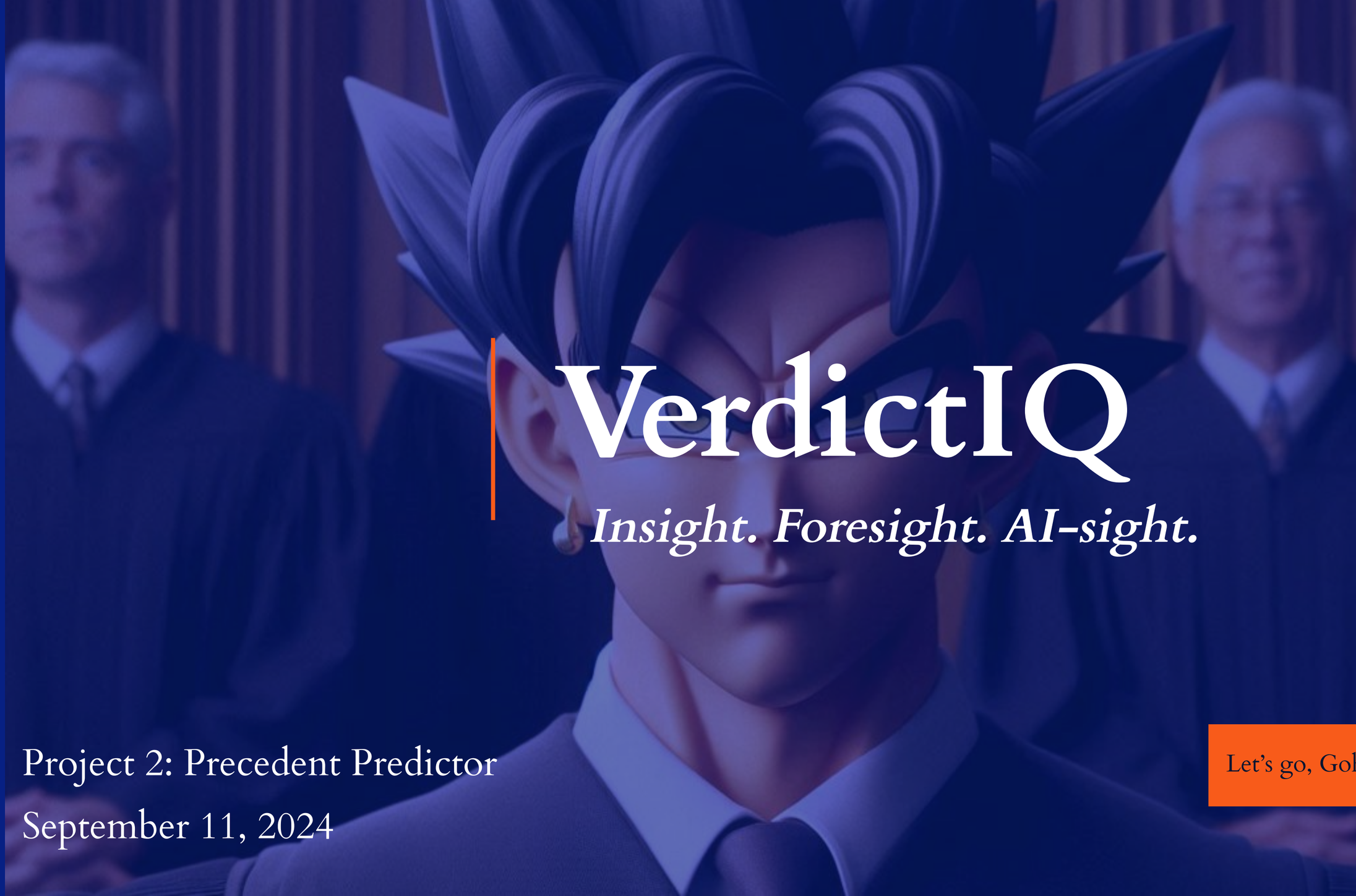


VerdictIQ

Insight. Foresight. AI-sight.

