6 The Supreme Court of Canada Leave Project

A Dataset and Machine Learning Model for Predicting Leave Application Outcomes

Paul-Erik Veel and Katie Glowach

6.1 Introduction

While much of this book is focused on appeals heard by the Supreme Court of Canada (SCC), most cases never get to the Supreme Court. The primary barrier is the leave application process. For the majority of cases, the Court must grant leave (permission) for cases to be heard, and the likelihood of getting leave in any case is low. In a typical year, the Supreme Court refuses to grant leave in 90% or more of cases in which it is sought.

In this chapter, we introduce Lenczner Slaght's Supreme Court of Canada Leave Project.[1] This project consists of a database that contains dozens of datapoints pertaining to every single leave application decided by the Supreme Court of Canada from January 1, 2018, onward. This dataset allows us to glean a variety of insights into the operation of the Supreme Court as an institution. We have also used this dataset to build a machine learning model – in essence, an algorithm that learns based on known training data, which is in turn used to make predictions regarding new data – to predict the likelihood of cases getting leave to the Supreme Court of Canada. While this type of tool is not commonly used in legal practice at present, which makes benchmarking performance difficult, this model has significant predictive utility and compares favourably to some commonly used screening tools in other practical domains such as medicine.

---

1 A more fulsome description of the project and other associated content can be found at https://litigate.com/data-driven-decisions.

6.2 An Overview of Seeking Leave to Supreme Court of Canada

In broad terms, the Supreme Court of Canada has jurisdiction over three kinds of cases: 1) federal references that go directly to the Supreme Court[2] (advisory opinions issued by the Supreme Court in response to particular questions posed by the federal government); 2) appeals in which there is an appeal "as of right" to the Supreme Court of Canada,[3] in which case a litigant can automatically bring an appeal to the Supreme Court, and 3) appeals in which the Supreme Court has granted leave to appeal.[4]

Cases to which the Court has granted leave are a particularly important subset of the Court's work, for several reasons. First, quantitatively, most cases heard by the Supreme Court of Canada are cases in which the Court has granted leave to appeal. Each year between 2012 and 2019, between 65% and 84% of the Supreme Court's docket consisted of cases in which the Court granted leave to appeal. The number of by leave cases heard by the Court dropped significantly in 2020 and 2021, likely as a consequence of the COVID-19 pandemic. But even in those years, by leave cases still represented over half the Supreme Court's docket.[5]

Second, cases in which the Supreme Court grants leave tend to be more jurisprudentially significant than cases that reach the Court as of right. For cases heard by leave, the Supreme Court has affirmatively decided, pursuant to the requirements of the *Supreme Court Act*, that the case raises sufficiently significant issues that it warrants being heard. By contrast, appeals by right lack any screening mechanism that would ensure that they raise jurisprudentially significant issues. Consistent with this, a recent study by one of the co-authors of this chapter found that most decisions from the bench – that is, decisions rendered without substantial reasons on the same day the appeal was heard – have been issued in decisions in which the Court heard the case as of right.[6] Consequently, the likelihood of the Court rendering full reasons that will ultimately

---

2 *Supreme Court Act*, RSC 1985, c S-26, s 53.

3 These include provincial references determined by provincial Courts of Appeal, as well as various criminal appeals. See *Supreme Court Act*, RSC 1985, c S-26, s 36; *Criminal Code*, RSC 1985, c C-46, ss 691–693.

4 *Supreme Court Act*, RSC 1985, c S-26, s 40. There are also provisions by which the provincial courts and the Federal Court of Appeal can grant leave for a case to be heard by the Supreme Court of Canada: *Supreme Court Act*, RSC 1985, c S-26, ss 37–37.1. However, this power is rarely invoked.

5 Supreme Court of Canada 2021 Year in Review, available at: https://www.scc-csc.ca/review-revue/2021/index-eng.html.

6 Alex Bogach, Jeremy Opolsky, and Paul-Erik Veel, "The Supreme Court of Canada's From-the-Bench Decisions", (2022) 106 SCLR (2d) 251–287.

make a case jurisprudentially significant is much higher for cases in which the Court has granted leave.

