

Hackathon November 2019: Severity Modeling Challenge

Background Severity Modeling

Severity modeling presents many challenges from a methodological standpoint. Severity modeling is seen in rating and RSP projects, where fitting parametric curves to severity distributions is a mandatory step towards smoothing the curve and appropriately modeling tail losses. However, the parametric and semiparametric modeling of severity distributions is challenging due to the limited number of observations and losses that tend to develop slowly over time.

Background RSP Ontario

The Risk Sharing Pool (RSP) is a program put in place by the Facility Association (FA). The general idea of the RSP is that each insurance company can transfer a limited number of vehicles they insure to an industry-wide pool where premiums and losses incurred from these vehicles are shared among all insurers. Only certain provinces have an RSP program and rules differ for each.

In 2017, the Co-operators decided to improve its approach with regard to the selection of vehicles for the RSP in Ontario by modeling the optimal premium and contrasting it against the actual premium. Risks with the largest difference between optimal and actual premiums are flagged as good candidates for the RSP pool. This approach has demonstrated significant improvement over the approach originally used to flag RSP candidates.

Problematic

The goal of identifying candidates for RSP requires us to model the optimal premium but in order to do so, the frequency and severity need to be modeled for each auto coverage considered under the RSP. Challenges to the modeling of severity are the relatively small number of claims among the insured and the skewed distribution of the claims (a few claims have very large values).

For rating purposes, most of the severity models that are used internally are GLMs that were developed in Earnix. In this hackathon, the objective is to model severity without imposing constraints on the methodological approaches and/or software, in order to see any potential gains in predictive accuracy.

Methods

1. Models: GLM, GAM, XGboost, Deep Neural Network, Linear Model, Lift Model...
2. Variables: You can use the variables provided and create or add new variables as you wish. For example, you could create new territories, integrate external data, consider the distance with the closest Coop office, the smoker status... The file "Description of Variables.xlsx" gives a brief description of a variable and another file, "IBC Vehicles Description mapping.pdf" details variables from IBC. Note that the VICC code is provided under "vehiclecode_cg_4" column.

Hackathon Description

1. You will be randomly assigned to teams prior to the beginning of the hackathon. Information is provided at the end of this document.
2. Two different severity datasets from RSP Ontario data will be provided – one containing data on Collisions (COL), the other containing data on Bodily Injury (BI).
3. The target variable is incurred_loss_m. Your models, one for each coverage, should be developed using the respective dataset that will be provided to you at the beginning of the hackathon. You can split them into train and test datasets, if desired.
 - a. RSP_ON_COL.csv
 - b. RSP_ON_BI.csv
4. Model Evaluation will be performed using the validation dataset.
 - a. Prediction files in .csv should be sent to gabrielle_vachon@cooperators.ca before Nov-14-2019 @ 23h59
5. The winner will be the team with the model that provides the best predictions on the validation dataset.
 - a. Metrics considered to determine the winning team are:
 - i. Root Mean Square Log Error (rMSLE)
 - ii. Mean Absolute Error (MAE)
 - b. Winners will be revealed on 15-Nov-2019. Eternal recognition from all BI colleagues is the ultimate prize!
6. Innovative solutions will be presented by teams during an upcoming lunch and learn. Details will be provided during the event.

Hackathon Teams

Team #1

Anne-Louise Ménard, Daniel Guerin, Nathan Rankin, Elyse Picard, Olivier Boivin

Team #2

Gabriel Alepin, Jérémie Boudreault, Etienne Chassé St-Laurent, François Laniel, Alexandre René

Team #3

Stephen Kane, Catherine Bernard, Maxime Lacroix, Sylvain Pannetier Lebeuf, Tomasz Rulka

Team #4

Guillaume Couture-Piché, Katia Charland, Luc Langevin, Ryan Dunlop, François Godbout

Team #5

Victor Lauzon, Antoine Djoufo, Eliot Gregoire, Kim Zhang, Guillaume Gautier

Team #6

Marion Grégoire-Duclos, Alvaro Guillen-Cuevas, Jing Xu, Idir Saidani, Frederick Guillot

Team #7

Veronika Gousseva, Jean-Mathieu Potvin, Paul Frosina, Joseph Mudge, Gabrielle Vachon

Team #8

Etienne Larrivée-Hardy, Jean-Sébastien Boutin, Steven Côté, Daniel Strongman, Alice Bergeron, Pénélope Ferron

