Kellogg School of Management
Northwestern University

S-Mobile: Predicting Customer Churn

Shu Ying Seng was a member of the first graduating class of the new Master's in Business Analytics program of the National University of Singapore (NUS). In contrast to many of her classmates who had little prior work experience, Shu Ying had worked for S-Mobile, a leading cellphone carrier in Singapore, for 7 years. Because she loved working there and because S-Mobile helped pay for her degree, she returned to the company 6 months ago.

Shu Ying's last job at S-Mobile before she attended NUS was to manage the retention desk at the company's call center. Her team's task was to persuade customers who called to leave S-Mobile to stay with the carrier instead. While Shu Ying's team could prove high "save" rates, she had always felt uneasy about this form of "reactive" churn management – she felt that it *trained* customers to threaten to leave in order to get discounts.

One of the key moments during her master's program was when Shu Ying discovered that customer analytics provided a compelling alternative to reactive churn management: Instead of waiting until a customer tried to leave, the company could be proactive and predict each customer's churn risk – before they even threatened to leave. This seemed like a much better approach because it allowed a company to act before customers were dissatisfied enough to want to leave, and retention offers had a much better chance of delighting customers. After all, how good could a retention offer look if it was given in response to a customer's threat to quit?

Given Shu Ying's experience in customer relations and the skills she acquired at NUS, she now managed an analytics team tasked with decreasing customer churn. She realized this was a big task. In the future, she might have to change the data that was collected, how customers were treated, what plans were available, etc.

Professor Florian Zettelmeyer prepared this case to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. The data used in the case was either publicly available or was simulated. Certain events and names are fictionalized for pedagogical purposes.

The first step, however, was to determine if existing data could be used to identify if some customers were more likely to churn than others. And if so, what marketing actions and offers could be used to reduce churn.

Shu Ying asked her team to pull data on a random sample of customers in order to build a predictive churn model. The response variable was whether a customer had churned in the last 30 days. The explanatory variables described customer characteristics and behavior over the 4 months preceding that period (i.e., the total time period covered in the data was 4 months + 30 days). See the data description at the end of this document.

The sample consists of three parts:

1. A training sample with 27,300 observations and a 50% churn rate ("training == 1")
2. A test sample with 11,700 observations and a 50% churn rate ("training == 0")
3. A representative sample with 30,000 observations and a churn rate of 2%, i.e., the actual monthly churn rate for S-mobile ("representative == 1")

The model would be used to generate churn predictions for 30,000 randomly chosen customers, i.e., the "representative sample" in the dataset, for whom Shu Ying had been authorized to evaluate the proactive churn management program.

To scale the predicted churn probabilities for use in the representative sample, the team could use (case) weights. However, not all models have an option to specify case weights or may even perform poorly when weights are applied during estimation. For this reason, a dataset with 1M observations is also available. Use the sample code in the starter notebook to access the data. Please do not include the 1M row data in your submissions.

The downside to using the dataset with 1M rows is, of course, that estimation time will increase substantially. You can use this larger dataset to re-estimate your chosen model and generate predictions for the representative sample.

The task

As Shu Ying briefed her team, she laid out what they would have to accomplish:

1. Develop a model to predict customer churn
2. Use model output to understand the main drivers of churn
3. Use insights on churn drivers to develop actions/offers/incentives
4. Quantify the impact of these actions/offers/incentives on the probability of churn
5. Decide which actions/offers/incentives to target to which customers
6. Evaluate the economics

Assignment guidelines

1. Develop a model to predict customer churn
   - Feel free to use any technique you like to predict churn. However, one of your models must be a logistic regression
   - Build models using the training data and explain your modeling choices

2. Use your model to describe the main drivers of churn and report on the key factors that predict customer churn and their relative importance.
   - Briefly discuss 5 key drivers of churn from your analysis in this step using Variable Importance (Permutation Importance) and Prediction or Partial Dependence plots

3. Use insights on churn drivers to develop actions/offers/incentives
   - Consider each of the 5 variable types, e.g., “Equipment characteristic”, “Customer usage”, etc. (see the data table at the end of this case). Discuss at least one variable from each group and propose action ideas that build on the plots from the previous question. In other words, what did you learn from the plots and how might you use those insights to reduce churn? No calculations are needed here, just your creativity. Feel free to discuss with ChatGPT to help come up with action ideas. If you use ChatGPT, please include your prompts here.