Finally, in most subject matters, effectively all the Court's jurisprudence consists of cases in which the Court has granted leave. For several decades, the Court's as of right jurisdiction has been limited to certain classes of criminal cases – those in which there is a dissent at the intermediate appellate court on a question of law, or those in which the intermediate appellate court has overturned an acquittal and substituted a conviction – as well as appeals of provincial references. Consequently, essentially every civil decision rendered by the Court over the last several decades occurred because the Supreme Court first granted leave.

The *Supreme Court Act* sets out the test for leave in s 40(1).[7] This provision contains two requirements for the Supreme Court to exercise its by leave jurisdiction. First, the decision below must be a decision of the "Federal Court of Appeal or of the highest court of final resort in a province". In most cases, this means either a provincial appellate court or the Federal Court of Appeal. However, in some cases in which there is no right of appeal to a provincial appellate court – for example, with respect to publication bans in criminal cases – leave may be sought from the decision of a provincial Superior Court.[8]

Second, the decision must be "by reason of its public importance or the importance of any issue of law or any issue of mixed law and fact involved in that question, one that ought to be decided by the Supreme Court or is, for any other reason, of such a nature or significance as to warrant decision by it". This is typically referred to as the "public importance" requirement. This requirement is the primary gatekeeper in most cases in which leave to appeal is sought to the Supreme Court of Canada. In practice, the "public importance" requirement means that the Court decides which cases it thinks are significant enough to warrant being heard. The Supreme Court does not typically issue reasons explaining why it did or did not grant leave to appeal.

Importantly, the Court grants leave in only a small minority of cases in which leave is sought. According to the Supreme Court's Year in Review 2021, in 2020 the Court granted 34 leave applications (approximately 8% of those filed) and dismissed 383 (approximately 92% of those filed). While the precise percentage varies from year to year, over the last five years the Supreme Court has consistently granted leave in less than 10%

---

7 *Supreme Court Act*, RSC 1985, c S-26, s 40(1).
8 *Dagenais v Canadian Broadcasting Corp.*, [1994] 3 SCR 835 at 858–862.

of cases in which leave has been sought.[9] Over the last ten years, the high-water mark for leave being granted was just under 13% in 2012.[10]

6.3 The Value of Quantitative Analysis of SCC Leave Applications

Supreme Court of Canada leave applications decisions are different from the appeal and reference decisions that form the subject of other chapters of this text, primarily because there are almost never reasons issued in leave applications. Consequently, the methods described in other chapters of this book, such as natural language processing tools or citation analysis, are not applicable to analysing Supreme Court of Canada leave application decisions. Instead, as described below, an analysis of Supreme Court leave applications must proceed on the basis of labelled leave application data.

Yet these same reasons also make a quantitative analysis of the Supreme Court's leave application process valuable. In appeals in which reasons are rendered, the possibilities of natural language processing and citation analysis compete with the conventional legal approach of closely reading a case. Such conventional reasoning will generally yield significant insights. Where a Court renders reasons for its decision, it provides a set of general principles that can be applied to future cases. Careful analysis through the conventional tools of legal reasoning can provide strong indications of how courts will rule in a range of future cases. While there will always be "hard" cases that may not be determined solely by reference to past decisions or other legal authorities, the existence of reasons will at the very least narrow the set of such "hard" cases.

By contrast, in leave applications, none of these methods work, because of the absence of reasons. Perhaps as a consequence, the leave process, and leave decisions have historically been understudied by the legal academy. A search of typical legal sources yields only a handful of academic articles pertaining to the leave process or outcomes.[11] Perhaps for similar

---

9 Supreme Court of Canada 2021 Year in Review, available at: https://www.scc-csc.ca/review-revue/2021/index-eng.html.

10 Supreme Court of Canada 2021 Year in Review, available at: https://www.scc-csc.ca/review-revue/2021/index-eng.html.

11 See, for example, Bertha Wilson, "Leave to Appeal to the Supreme Court of Canada" (1983) 4:1 Advoc Q 1; Geoff R. Hall, "Applications for Leave to Appeal: The Paramount Importance of Public Importance" (1999) 22:1 Advoc Q 87; Bruce Ryder and Taufiq Hashmani, "Managing Charter Equality Rights: The Supreme Court of Canada's Disposition of Leave to Appeal Applications in Section 15 Cases, 1989–2010" (2010) 51 SCLR (2d); Denise Cooney, "An Absence of Reason: Why the Supreme Court of Canada Should Justify Dismissing Applications for Leave to Appeal" (2012) 70:1 UT Fac L Rev

reasons, when legal practitioners have tried to predict whether cases will get leave, their predictions have been limited to vague and imprecise heuristics, such as the observation that cases that raise constitutional issues are relatively more likely to get leave.

