Centre for
Operations Excellence

Predicting Claim Outcomes using Text Analytics

WorkSafeBC is dedicated to promoting health and safety for all workers of British Columbia. Part of their initiative to protect workers and employers is to provide no-fault insurance. In the event of any work-related injury, disease, or disability, workers can file a claim that WorkSafeBC will then review. Approved claimants are compensated for time off work and medical expenses associated with the work injury.

Problem and Opportunity

To properly manage these claims, the Claim Management System (CMS) was implemented in 2009. All information necessary to process a claim is stored in this system, including forms and letters provided by the worker, their employer, and medical practitioners. These documents have not been stored in accessible formats and so previous analyses have revolved around structured data coded in the system. This project aimed to determine if unstructured text data from claim forms and letters could improve the performance of existing models, which rely solely on structured data, in predicting claim severity.

Approach and Solution

Documents received within 14 days from registration were included in the training data, allowing enough information to be gathered but still having room to make early interventions. Relevant text within these documents was then extracted using Python, normalized through tokenization, stopword removal, n-gram training, and lemmatization, and finally converted to TF-IDF weighted vectors.

Using only the processed text data, predictive and topic models were trained. Predictive models commonly used in practice such as Naïve Bayes and Logistic Regression were explored to identify claims that were highly likely to become more complex than the rest. To validate model results and provide more tangible basis for action steps, topic modeling was used to find inherent themes and topics in the claim text that represent early indicators of claim complexity.

Benefits

At present, WorkSafeBC uses a set of business rules to allocate claims to different types of case officers. However, capitalizing on information obtained from claim forms and letters within a short time horizon showed great potential as it solely outperformed existing models predicting claim severity using structured data. This will enhance the structured data in supplementing the existing business rules by providing a stronger basis for early intervention and sound routing decisions. Having the right people handle the right claims as soon as possible can ultimately lead to a smoother, faster recovery for BC workers.

Value found in the text data can be further expanded by using more advanced parsing and learning techniques, and by integrating the text models into existing structured data models. Delivered software code for document models and processing pipeline will facilitate adoption and further development once text data sources are made accessible on an ongoing basis within WorkSafeBC.

UBC SAUDER SCHOOL OF BUSINESS
2053 Main Mall, Vancouver, BC V6T 1Z2
604.822.1800 | info@coe.ubc.ca | www.sauder.ubc.ca/coe