Project 2: Precedent Predictor

September 11, 2024

Let's go, Gohan!

VerdictIQ Founders

Dhruba Chakrabarti

Dipesh Pandya

Alexis Wukich

Executive Summary

Our Problem

- How can we demonstrate that the legal industry can benefit from machine learning?
- Can we use historical case data to train an algorithm to accurately predict an outcome of a U.S. Supreme Court case?”

Our Approach

1. Initial Clean-up and Exploration of our Dataset
2. Preprocessing our Data
3. Streamlining our Process: Training, Testing, and Evaluating Multiple Models and Multiple Variables
4. Using our Models to Predict Outcomes

Challenges & Opportunities

1. Starting with a clean dataset is a huge advantage
2. Multiclass data can be very difficult to manage, especially with the increased number of classes
3. There are endless opportunities to develop and grow with the data

Our Dataset

- The Supreme Court Database from Washington University Law
- Legacy Database contains cases from 1791 to 2022
- **Modern Database** contains cases from 1946 to 2022
  - optimized for data analysis
  - ~14,000 rows of data, 50+ columns
  - Most variables' values were coded using a well documented, variable-specific numerical scoring system
  - Started with a very clean, up-to-date, and well-maintained database

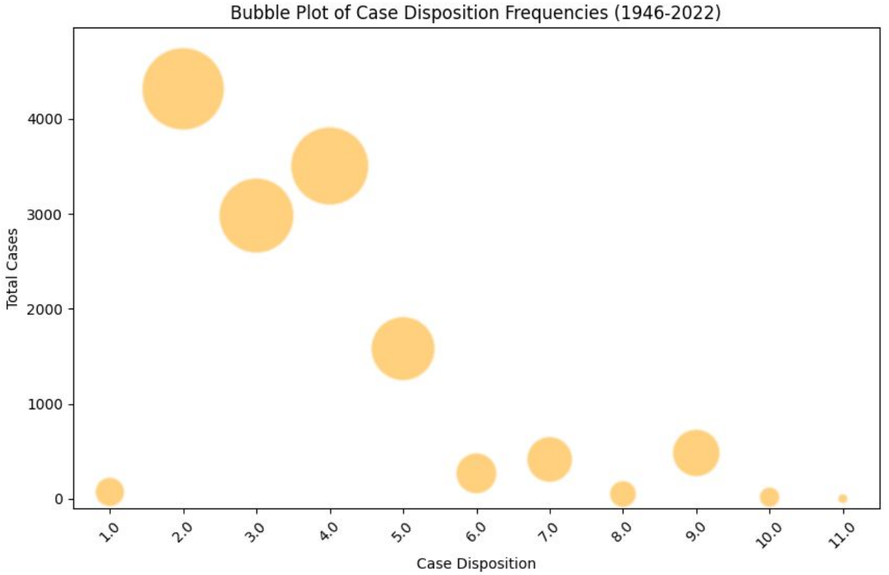
Primary Goals of Clean-up & Exploration

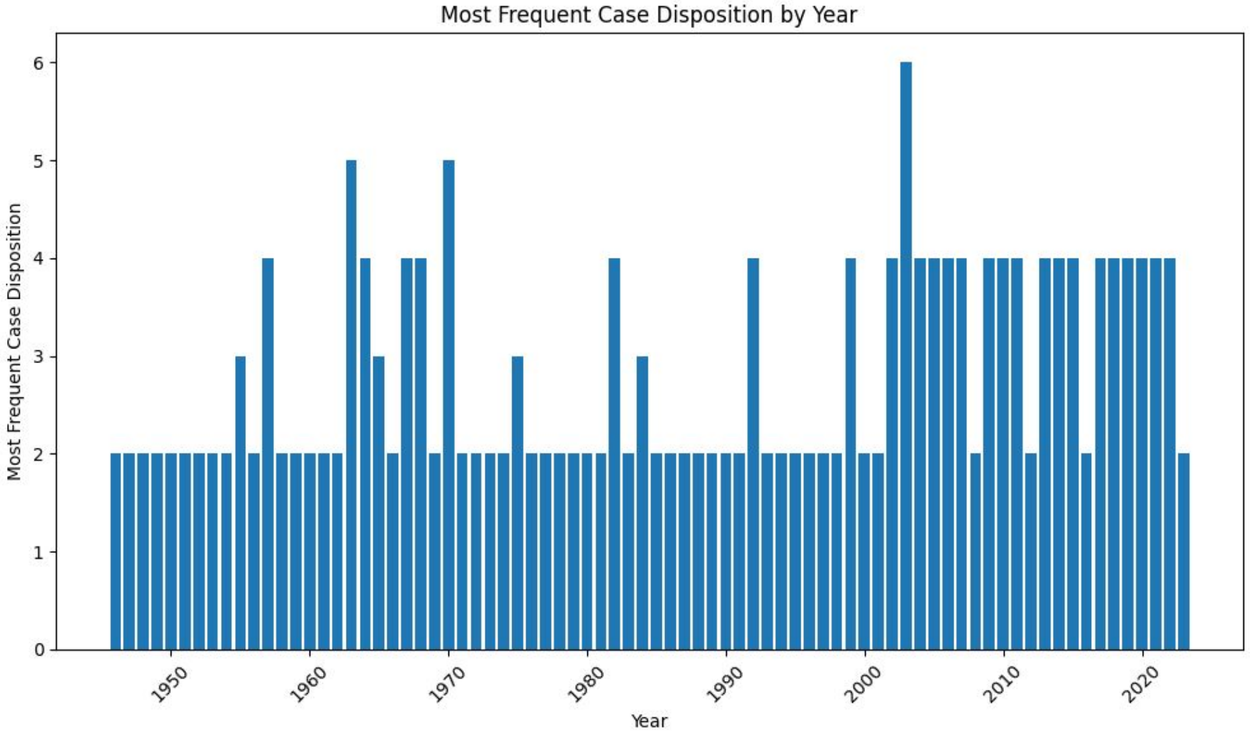
- identifying our **target "outcome" variable**
- feature reduction
- create a methodology for NaN values

Exploration Question #1:

What is the most common case disposition?

>**Answer**: Using this bubble plot we determined that “2”, or “affirmed”, followed by “4” and “3”, “reversed and remanded” and “remanded”, respectively, were the most common disposition from the Supreme Court from the 1946 term to the 2022 term.