As such, our collective knowledge regarding the leave process is limited. For example, Netolitzky has done valuable work exploring the difficulties that self-represented litigants face in getting leave to the Supreme Court.[12] Ryder and Hashmani found that, from the late 1990s onward, the rate of successful leave applications in cases raising equality rights under s 15 of the Charter had dropped, and governments generally had better prospects of obtaining leave than parties seeking to advance equality rights.[13] This prior work provides valuable insights into leave applications in very particular contexts, but it does not provide general insights into leave applications as a whole.

There is significant value in better understanding the leave application process and outcomes. As described above, the leave process is the gatekeeper for most of the Court's jurisprudence, and the vast majority of its jurisprudentially significant decisions. Understanding how and why cases come to be at the Supreme Court of Canada has implications for the evolution of that Court's jurisprudence. For practitioners, the stakes are also significant. Various factors may motivate clients to seek leave, such as an attempt to preserve their liberty, child protection or custody issues, or significant financial considerations. Yet the costs of seeking leave can be significant. While we are not aware of any study quantifying the costs of leave applications, they range from thousands to tens of thousands of dollars, depending on the particular practice area and lawyers' rates. However, the likelihood of success on a leave application is low, on average. Improving our understanding of what factors drive leave decisions could lead to better screening by lawyers in relation to whether it is worth seeking leave in a particular case. Screening out even a modest percentage of leave applications on the basis that they have no reasonable prospect of

---

41; Donald J. Netolitzky, "The Walking Wounded: Failure of Self-Represented Litigants in 2017 Supreme Court of Canada Leave to Appeal Applications" (2021) 58:4 Alta L Rev. 837; "Justice Suzanne Côté's Reputation as a Dissenter on the Supreme Court of Canada" SCLR: Osgoode's Annual Constitutional Cases Conference 88. (2018) (suggesting that the leave process may have changed as a result of Justice Côté).

12 Donald J. Netolitzky, "The Walking Wounded: Failure of Self-Represented Litigants in 2017 Supreme Court of Canada Leave to Appeal Applications" (2021) 58:4 Alta L Rev 837.

13 Bruce Ryder and Taufiq Hashmani, "Managing Charter Equality Rights: The Supreme Court of Canada's Disposition of Leave to Appeal Applications in Section 15 Cases, 1989–2010" (2010) 51 SCLR (2d).

success could save millions of dollars annually in expenses to litigants, let alone the effort and expense saved by the Court as an institution.

6.4 The Supreme Court of Canada Leave Project

We at Lenczner Slaght have sought to fill this gap in knowledge of leave applications with our Supreme Court of Canada Leave Project. The project consists of a hand-coded dataset of information pertaining to every leave application decided by the Supreme Court of Canada from January 1, 2018, onward.[14] Our database is updated with new cases as they are released each week.

In our database, we collect four categories of information about every leave application:[15]

1. Information about the parties – For example, the names of the parties, the number of applicants and respondents, information about what type of player each party is (e.g. individual, corporation, government, etc.), each party's role in the litigation (e.g. plaintiff vs defendant), and whether they were represented by counsel in their Supreme Court leave application.
2. Information about the case – For example, the general area of law into which the case falls, whether it is a class action, and whether the case is also proceeding as of right.[16]
3. Information about the lower court decision – For example, which Court decided the decision below, whether there was a dissent or concurrence at the court below, whether the intermediate appellate court overturned another lower court decision, whether there were

---

14 There is no particular significance to starting the dataset as of January 1, 2018. Our goal was that, before the project was first put to use in early 2021, we wanted to have at least 1,500 leave decisions in our dataset, including at least 100 cases in which leave was granted. We felt that this would provide a large enough set. We determined that we could meet those criteria by including all cases from January 1, 2018, onward. While, in general, more data is better, it is highly labour-intensive to build such a database manually, so we limited how far back we went.

15 All of the data in our database is obtained by a human review of publicly available sources. The sources consulted are: 1) the Supreme Court of Canada's news releases, published on https://scc-csc.lexum.com/; 2) case information from the Supreme Court of Canada's website at https://www.scc-csc.ca/; and 3) decisions of lower courts.

