *Exhibit 1*); however, at this point the model is not yet complete. In its current format, it only gives us a t score, as shown: $t = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots$

The t score is not very helpful in itself, but it can be converted into a Bernoulli Probability by applying this formula giving us our scoring model:

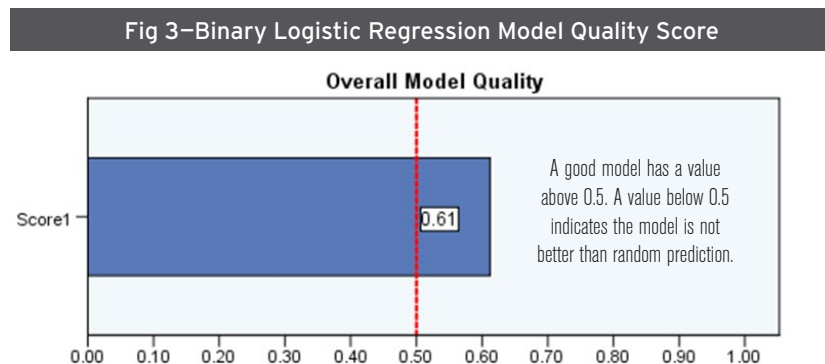
$$p(y = 1|t) = \frac{e^t}{1 + e^t}$$

With the scoring model complete, we can apply it to our validation sample and evaluate its predictive performance. We exported the t score model using the *Propensity to Purchase* option in the Direct Marketing module in SPSS, and applied this to our validation sample using SPSS's *Apply Scores from Model* feature.

As a result, each customer received a Predicted Value and a Selected Probability (that $y=1$). We then grouped the Selected Probability values into seven different bins (Figure 2).

Fig 2–Distribution Of Churn Probability				
Valid probability of churn	Frequency	%	Valid %	Cumulative %
<40%	12698	18.0	18.0	18.0
40-50%	26151	37.1	37.1	55.2
50-60%	22616	32.1	32.1	87.3
60-70%	7361	10.5	10.5	97.7
70-80%	1420	2.0	2.0	99.7
80-90%	173	.2	.2	99.7
90-100%	18	0	0	100.0
Total	70437	100.0	100.0	

We noted that more than 1,000 customers were not given a churn score. Initially, we used the age of the head of household as a variable factor; however, more than 20,000 customers were shown to either not have a listed age (hence no score), or an age of 0 (impossible), so age was not a useful variable.



After removing age as a variable factor, we ran the *Backwards: Conditional* binary logistic regression again (see Figure 3). We also computed a new “accuracy test” variable, which was the sum of the value of the actual churn result (0 or 1) and two times the churn value predicted by our model (0 or 2).

- Values of 0 = the model correctly predicted someone would not churn (42.7%)
- Values of 1 = the model incorrectly predicted someone would not churn (12.4%)
- Values of 2 = the model incorrectly predicted someone would churn (28.4%)
- Values of 3 = the model correctly predicted that someone would churn (16.4%)

“mobile operators can’t rely on adding new high-value contract subscribers to boost their bottom lines.”

Overall, the model was accurate 59.1% of the time in predicting whether or not a customer would churn ($42.7 + 16.4 = 59.1$). We then decided that the profitability of each customer was important to consider as well.

THE PROFITABILITY FACTOR

Sprint CEO Dan Hesse recently confirmed the importance of the Profitability Factor. In his April 2013 conference call with shareholders, he explained that mobile operators can’t rely on adding new high-value contract subscribers to boost their bottom lines. Instead, companies will need to focus on retaining the customers they have, while encouraging individual subscribers to spend more.

That said, the goal for any voluntary churn reduction program is revenue stabilization and augmentation. Looking at a subscriber’s profitability is arguably a better model for developing a targeted marketing program than utilizing overall subscription numbers as a benchmark. Not all subscribers are equal in value; in fact, based on cost vs. profitability, there are some customers that it might be wiser for a service provider to proactively lose from a revenue standpoint.

By considering the profitability of each of customer in addition to his or her likeliness to churn, we can determine if some churn is desirable, identify high value customers who should be proactively targeted to remain loyal, and develop an overall action plan.

