Predicting Supreme Court Decision Making
Data Mining Final Project Proposal
Monica Rajendiran (mkr2ff)

The Supreme Court in its current form exists as the pinnacle of federal judicial decision making, the final interpreter of the United States Constitution. It has ultimate appellate jurisdiction, meaning that it has the ability to review and change the decisions made by lower courts. In its usual form, the court consists of a Chief Justice that presides over 8 other justices that are categorized as conservative, liberal, or moderate in their interpretation of the law and Constitution. Justices are nominated by the President and confirmed by the Senate but they are expected to be independent and nonpartisan in their interpretation of the Constitution and their subsequent passing of judgements on cases that bubble up to them from the lower courts [1].

The decisions of Supreme Court cases have been analyzed by legal experts, scholars, and hobbyists for decades but recently their decisions are also coming under the lens of applied machine learning and statistical modeling techniques. This is a challenge because the cases that Supreme Court justices get are often unclear and complex in light of Constitutional Law. On top of that, the issues are handled by human justices with deep, unique, and often evolving ideologies and interpretations which can be hard to quantify.

Although Supreme Court justices are expected to be independent in their decisions, it is clear from legal and historical analysis that external influences that can potentially be quantified affect the justice's decisions. There are several studies that focus on studying the Supreme Court and attempting to predict the voting behavior of the Court in any given year with a given case. For example, a law professor from the University of Pennsylvania, along with a team of legal scholars and political scientists, undertook a study to predict cases in the 2002 - 2003 Supreme Court term using classification trees based on ideological preferences and past voting history [2]. Another study looked at six decades of Supreme Court votes, variables, and cases to train a random forest model for generalized Supreme Court predictions [3]. These two studies boast an accuracy rate between 70 and 75% but interestingly, a hobbyist from Queens, New York and a legal correspondent at New York Times and Yale Law School claim to accuracy rates just as high without the use of a statistical model, instead relying on their memories of justices' past decisions, opinions, and oral arguments [4]. Jacob Berlove, the hobbyist from Queens, has correctly predicted Supreme Court cases for the past 3 years, more than 80% of the time. The grounds for predicting Supreme Court Decisions are still evolving and current statistical models rely solely on variables that arise out of the Supreme Court such as previous voting history and justice demographics and ideologies.

I propose the study and creation of a statistical model that uses many of the variables that existing Supreme Court predictors use, with the addition of analysis and inclusion of variables from the political environment, including the makeup of Congress, the party of the Presidency, and public opinions on certain issues. I would like to also include variables related to metrics that define the state of the country at a particular court term, such as economic, social, or foreign policy. I would also like to include an analysis on Supreme Court oral arguments and include them into my statistical model through text mining. Analysis will in general reveal what variables work and do not work in predicting Supreme Court decisions, but I intend to try and include a breadth of factors that have not already been considered (e.g.

oral arguments and political environment factors). The following is a table of potential predictors and responses I intend to model.

Predicting Supreme Court Decisions **Predictors** (List is not Exhaustive)	
Predictor Topic	Example
Justices Personal and Professional Information	Justice's Alma Matter, Justice's Previous Political Leaning
Justices Previous Voting Decisions	Justice's Voting History on Specific Case Topics (e.g. "3 affirmations for liberal freedom of speech cases)
Case History	Lower Court's Ideology
Oral Argument	Key Words using TD-IDF
Political Environment	Makeup of Congress (% Liberal vs. Conservative), Party in Presidency
Economic Environment	GDP, Employment Numbers, Recession Status
Predicting Supreme Court Decision **Responses** (List is not Exhaustive)	
Response	Description
Case Decision (Affirm/Reverse)	Binary Response
Voting Ratio (Number of Votes to Affirm vs. Number of Votes to Reverse	Numerical Response

I intend to primarily use classification trees and Random Forests as methods for predicting Supreme Court decisions. As I build out my data exploration and modeling, I would like to explore other methodologies that could successfully be used. I intend on creating a generalizable prediction model so my evaluation metrics will be tested on chronologically different Supreme Court terms and cases, but the 2016 Supreme Court term will be the primary validation set.

I am interested in this problem because I have a strong interest in the laws that we have created, how they are enforced and interpreted, and how they impact the country, especially those laws that pertain to social or environmental issues. I see the ability to successfully automatically predict court cases as one that can lead to a changed interpretation of how legal experts, scholars, and justices themselves see the law and the effect that human biases have on it.

References

1. Hamilton, Eric. "Politicizing the Supreme Court." *Stanford Law Review Online*65 (2012): 35.
2. Ruger, Theodore W., et al. "The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking." *Columbia Law Review* (2004): 1150-1210.
3. Katz, Daniel Martin, Michael James Bommarito, and Josh Blackman. "Predicting the behavior of the supreme court of the united states: A general approach." *Available at SSRN 2463244* (2014).
4. Greenhouse, Linda. "Press room predictions." *Perspectives on Politics* 2.04 (2004): 781-784.