16 In the majority of appeals as of right, the Supreme Court's jurisdiction arises from the fact that the case is a criminal law matter where there was a dissenting judge at the Court of Appeal. However, in those cases, the Supreme Court will only hear an appeal as of right on the particular issue(s) on which a judge dissented at the Court of Appeal. Consequently, if an appellant who has an appeal as of right wishes to raise issues other than the issues on which a judge dissented, they must seek leave to appeal in respect of those issues.

interveners, how many judges decided the decision below, the length of the decision, the number of citations to other legal sources in the decisions, the number of days of argument, and the length of time it took to decide the case.

4. Information about the leave process pertaining to that case – For example, the date of the decision below, the date that leave was applied for, and the date the leave application was decided. These data points allow us to explore seasonality effects (e.g. whether the Court is more likely to grant leave in particular months) as well as autocorrelation effects (e.g. whether there is some temporal relationship between successful leave applications). These data points also allow us to generate other variables, such as the number of leave applications decided on any particular day.

As of September 1, 2022, our dataset contained information about dozens of different variables for 2,234 leave decisions.

6.5 Insights from the Data

We put that dataset to two broad uses. The first involved extracting general insights about the leave process and outcomes at the Supreme Court of Canada. As a simple example, we use this data to understand how long it takes for leave applications to be decided at the Supreme Court of Canada. Our data shows that in 2021, the median time from application to decision was 136 days (approximately four and a half months). This is one metric that can be helpful in evaluating the Court's performance over time and advising clients.

While just knowing a median is helpful, knowing the full distribution of timelines to decision is usually more helpful. Figure 6.1 depicts a histogram of the time from Supreme Court leave application to decision in 2021. The bin widths are 30 days: that is, each bar represents the number of cases decided within increasing 30-day periods from the date the leave application was filed. This shows that the distribution is wide. Figure 6.1

This data shows that it is not unusual for cases to be decided more quickly or more slowly. Indeed, 22% of leave applications decided in 2021 took more than six months to be decided, and 5% of leave applications took more than ten months to be decided. This data can be relevant to parties considering whether to pursue a leave application.

Going a step further, we also use our dataset to relate facts about the leave process to leave outcomes. For example, some Supreme Court watchers have commented on social media in relation to a phenomenon
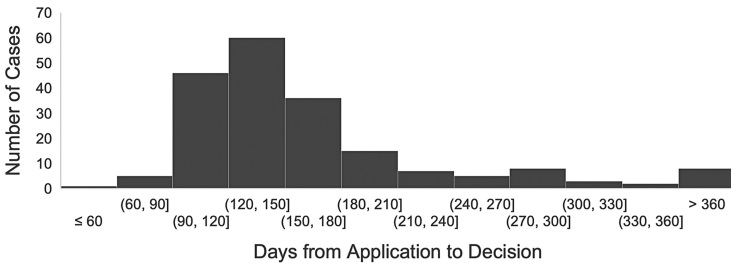
Figure 6.1 SCC leave applications - days from application to decision

at the Supreme Court that they have labelled "clearing the decks":[17] in essence, the Court has a tendency to decide, from time to time, a disproportionately large number of leave applications on a single day and to dismiss most or all of them. Our dataset allows us to empirically investigate whether this phenomenon does exist. Interestingly, it does: the likelihood of a randomly selected case being granted leave depends on the number of other leave applications being decided on the same day. To put it in concrete terms, over our entire dataset (up to September 1, 2022), leave applications decided on days on which ten or fewer cases were being decided had a 10.6% chance of being granted leave, while leave applications decided on days on which 20 or more cases were being decided had just a 4.7% chance of being granted leave.[18] Notably, this "clearing the decks" effect persists even after we control for other variables that impact the likelihood of leave being granted.

Finally, our dataset allows us to test whether certain factors impact the likelihood of cases getting leave to the Supreme Court of Canada. We had hypothesized that a variety of factors might be associated with a higher likelihood of getting leave and, indeed, we found that a number of strong associations were in line with our hypotheses. Without describing all of our findings in detail – which would be beyond the scope of this chapter – below are some of the main conclusions we reached.