In order to analyze the data in a way that provides insight into customer profitability, we used the customer lifetime value (CLV) approach to perform an annual profit & loss analysis for each subscriber. Because the data is missing actual costs, we made the following assumptions:

- Monthly COGs = 40% of monthly revenue
- Monthly Directas Cost/Call = \$0.50
 - Director assisted call representative wage = \$10/hr
 - Director assisted call representative productivity = 20 calls/hr
- Monthly Custcare Cost/Call = \$1
 - Customer care service representative wage = \$10/hr
 - Customer care service representative productivity = 10 calls/hr
- Monthly Retcalls Cost = $\$2.50 * (\# \text{ of Calls} / \# \text{ of Months in Service})$
 - Retention team call representative wage = \$15/hr
 - Retention team call representative productivity = 6 calls/hr
- Monthly Phone Cost = $\$12.50 * \# \text{ of Phones}$
 - Average phone cost to company = \$300
 - Average phone life = 24 months

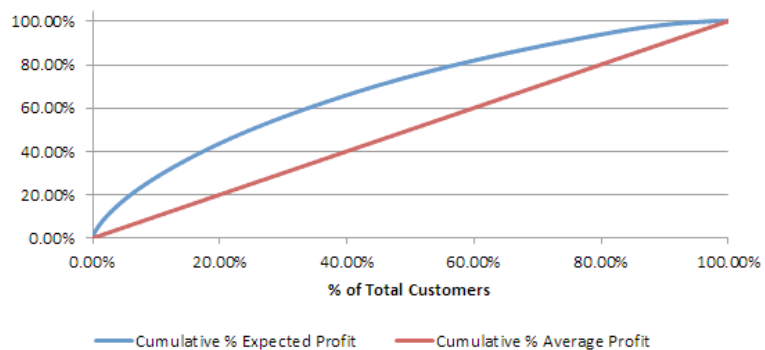
We then took the profitability for each customer and discounted it using a hurdle rate of 10% (we assumed an annual cost of capital of 10%). Using the calculated annual value for each customer and each customer's credit rating, we calculated each customer's ongoing monthly expected profit by using the formula: $\text{Monthly Expected Profit} = (\text{Monthly Revenue} - \text{Monthly Cost}) - (\text{Monthly Cost} * \text{Default Rate})$

Credit Score	300 - 549	550 - 599	600 - 649	650 - 699	700 - 749	750 - 799	800 - 850
Default Rate	71%	51%	31%	15%	5%	2%	1%

For default rates, we used the national average credit default ratings (left).

With each customer's Monthly Expected Profit calculation, we could now compare an individual's profitability. We were astounded to learn that **while 35.37% of the company's customers are more profitable than the average customer, these customers generate more than half (61.52%) of the cumulative Monthly Expected Profit** (Figure 4).

Fig 4–Cumulative Profitability



To better visualize the distribution of profitability, we grouped customers into seven bins according to each customer's expected profit. We observed that 10% of customers are significantly more profitable than the average customer (>1 standard deviation, bins 5-7) and 737 customers appear to be unprofitable (bin 0).

RECOMMENDATIONS

By melding the Churn Model with the Profitability Factor, the service provider gets a useful framework for making more informed marketing decisions, and can concentrate retention efforts on those highly profitable customers who are most likely to churn. These are the customers the carrier cannot afford to lose. Unprofitable subscribers who are likely to churn can either be let go as customers or can be enticed to move to a profitable status.

“Unprofitable subscribers who are likely to churn can either be let go... or enticed to move to a more profitable status.”

Fig 5–Customer Profitability Distribution

Valid	Frequency	%	Valid %	Cumulative %
Negative	737	1.0	1.0	1.0
> -1 Std Dev	2801	4.0	4.0	5.0
< -1 Std Dev	41986	59.6	59.6	64.6
< +1 Std Dev	17810	25.3	25.3	89.9
> +1 Std Dev	4348	6.2	6.2	96.1
> +2 Std Dev	1548	2.2	2.2	98.3
> +3 Std Dev	1207	1.7	1.7	100.0
Total	70437	100.0	100.0	

Using the seven bins from the Churn and Profitability scores, each customer’s combined score is used to assign him or her to a quadrant. For example, a customer with a Profitability score of 7 and a Churn score of 5 gets a combined score of 75. Each two-digit score can be read as a set of X,Y coordinates.

Fig 6–Profitability/Churn Action Matrix

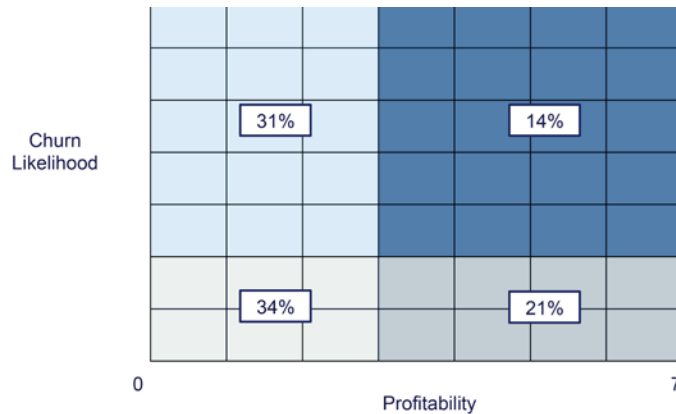


The quadrant boundaries were created from the company data set at break points driven by business needs. The Churn axis breaks between 2 and 3 where there is a 50% or more probability of churn. The Profitability axis breaks between 3 and 4 where the mean profitability was calculated. Figure 7 illustrates the breakdown of the quadrants and the customer distribution.