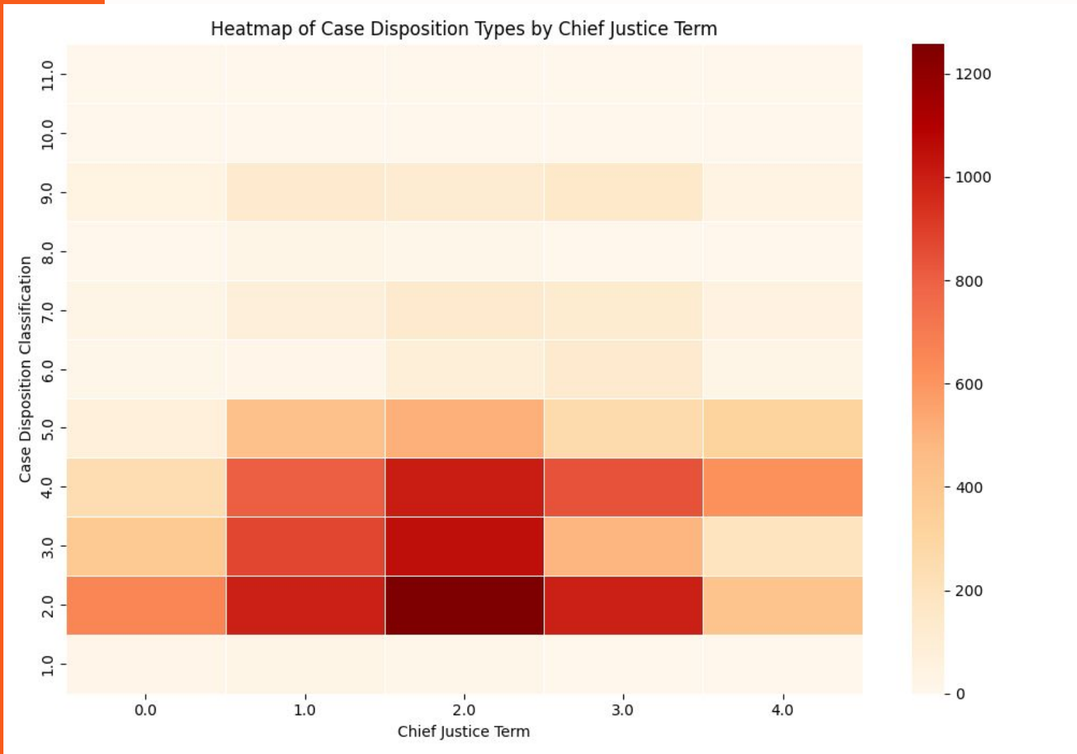
Exploration Question #2:

Have case dispositions changed over time?

>**Answer**: Using this bar graph we determined that “2” (“affirmed”), followed by “4” (“reversed and remanded”) were the case dispositions that appeared most frequently from the 1946 term to the 2022 term.

Exploration Question #3:

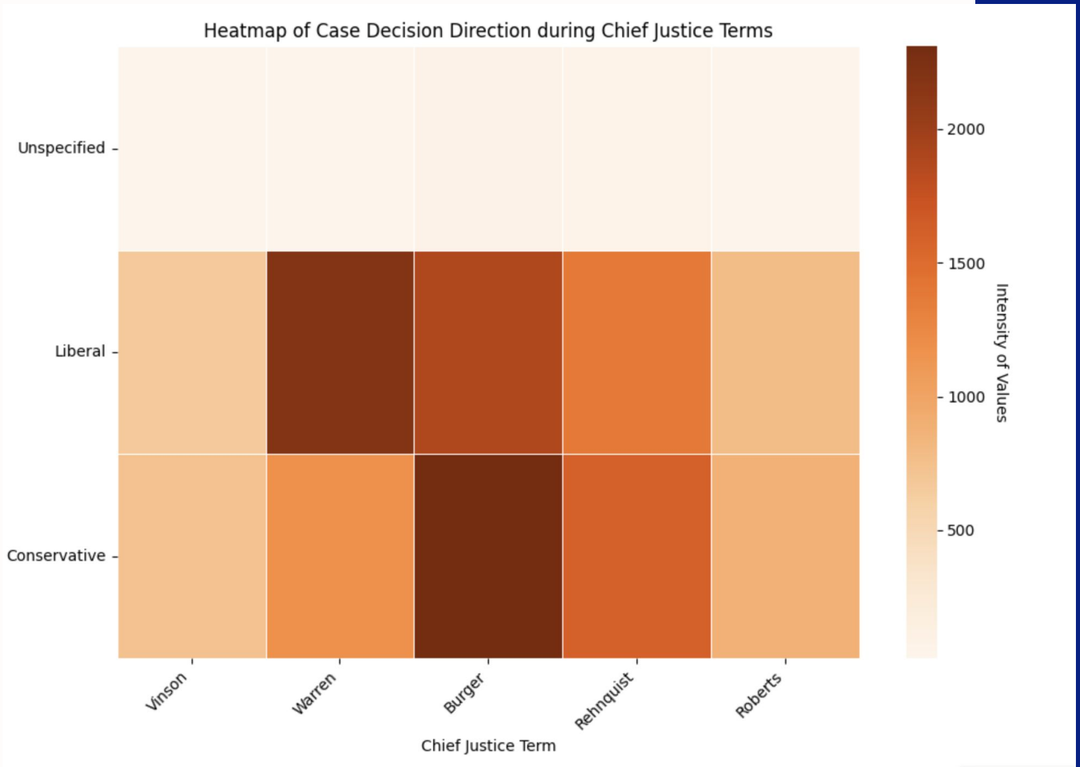
How have case dispositions changed over time, based on who is serving as Chief Justice?



