

## MITOCW | MIT15\_071S17\_Session\_4.2.01\_300k

---

In this lecture, we'll see how analytics can be used to predict the outcomes of cases in the United States Supreme Court.

This seems like a very unconventional use of analytics, but in 2002 a group of political science and law academics decided to test if a model can do better than a group of experts at predicting the decisions of the Supreme Court.

In this case, a very interpretable analytics method was used, called classification and regression trees.

The legal system of the United States operates at the state level and at the federal or country-wide level.

The federal level is necessary to deal with cases beyond the scope of state law, like disputes between states, and violations of federal laws.

The federal court is divided into three levels-- district courts, circuit courts, and the Supreme Court.

Cases start at the district courts, where an initial decision is made about the case.

The circuit courts hear appeals from the district courts, and can change the decision that was made.

The Supreme Court is the highest level in the American legal system and makes the final decision on cases.

The Supreme Court of the United States consists of nine judges, or justices, who are appointed by the President.

This image shows the nine Supreme Court justices from the time period 1994 through 2005.

This was the longest period of time with the same set of justices in over 180 years.

The people appointed as Supreme Court justices are usually distinguished judges, professors of law, or state or federal attorneys.

The Supreme Court of the United States, or SCOTUS, decides on the most difficult and controversial cases in the United States.

These cases often involve an interpretation of the Constitution, and have significant social, political, and economic consequences.

There have been many significant and groundbreaking decisions made by the Supreme Court.

These are a few notable decisions that were made.

In 1942, the Supreme Court decided on the *Wickard v. Filburn* case.

This case recognized the power of the federal government to regulate economic activity.

Filburn was a farmer, who was growing wheat for on-farm consumption.

However, the US had established limits on wheat production, and Filburn was exceeding those limits.

So even though the extra wheat he was producing was for his own use and he had no intention of selling it, he was forced to destroy it.

In 1973, the Supreme Court decided on the Roe v. Wade case, one of the most well-known cases to this day.

They decided to legalize abortion, and by doing this, prompted a national debate that continues today about the legality of abortion.

In 2000, the Supreme Court actually decided the outcome of the presidential election.

The race was so close in the state of Florida, that a recount of the ballots was required.

But the Florida Secretary of State certified that President Bush was the winner before the recount could be completed.

The case then went to the Supreme Court where it was ruled that all ballots needed to be recounted.

But since this could not be done before the winner had to be declared, President Bush won the state of Florida, and thus, the presidency.

A very recent case from 2012 dealt with the Patient Protection and Affordable Care Act, commonly called ObamaCare, which requires most Americans to have health insurance.

The Supreme Court upheld this requirement.

Since non-profits, voters, and anybody interested in long-term planning can benefit from knowing the outcomes of the Supreme Court cases before they happen, legal academics and political scientists regularly make predictions of Supreme Court decisions from detailed studies of the cases and individual justices.

In 2002, Andrew Martin, a professor of political science at Washington University in St.

Louis, decided to instead predict decisions using a statistical model built from data.

Together with his colleagues, he decided to test the model against a panel of experts.

They wanted to see if an analytical model could outperform the expertise and intuition of a large group of experts.

Martin used a method called classification and regression trees, or CART.

In this case, the outcome is binary.

Will the Supreme Court affirm the case or reject the case?

He could have used logistic regression for this, but logistic regression models are not easily interpretable.

The model coefficients in logistic regression indicate the importance and relative effect variables, but do not give a simple explanation of how a decision is made.

In this lecture, we'll discuss the method of CART, and a related method called random forests.

We will then see if these methods can actually outperform experts in predicting the outcome of Supreme Court cases.