

2017 Travelers Case Competition

Kangaroo Auto Insurance Company Modeling Problem

October, 2017



1) Business problem

You work for Kangaroo Auto Insurance Company, an Australian company. Your business partner, who is not at all familiar with statistics, would like you to create a retention model based on historical policy data. Your business partner is concerned about the company's retention rates as well as the key drivers that cause policy cancellations.

For this case competition, your group is tasked with identifying policies that will cancel before the end of their term. With the help of a predictive model, your group will be able to predict those policies that are most likely to cancel as well as understand what variables are most influential in causing a policy cancellation.

2) Data Description

The Kangaroo data set is based on 4 years of property insurance policies from 2013 to 2016. There are roughly 7500 policies in the training data and each policy only has one observation. There are almost 2000 policies that were canceled during the effective term. In the test data, the cancelation indicator is NA. Your work is to build a model on the training data and use your best model to predict the cancelation indicator for each policy in test data.

Variable Descriptions:

<i>ID:</i>	Policy ID (cannot be used in model)
<i>tenure:</i>	Number of years with Kangaroo
<i>claim.ind:</i>	Occurrence of claim (0=no, 1=yes)
<i>n.adults:</i>	Number of adults in the property
<i>n.children:</i>	Number of children in the property
<i>ni.gender:</i>	Gender of policyholder
<i>ni.marital.status:</i>	Marital status of policyholder (0=no, 1=yes)
<i>premium:</i>	Price of the policy
<i>sales.channel:</i>	Medium through which policy was purchased
<i>coverage.type:</i>	Type of coverage
<i>dwelling.type:</i>	Type of dwelling
<i>len.at.res:</i>	Length at residence (how long policyholder lived at property)
<i>credit:</i>	Financial credit level of policyholder
<i>house.Color:</i>	Color of house
<i>ni.age:</i>	Age of policholder
<i>year:</i>	Year of the policy
<i>zip.code:</i>	Zip code of the property
<i>cancel:</i>	cancelation indicator (0=no, 1=yes). This is the response variable

3) Modeling

- Each group may have at most 5 people and will:
 - Work together within group on data analysis, but not between groups
 - Build a model to predict the retention rate and submit the predictions on the test data.
 - You can build the model with any software
 - Submit the prediction of test data as a csv file. Format of submission will be provided with data
 - Prepare presentation slides to summarize the prediction and analysis results. You do not need to explain the problem, just summarize what you did and what you found
 - Work with your group members and ask for clarification from the email at the end of this introduction

4) Model Evaluation

Models will be evaluated by c-statistics, which is a measure of how well your model does at rank ordering those most likely to cancel their policy. We will calculate your score once you submit your result. Please make sure the format is the same as the submission sample. R Package to calculate c-statistic will be provided.

5) Timeline

- October 13th – Deadline to register
- October 16th – Data set and business problem distributed via email
- October 27th – First model submission due (optional)
- November 3rd – Final model submission due
- Week of 11/6 – Group Presentations, exact date TBA

6) Benchmark Model

Score from a GBM model will be the competition benchmark. Any team scoring better than the benchmark score will be invited to present their model.

7) Presentation Instructions

- What methods did you consider (you don't have to actually try all of these methods; just ones that you think would work for this problem)?
- What method did you choose in the end, and why?
- How did you do your variable selection?
- What other variables not in the data set do you think might be useful?

8) Registration

Register by Friday October 13th using survey links below. Each individual should submit their own registration.

Downtown Campus: <https://pollev.com/surveys/JYIuhZBnq>

Storrs Campus: <https://pollev.com/surveys/qM4NnGPFs>

9) Contacts

Kathy Ziff: KZIFF@travelers.com

Yi Cao: YCAO@travelers.com

Tiran Chen: TCHEN3@travelers.com