>**Answer**: Using the heatmap we could determine which Chief Justice presided over specific case dispositions. For example, Chief Justice Burger prised over the most cases that resulted in a case disposition "2" ("affirmed").

Exploration Question #3:

Can we categorize the ideological direction of case decisions based on the term of the Chief Justice?

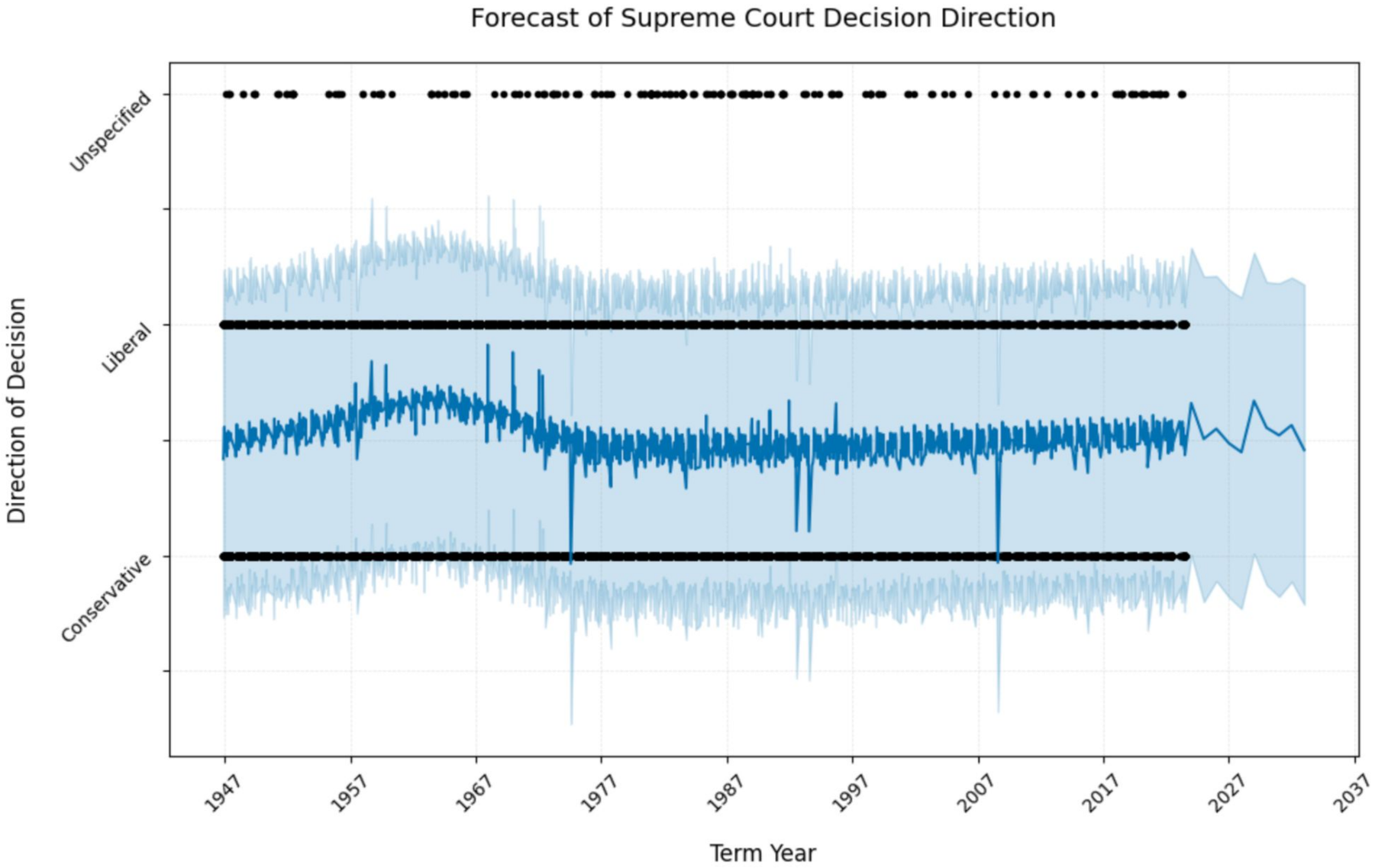


>**Answer:** The heatmap demonstrated that the Court had been rather ideologically balanced during the Chief Justice