Paul-Erik Veel
416-865-2842
pveel@litigate.com

March 22, 2021

The SCC Leave Project: A Machine Learning Algorithm to Predict the Likelihood of Getting Leave to the Supreme Court of Canada

In the spring of 2020, we posted an analysis of the Supreme Court of Canada's 2019 Year in Review and summarized some of the statistics found in that document. Unfortunately, the general statistics found in the 2019 Year Review were high level and limited the analysis we could provide about what was happening at the Supreme Court of Canada. We also could not find any publicly available and current datasets that would allow us to provide a more detailed analysis.

So, under our newly launched Data-Driven Decisions program, we decided to create our own dataset, and that led to what we call the Supreme Court of Canada Leave Project. In brief, we built a database that contains information about every leave application decision released by the Supreme Court of Canada from January 1, 2018 onward. At the time of launch, this includes over 1,600 decisions, but the total number of cases increases each week as more leave decisions are released.

Our database captures a variety of variables pertaining to each of those decisions that fall into four main categories:

- Information about leave application decisions themselves (such as whether leave was granted and how long it took for the decision to be rendered);
- Information about the parties to leave applications;
- Information about the type of case; and
- Information about the underlying decision from which leave is being sought.

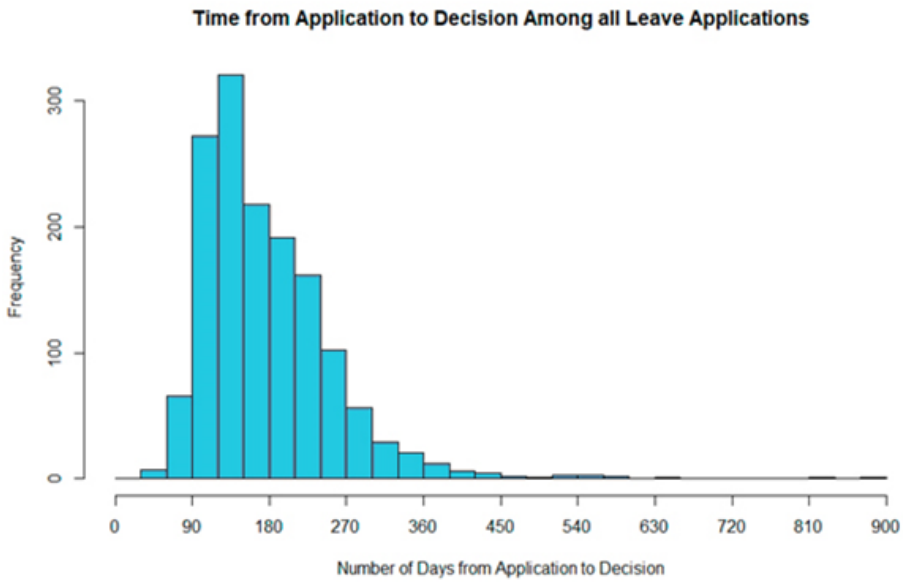
Key Uses of Our Dataset

Broadly speaking, we see this dataset as having three key uses.

First, it allows us to filter and quickly identify cases that meet specified criteria. For example, if you wanted to know every private law case in which a plaintiff sought and obtained leave

over the last 24 months, we could pull that report quickly. This helps us analyze cases where leave was successful or unsuccessful more quickly and can improve the quality of our applications for leave.

Second, this data gives us some quantitative insights about the process of getting leave to the Supreme Court of Canada. For example, clients sometimes want to know how long it takes from the time a leave application is brought to the decision, and we can easily generate data about how long that takes. We can also present it visually so that it's comparatively easy to understand. For example, below is a histogram that shows the length of time between the date of leave application and the date on which leave was decided for the cases in our database:



Again, with our dataset, we can slice and dice the data to provide statistics and visualizations pertaining to a range of variables of interest.

Third, and perhaps mostly importantly, this dataset has allowed us to develop a statistical model, based on machine learning techniques, that allows us to estimate the probability of any particular case being granted leave to the Supreme Court of Canada. In short, our model can tell you whether, extrapolating from relevant features of historical data, your case has a less than 1% chance of getting leave, a 15% chance of getting leave, or a 60% chance of getting leave. No statistical model will be a perfect representation of the real world; this will never replace our lawyers' expert judgment. However, it is an important complement. For a particular client considering whether to apply for leave, it might be that a 30% chance of getting leave is worth

it, while a 5% chance of leave is not. Our machine learning model can readily distinguish between those cases that have only a very small chance of leave being granted and those that have a moderate chance of leave being granted.

Upcoming Leave Predictions

Using our model, we will be providing our guesses as to the probability that particular cases will get leave, shortly before leave decisions come out. We'll use the model to help advise our clients deciding on whether to seek leave, but externally we'll also post the results of that model to provide a look at upcoming leave decisions. Keep watching this blog for our upcoming predictions.

If you're reading our blog and following the predictions of our model, there are three things we should flag at the outset.

First, our model is based extrapolating from previous data. An underlying assumption of any predictive model based on past data is that the data-generating process in the future is going to be roughly the same as what it was in the past. In our case, that means we assume that the Court is going to keep on approaching leave decisions like it has in the past. If the judges of the Court suddenly decide to approach leave applications very differently, all bets are off. We think this is unlikely, but it's worth flagging.

Second, whenever you're looking at probabilities, it's important to think about those correctly. Civil lawyers in particular are used to thinking about things on a balance of probabilities: on that standard, if something is 51% likely, then it happened, and if it's 49% likely, it didn't. But that's not how probabilities actually work. A leave application with a 60% probability of being granted will not always get leave, and 5% case will not always never get leave. Rather, on average, a 60% case will leave three out of five times, while a 5% case will get leave one out of 20 times. To put it differently, it's not an outlier for a 1% case to get leave every once in a while; it's only an outlier if 1% cases are getting leave a lot more frequently than the model would expect.

So what does that mean? It means that when you look at our model's predictions, it's important to understand what they mean and do not mean. The most likely prediction in any given week isn't always going to get leave, and low probability cases sometimes are going to get leave. That's just the nature of probabilistic models.

Third, as with any machine learning model, we expect that our model will improve over time as it includes more data and we make additional refinements.

With those caveats, we hope our model will be both interesting and useful. We're looking forward to further expanding the range of legal topics on which we can provide quantitative data analytics through our Data-Driven Decisions program.