First, we found that indicators of disagreements at the courts below were associated with an increased likelihood of leave being granted. Thus, the Court of Appeal allowing the appeal from the lower court's

17 See, for example, @jopolsky, February 24, 2022, "Big week with 14 leave applications. Another instance of clearing the decks?"; @jopolsky, March 17, 2022, "As predicted. 15 up. 15 down. Clearing the decks indeed".
18 Using a difference in proportion test, the difference between two is statistically significantly greater than 0 at the p<0.001 level.

decision, as well as the presence of dissents and concurrences at the Court of Appeal, were each associated with a higher probability of the Supreme Court granting leave. This is not surprising: a case in which different judges disagreed on the outcome is more likely to be associated with an underlying legal issue of public importance.

Second, indicators that the Court of Appeal perceived a case as having increased importance or complexity were also associated with an increased likelihood of the Supreme Court granting leave. For example, we found that, all else being equal, longer reasons, more citations to other sources, and a longer period of time under reserve at the Court of Appeal were each associated with an increased likelihood of the Supreme Court granting leave.

Third, we had speculated that in criminal cases, the Supreme Court was more likely to grant leave in cases in which the Crown seeks leave, rather than in cases in which the accused seeks leave.[19] This hypothesis turned out to be well founded: Crown applications for leave were more likely to be granted than applications by criminal defendants, and this effect persisted and remained significant even after controlling for other variables.

Finally, and unsurprisingly, the self-representation of an applicant in a leave application was strongly associated with leave not being granted. Indeed, in our entire dataset between January 1, 2018, and September 1, 2022, the Supreme Court did not grant leave in a single case in which the applicant was self-represented on the leave application.

Yet not all of our hypotheses turned out to be well founded. First, many Supreme Court watchers would likely predict that it is easier to get leave in cases that raise certain types of legal issues (for example, constitutional law cases). We have found little evidence that area of law by itself, after controlling for other factors, has a substantial impact on the likelihood of getting leave.[20]

---

19 There are several different reasons why this might be true. It may be that the Supreme Court of Canada more seriously entertains requests from the Crown to hear cases than it does from criminal defendants. It may be that the Crown is more selective in seeking leave – that is, it tends to seek leave relatively more often in those cases that are likely to get leave. Or it may be that cases in which the Crown seeks leave somehow otherwise implicate the public interest to a greater extent on average than do cases in which defendants seek leave.

20 There may be more complex relationships between the nature of the legal issues at play in the case and the likelihood of getting leave. Indeed, our current black-box machine learning model for predicting the likelihood of leave applications being granted, described below, does include some features relating to the legal issue at play in a case. This suggests that area of law does factor into the likelihood of a case getting leave, but not in a simple way.

Second, we had hypothesized that class action cases were more likely to get leave. Class actions tend to be large cases – both in terms of the quantums at stake and the number of persons affected. Moreover, they often raise complex procedural and substantive issues. Consequently, it seemed reasonable that they would be relatively more likely to get leave to appeal. However, we found no evidence that class actions were statistically significantly more likely to get leave to the Supreme Court of Canada than other types of cases, after controlling for other factors.

Finally, we had conjectured that it might be easier to obtain leave to appeal from decisions of some provincial courts of appeal than others. For example, there are three judges from Ontario and three from Quebec on the Supreme Court of Canada, and we might have therefore expected that they would form a critical mass of judges who would have a greater interest in cases originating in those provinces. However, we have found no evidence to date of any difference in the likelihood of leaving being granted based on the province of origin of a case.

6.6 Our Supreme Court of Canada Leave Prediction Model

While these insights help us understand the Supreme Court institutionally, the second use to which we put our dataset is more practical: predicting the likelihood that particular cases will get leave to the Supreme Court of Canada. A well-performing model for predicting the likelihood that particular cases will get leave has significant potential utility in practice. Such a model could be used as a first pass screening mechanism: that is, it could identify a subset of cases with a sufficiently low chance of getting leave that no further analysis would be warranted. Such a model could also provide a second opinion function. If a lawyer judges that a case has a high likelihood of being granted leave, but the model predicts a low probability, that may lead one to investigate further to decide whether a leave application is worth it. By contrast, if the predictive model is aligned with the lawyer's opinion, that is a datapoint that supports the reasonableness of the lawyer's opinion. Finally, a model can provide a quantitative prediction. While some clients may want a probability estimate of their chances of getting leave, lawyers may have difficulty formulating their advice in probabilistic terms. A predictive model could help ameliorate that situation.