Vinson years, shifted more liberally during Chief Justice Warren's tenure, before rebounding more conservatively with Chief Justice Burger. The court continued to lean conservatively (though less so) during Chief Justice Rehnquist's term, In the most recent years, under the tutelage of Chief Justice Roberts, the court has returned to a much more balanced position.

Exploration Question #5:

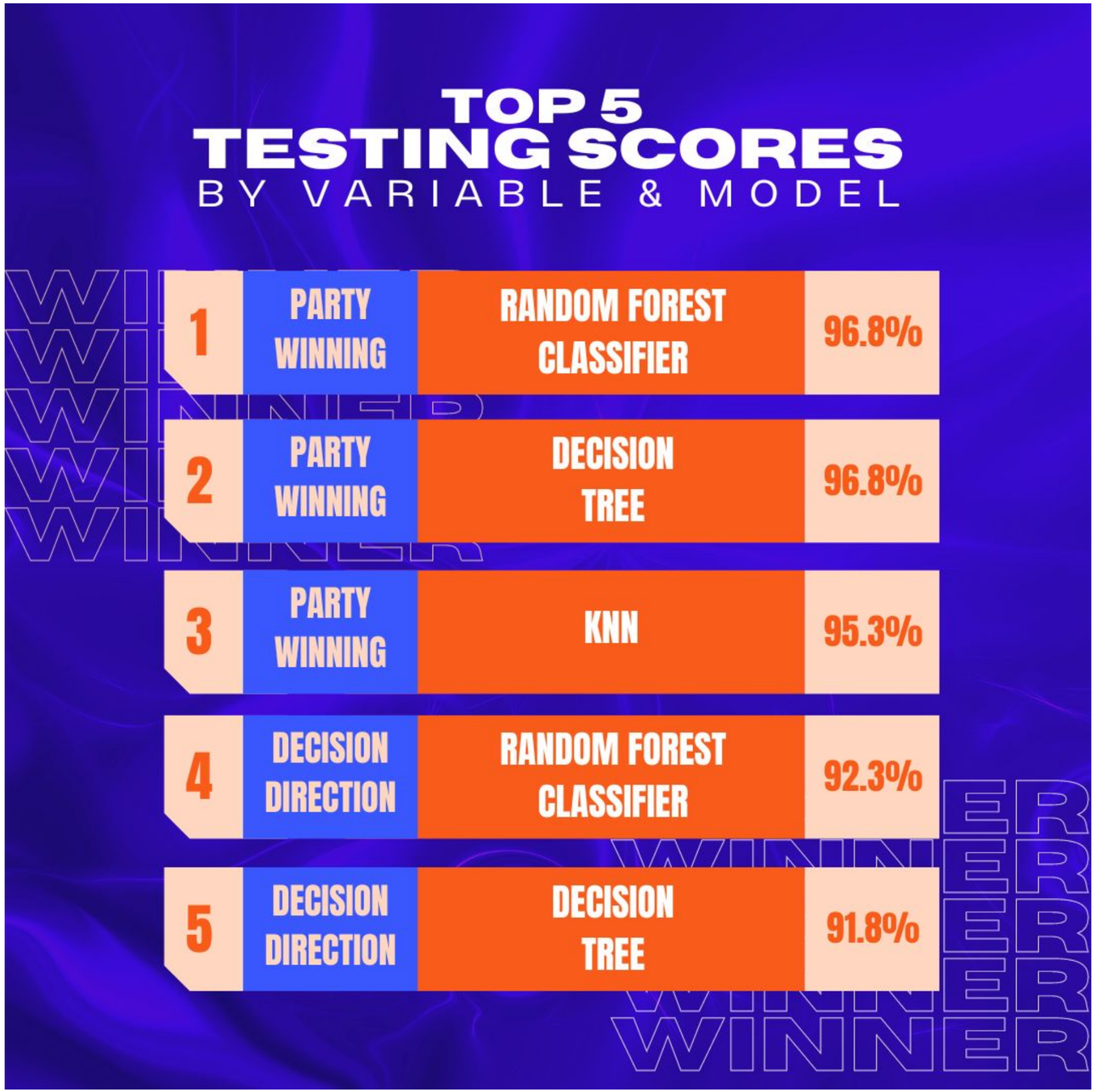
Can we use this data to predict the ideological direction of future decisions?