4. Quantify the impact of these (5) actions/offers/incentives on the probability of churn
   - Either (i) predict the effect of a churn driver (similar to what we did for Pentathlon NPTB) or (ii) suggest how you might set up an experiment (RCT) to evaluate the action/incentive/offer in the field
   - Generate predictions for the representative sample
   - Since it is not feasible to execute an RCT, describe how you would set up such an experiment and then make assumptions about the possible results and impact on churn that you can use in steps 5 and 6

5. Decide which of the 5 actions/offers/incentives to target to which customers
   - For each action/offer/incentive specify the criteria used to select customers. Will you apply the action/offer/incentive to all customers, or a subset? Motivate your approach
   -

6. Evaluate the economics (CLV):
   - For 3 actions/offers/incentives provide a comprehensive evaluation of the profitability implications using a 5-year (60 month) time window

Deliverables

1. Please upload a Jupyter notebook describing your work on each of the steps listed above through GitHub and GradeScope. The text in the report, excluding exhibits, should **not** be more than the equivalent of 3 single-spaced pages in Word (approx. 1500 words).

2. In addition to your Jupyter notebook with analysis and text, with your team, create a video presentation on the <u>top 3</u> actions/offers/incentives. At least one of these should be actionable at the retention stage. For each action/offer/incentive provide:
   - A description of the action/offer/incentive
   - An estimate of the impact of the action/offer/incentive on the probability of churn
   - The criteria used to select customers for the action/offer/incentive
   - The economic justification for the action/offer/incentive

   Target your presentation to the senior data scientist at the company and the chief product manager for S-mobile. You should explain the technical details of your work and provide context on the proposed next steps in decision making. Please use **S-mobile: Predicting Customer Churn (Video submission)** to upload the link to your group video on Panopto (10 points)

3. Generative AI (**5 points**): Describe **in detail** how your team used Generative AI-tools like ChatGPT to support your work on this case. Provide pdfs and/or screenshots of your "discussions" with these tools and comment on what things did and did not go well. Make sure to add discussion about your thought process and how you tried to maximize the benefits from using these tools. Also add any questions you may have about the assignment and the support you received from GenAI so we can discuss these topics in class.

   Note: No matter how you used Generative AI-tools, you are expected to fully understand all elements of the case solution submitted by your group. Any group member may be called on in class to walk us through your thought process and how different parts of your code work and how you arrived at your solution.

Hints:

- Check that the average predicted churn probability from your model for the representative sample is equal to 2%
- You may assume that S-mobile has 1 million subscribers. Other assumptions will be needed so please be explicit and provide a rationale.
- The key learnings from this case come from working through steps 2 - 6. Although I expect you to do a good job on step 1, I recommend you spend most of your time on steps 2 – 6.

If you have questions, please don't hesitate to post on Piazza or ask during a work session!

Data Description

variable	description	type	variable type
customer	Customer ID	Character	ID
churn	Did customer churn in last 30 days? (yes or no)	Factor	Response variable
changer	% change in revenue over the most recent 4 month period	Integer	Usage trend
changem	% change in minutes of use over the most recent 4 month period	Integer	Usage trend
revenue	Mean monthly revenue in SGD	Integer	Customer usage
mou	Mean monthly minutes of use	Integer	Customer usage
overage	Mean monthly overage minutes	Integer	Customer usage
roam	Mean number of roaming calls	Integer	Customer usage
conference	Mean number of conference calls	Integer	Customer usage
months	# of months the customer has had service	Integer	Customer usage
uniqsubs	Number of individuals listed on the account	Integer	Customer usage
custcare	Mean number of calls to customer care	Integer	Customer action
retcalls	Number of calls by the customers to the retention team	Integer	Customer action
dropvce	Mean number of dropped voice calls	Integer	Quality
eqpdays	Number of days customer has owned current handset	Integer	Equipment characteristic
refurb	Handset is refurbished (no or yes)	Factor	Equipment characteristic
smartphone	Handset is a smartphone (no or yes)	Factor	Equipment characteristic
creditr	High credit rating as opposed to medium or low (no or yes)	Factor	Customer characteristic
mcycle	Subscriber owns a motorcycle (no or yes)	Factor	Customer characteristic
car	Subscriber owns a car (no or yes)	Factor	Customer characteristic
travel	Has travelled internationally (no or yes)	Factor	Customer characteristic
region	Regions delineated by the 5 CDC Districts (e.g., CS is Central Singapore)	Factor	Customer characteristic
occupation	Categorical variable with 4 occupation levels	Factor	Customer characteristic
training	1 for training sample, 0 for validation sample, NA for representative sample	Integer	Sample selection
representative	1 for representative sample, 0 for training and validation sample	Integer	Sample selection