With those goals in mind, we built a machine learning model that learns from the existing data in the dataset to predict the probabilities of new cases getting leave. A machine learning model is, at its most general, any algorithm that builds a model by learning on the basis of pre-existing training data, which is then used to make predictions regarding other test data.

Machine learning models can take various forms. Relatively simple statistical models, such as ordinary least squares regression (for continuous data) and logistic regression models (for binary outcomes), can be considered machine learning models when built on the basis of a set of training data and then applied to test data. Such models are simple and easily interpretable, because there is some form of linear relationship between the input variables and the output. Readers will have no doubt seen the simplest form of linear regression, which is simply a straight line drawn through a cloud of data points in a way that best approximates a relationship between two variables. On the other end of the spectrum are blackbox models. These are machine learning models that may make more accurate predictions, but at the expense of interpretability; the complex relationships of the variables and how they relate to each other are determined by an algorithm and cannot be understood by humans, even those who design the models.[21]

The outcomes of leave applications are a binary variable: a party either gets leave, or it does not. Machine learning models that predict what category a particular case will fall into are called classification models. However, in our model, we do not predict the binary outcome of whether or not a case will get leave. Rather, the output of our model is the *probability* that a particular case will get leave. We believe probability predictions are more useful than a simple yes/no prediction, as probabilities can be more easily combined with lawyers' judgments to render an overall assessment of the likelihood that a case will get leave.

When we present such data publicly, we often group cases into four categories:

1. Cases to Watch – These are cases in which our model predicts a greater than 25% chance that leave will be granted. The chance that these cases will be granted leave is much better than average. While cases included in this category will not all get leave, they are worth watching as strong candidates.
2. Possible Contenders – These are cases in which our model predicts between a 5% and 25% chance that leave will be granted. These cases have an average to somewhat above-average chance of getting leave.

---

21 Full details of such black-box models are beyond the scope of this chapter, but examples include Random Forest models (which attempt to aggregate the results of hundreds or thousands of randomly constructed decision trees on subsets of the training data) and neural network models (which are inspired by the information-processing approach used by neurons in the human brain). For a fulsome discussion of various machine learning algorithms, see Max Kuhn and Kjell Johnson, *Applied Predictive Modeling* (New York: Spring Science+Business Media, 2013).

While most cases in this category will not get leave, we expect to see a healthy minority of cases in this category being granted leave.

3. Unlikely Contenders – These are cases in which our model predicts between a 1% and 5% chance that the case will get leave. The safe bet is against leave being granted in these cases, but we do expect to see leave being granted from time to time.

4. Long-Shots – These are cases in which our model predicts a less than 1% chance that the case will get leave. Although it will happen from time to time, it would be a significant outlier for our model for these cases to be granted leave.

While the particular probabilities that form the boundaries of these categories are arbitrary, such categories can be helpful in guiding lawyers' reasoning. For example, it would not be unreasonable for a lawyer to apply a heuristic that they presumptively recommend seeking leave to appeal in any Cases to Watch, while presumptively recommending against seeking leave to appeal in any Long-Shots. Cases falling into the intermediate categories may be worth further consideration, taking into account the lawyer's own judgment of the likelihood of success as well as the stakes of the case.

We have applied different machine learning models to our leave application data at different times. For example, the primary model we used for predictions between October 2021 and March 2022 was a logistic regression model with 11 independent variables. This model performed reasonably well in making predictions. Moreover, because of the ease of conventional statistical analysis with a logistic regression model, we were able to gain insights into the factors that impact the likelihood of leave being granted, including what factors are statistically significant and what the effect size is. That model helped us identify several factors that were statistically significantly associated with getting leave to the Supreme Court, many of which are described above.

However, as of April 2022, we began to use a gradient-boosting model for our predictions, as implemented in the R package *xgboost*. *Xgboost* is a commonly used open-source machine learning package whose powerful performance in a variety of contexts has been recognized.[22] Our current model (as of September 1, 2022) is trained on 16 variables in the dataset. Based just on those 16 variables, our model performs strongly.