>**Answer**: Our forecasted model predicts will follow a relatively “seasonal” pattern over the next four years, trending slightly more conservative around 2024, before returning to a more balanced ideological direction around 2028, becoming slightly more conservative again until around 2032-2032, and then again returning to a more balanced ideology. This trend seems to fall in line with the four-year election cycle.



Streamlining our Process: Training, Testing, and Evaluating Multiple Models and Multiple Variables

	target_column	model_name	function_name	training_score	testing_score	accuracy_score
0	partyWinning	KNN	knn	0.961794	0.953233	0.953233
1	partyWinning	Logistic Regression	lr	0.778655	0.785508	0.785508
2	partyWinning	Random Forest	rf	0.986238	0.967956	0.967956
3	partyWinning	Decision Tree	dt	0.986238	0.967667	0.967667
4	decisionDirection	KNN	knn	0.902897	0.886547	0.886547
5	decisionDirection	Logistic Regression	lr	0.629872	0.625289	0.625289
6	decisionDirection	Random Forest	rf	0.941488	0.923499	0.923499
7	decisionDirection	Decision Tree	dt	0.941488	0.918014	0.918014
8	decisionType	KNN	knn	0.842075	0.829388	0.829388
9	decisionType	Logistic Regression	lr	0.832259	0.832564	0.832564
10	decisionType	Random Forest	rf	0.878645	0.852771	0.852771
11	decisionType	Decision Tree	dt	0.878645	0.849307	0.849307
12	caseDisposition	KNN	knn	0.632759	0.605081	0.605081
13	caseDisposition	Logistic Regression	lr	0.571552	0.586894	0.586894
14	caseDisposition	Random Forest	rf	0.683187	0.633372	0.633372
15	caseDisposition	Decision Tree	dt	0.683187	0.634238	0.634238



Using our Models to Make Predictions

Video Link

References & Credit

CitationsHarold J. Spaeth, Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2023 Supreme Court Database, Version 2023 Release 01. URL: http://supremecourtdatabase.org.

(Supreme Court Database Online Code Book)[http://scdb.wustl.edu/documentation.php?s=1]

Thank you to the following individuals for helping us when the code was tricky, or we wanted to stretch our skills!
Julia Kim for helping us when we were stuck
Anthony Inthavong for introducing us to Flask and encouraging us to try!
Zohaib Khawaja, our favorite, former T.A., for whom we customized this ridiculous presentation that he will never see
And, relatedly, Microsoft Image Creator

* “Quiestions” is legalese for questions. Or perhaps, there are perils to using AI image generators to generate text.