Within each quadrant, marketing efforts would be prioritized based on Expected Churn Loss:

$$\text{Expected Churn Loss} = \text{Monthly Expected Profit} * \text{Probability of Churn}$$

Fig 7–Profitability/Churn Quadrant Distribution



“...churn predictors could help determine how the company could retain a valued customer...”

As an example, this calculation could help determine if a Profit7-Churn6 individual should be addressed before a Profit6-Churn7 customer.

The actionable predictors of churn can now be used to develop specific marketing actions appropriate to each customer. Here are a couple of ways churn predictors could help determine how the company could retain a valued customer:

- **Subscriber has called customer service.** The customer service team would know (via integrated CRM) that this customer is a priority; likewise, during workload management for callbacks, profitable customers who are likely to churn would be at the top of the work queues.
- **Higher unique subscribers:** This would indicate that the customer does a lot of business and will regularly shop for volume discounts. The service provider could discount additional lines and offer to consolidate any lines the subscriber has with competitive carriers.
- **Children in household:** The carrier could offer discounted family plans to encourage children to join the plan instead of choosing a competing offer.
- **High credit rating:** Customers with high credit ratings can easily leave and are not locked into a plan. The service provider can offer early discounts to extend a 24-month commitment, a few months prior to plan expiration.

In conclusion, while Churn and Profitability/CLV models are not new, this unique quantitative approach of combining churn with the Profitability Factor is a new twist. **Through a simple scoring model similar to RFM, service provider marketers can now prioritize their retention actions for the best customers.**

EXHIBIT 1- CHURN SCORING VARIABLES

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	53178.922 ^a	.042	.056
34	53201.048 ^a	.041	.055

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 34 ^a	revenue	.00173783	.001	5.082	1	.024	1.002
	mou	-.00025644	.000	38.659	1	.000	1.000
	recchrge	-.00289557	.001	11.215	1	.001	.997
	overage	.00085752	.000	10.107	1	.001	1.001
	roam	.00688599	.002	11.818	1	.001	1.007
	changem	-.00049513	.000	88.434	1	.000	1.000
	changer	.00217132	.000	35.969	1	.000	1.002
	dropvce	.00795962	.002	24.849	1	.000	1.008
	blckvce	.00311351	.001	8.315	1	.004	1.003
	unansvce	.00083462	.000	4.113	1	.043	1.001
	custcare	-.00602921	.003	5.765	1	.016	.994
	threeway	-.03294184	.011	8.483	1	.004	.968
	outcalls	.00102671	.001	3.492	1	.062	1.001
	incalls	-.00293776	.001	8.980	1	.003	.997
	peakvce	-.00059728	.000	8.739	1	.003	.999
	months	-.01953363	.002	97.397	1	.000	.981
	unqsubs	.19068780	.020	91.896	1	.000	1.210
	actvsubs	-.21070478	.027	58.730	1	.000	.810
	phones	.05159926	.013	16.382	1	.000	1.053
	eqpdays	.00145299	.000	404.793	1	.000	1.001
	children	.10761678	.027	16.322	1	.000	1.114
	credita	.06878241	.030	5.154	1	.023	1.071
	creditb	.08967838	.034	6.871	1	.009	1.094
	creditc	-.11003307	.041	7.277	1	.007	.896
	creditd	-.30767517	.040	58.441	1	.000	.735
	prizmub	-.05371995	.022	5.749	1	.016	.948
	refurb	.23448797	.031	55.427	1	.000	1.264
	webcap	-.15690567	.037	17.705	1	.000	.855
	marryun	.09235302	.027	11.780	1	.001	1.097
	mailres	-.15549665	.025	37.408	1	.000	.856
	newcelly	-.06607626	.026	6.305	1	.012	.936
	income	-.01062298	.004	6.156	1	.013	.989
	creditad	-.06110406	.032	3.570	1	.059	.941
	setprcm	-.10878176	.040	7.383	1	.007	.897
	setprc	.00062148	.000	4.925	1	.026	1.001
	retcall	.72729224	.057	160.784	1	.000	2.069
	Constant	-.00796744	.081	.010	1	.921	.992