The "performance" of machine learning models can be difficult to assess in an intuitive manner. This is particularly true when the variable of

---

22 In brief, *xgboost* is an ensemble learning method that uses an iterative series of decision trees in a way that can take a set of individually weak variables and combine them into a strong model overall.

interest is a binary variable that is very unbalanced (that is, one outcome occurs substantially more often than the other). In such cases, of which Supreme Court of Canada leave decisions are an example, accuracy (or how often a prediction is correct) can be an unhelpful measure. For example, if only 8% of cases get leave, a model that predicts that no case will ever get leave is accurate 92% of the time. While that sounds like a very accurate model, that metric is essentially useless in practice.

A better way of thinking about model performance in those circumstances is to consider the trade-off in a model between sensitivity and specificity. Sensitivity refers to the true positive rate: how often the model correctly identifies a case as likely to get leave, when it actually does get leave. Specificity refers to the true negative rate: how often the model correctly identifies a case as not likely to get leave, when it in fact does not get leave. There is invariably a trade-off between sensitivity and specificity in any given model: the more true positives the model correctly predicts, the more false positives it will likely predict as well.

One measure of the performance of a machine learning model in classification problems that combines both of these measures is known as the area under the curve (AUC). As the name suggests, the AUC measures the area under a particular curve, the Receiver Operator Characteristic (ROC) curve. The ROC curve provides a plot of a model's sensitivity vs 1 minus sensitivity at every threshold value. The AUC of a model ranges between 0.5 and 1, with values closer to 1 indicating a better model. An AUC of 0.5 indicates that the model has no predictive power; put simply, it is no better at predicting than flipping a weighted coin. By contrast, an AUC of 1.0 indicates a perfect model with both 100% sensitivity and 100% specificity.[23] An AUC of 1.0 is impossible in any real-world context. However, the classical description holds that an AUC between 0.7 and 0.8 is acceptable, an AUC between 0.8 and 0.9 is excellent, and an AUC above 0.9 is outstanding.[24] By way of comparison, the Framingham risk score, a tool used in medicine for predicting the likelihood of certain cardiovascular events, has been reported to have a real-world AUC of between 0.62 and 0.78.[25]

As of September 2022, our predictive model, built using *xgboost*, when trained on 70% of the data (randomly selected) and then used to predict the remaining 30%, has an AUC in the range of 0.85. This shows that our

---

23 In contrast to a naïve measure like accuracy, ROC curves and AUC curves are insensitive to disparities in the proportions of the different classes: Max Kuhn and Kjell Johnson, *Applied Predictive Modeling* (New York: Spring Science+Business Media, 2013) at 264.

24 D.W. Hosmer and S. Lemeshow, Chapter 5, *Applied Logistic Regression*, 2nd ed. (New York: John Wiley and Sons, 2000), pp 160–164.

25 See, for example, Asaf Bitton and Thomas Gaziano, "The Framingham Heart Study's Impact on Global Risk Assessment" Prog Cardiovasc Dis. 2010; 53(1): 68–78.

machine learning classification model has significant utility in predicting the outcomes of leave applications. The strong performance of our model is not based on being able to directly measure whether a case raises issues of public importance. Rather, our model works because we have successfully identified a range of factors that correlate, to various degrees, with cases that the Supreme Court considers to be of public importance.

Our model is not, and never will be, perfect in predicting outcomes, for several reasons. First, our model does not incorporate all the data that is important in determining the outcome of leave applications. For example, our model does not presently incorporate any information about the quality of the leave application; assuming that the quality of lawyering matters even slightly to the likelihood of leave being granted, our model does not take this into account. We hope to be able to expand our dataset in the future to include such information. Second, even with access to every conceivable piece of information, our model would still almost certainly not be perfect, because it is modelling the behaviour of a small number of human beings. While the social sciences have shown us that human behaviour demonstrates regular patterns that can be modelled, such models will never be perfect. The fact that a model is not perfect does not mean it is not useful. In this case, there is good evidence that our model performs well enough to be a useful complement to lawyers' predictions.

6.7 Conclusion

The standard for being granted leave to appeal to the Supreme Court of Canada is the nebulous "public importance" standard. As a discretionary standard applied by a Court that does not provide any reasons for its decision, the leave application decision process is opaque. Yet that does not mean that leave application decision-making is devoid of patterns or completely unpredictable. On the contrary, our dataset has allowed us to identify several factors that are statistically correlated with getting leave to the Supreme Court. Our machine learning model, using just 16 variables, provides sufficient predictive insights to be useful as either a screening tool or a second opinion. We hope to improve the model's performance in the future through the collection of additional data.