**Using Natural Language Processing to improve claims handling process.**

In 2017 the Data Science team at The General® was tasked with building a model to predict attorney involvement in the claims process. The model has evolved several times since the initial ask: beginning with a simple logistic model, evolving to more complex classification techniques, and in 2018 adding a text component. We have recently integrated the model into our claims system. After a three day lag period (necessary to gather more information about the claim and populate the text fields), the model identifies claims that are automatically sent for review. Currently, only about 60-65% of these claims are accepted (claims acceptability rate, or CAR) and we need your help to increase that to 75%.

We quickly realized that simply predicting whether or not a claim would eventually have attorney involvement was relatively straightforward, and not that helpful to the business. The challenging aspect was to predict when a specialized claims group (Non-Attorney Repped Bodily Injury, or NARBI) could mitigate the risk of representation and act before it occurred. Claims that become “repped” are expensive to the business: they take our adjusters more time with increased overhead and generally have higher paid losses. It’s also generally a disadvantage to our customers: it takes longer for their claims to resolve – a poor customer experience, and they owe a significant portion of the payment (30-35%) to their lawyer.

The NARBI team is a talented group of adjusters trained in working with customers to help them resolve their claims quickly and fairly, without getting an attorney involved. The team only takes on specific claims they believe their efforts will have an impact on. For example, if the injury is more severe – such as an injury included a fracture or worse – that claim will likely end up being litigated and the team won’t accept it. They also only take on claims where a bodily injury occurred; so if the claim only involves property damage, they won’t accept it. In our logic, we have many built in filters to account for this: the location of the claim, how many vehicles were involved, the severity of the injury, along with others.

That’s where you come in! We know there are factors highly indicative of if this team will accept a claim we are currently not accounting for, or don’t have a good method for predicting. Our primary question of interest is if we can predict who was at fault. Adding this logic to our model would go a long way to improving CAR. We are also interested in predicting loss cause, severity type, and (as part of a future “claims triage” project) what claim group the claim should be routed to. These variables are all labeled in the data set and defined in the data dictionary in DropBox.

You’ll notice the data set includes both claims and exposures. Think of claims as the “parents”, with exposures as the “children” – a claim can have multiple exposures, but not vice versa. These are labeled as such in the data. The NARBI team works on the entire claim – so even if just one of the exposures is flagged by the model, they’ll still work the entire claim.

**Common abbreviations in text fields**

* DNOP/VNOP – driver/vehicle not on policy
* CV/IV – claimed vehicle/insured vehicle

**Additional Resources**

AutoRegressed blog post on project:

http://autoregressed.com/2018/11/19/using-concept-detection-with-word-embeddings-to-improve-our-claims-handling-process/