Predicting the occurrence, magnitude, and outcome of Securities Class Action Lawsuits

Ish Singham | Banyan Risk Advisor: Michael Densham | Faculty Advisor: Professor Bertsimas

Project Scope, Context, Datasets

Problem and Scope

Project Objective: Build a predictive risk detection and rating system for publicly traded director's and officer's insurance to decrease exposure to low margin or negative margin trades for Banyan's risk capital provider(s) while reducing the leakage of capital, time, and reputation.

Capstone Sponsor Firm Context and Motivation

Banyan Risk: Banyan Risk, Ltd is a specialty insurance Managing General Agent (MGA), focused on providing insurance risk solutions for Directors and Officers (D&O). This industry has seen a rising trend in activity, through security class action lawsuits across the globe. From alleged misrepresentation to fraud and litigation, claims have become a growing source of loss to customers, shareholders, and underwriters over the last 30 years. Banyan Risk Ltd. are working to provide the best-in-industry risk rating model, to support the identification of early warning signs to both decrease exposure to negative margin risk reward trades for their supporting capital risk pool(s).

Banyan Risk currently has around **100 clients**. Each of these firms could receive a claim for a several million-dollar lawsuit during any given point in time. Banyan would like to improve the accuracy of their risk assessment rating models, in order to minimize losses in the future. Their current risk-assessment methodology is largely reliant employees' industry experience, and the rating models of other industry-leading firms.

Banyan Risk would like to pursue an alternative methodology as they have determined that it is impossible to stay ahead of the industry if their models mimic those of competitors.

### Data Description

Providers:
1. S&P500 CapitalIQ Pro (CIQ)
2. Institutional Shareholder Services (ISS) Securities Class Action Historical Cases

Intended Dataset Purposes: CIQ dataset will provide point-in-time market data and historical financials for all public firms over the last 50 years. ISS dataset provides insights into case details, such as legal stakeholders, as well as case outcome and final settlement value.

Using both datasets, I aim to predict the likelihood of any given firm being subject to securities class action litigation, given their historical financial performance, sector, and other variables. After assessing the risk of an individual company facing litigation, I will then analyze the likelihood of potential outcomes and, when relevant, forecast the most likely settlement value incurred by the firm, and its capital providers.

### Director's and Officer's Insurance Industry Landscape

Banyan Risk compete in the middle layer of this industry. They are tasked with measuring risk and pricing insurance accordingly. There are a multitude of firms in this layer, competing for a finite number of prospective clients. It is crucial that Banyan develops a rater model that sets it apart from other industry players.
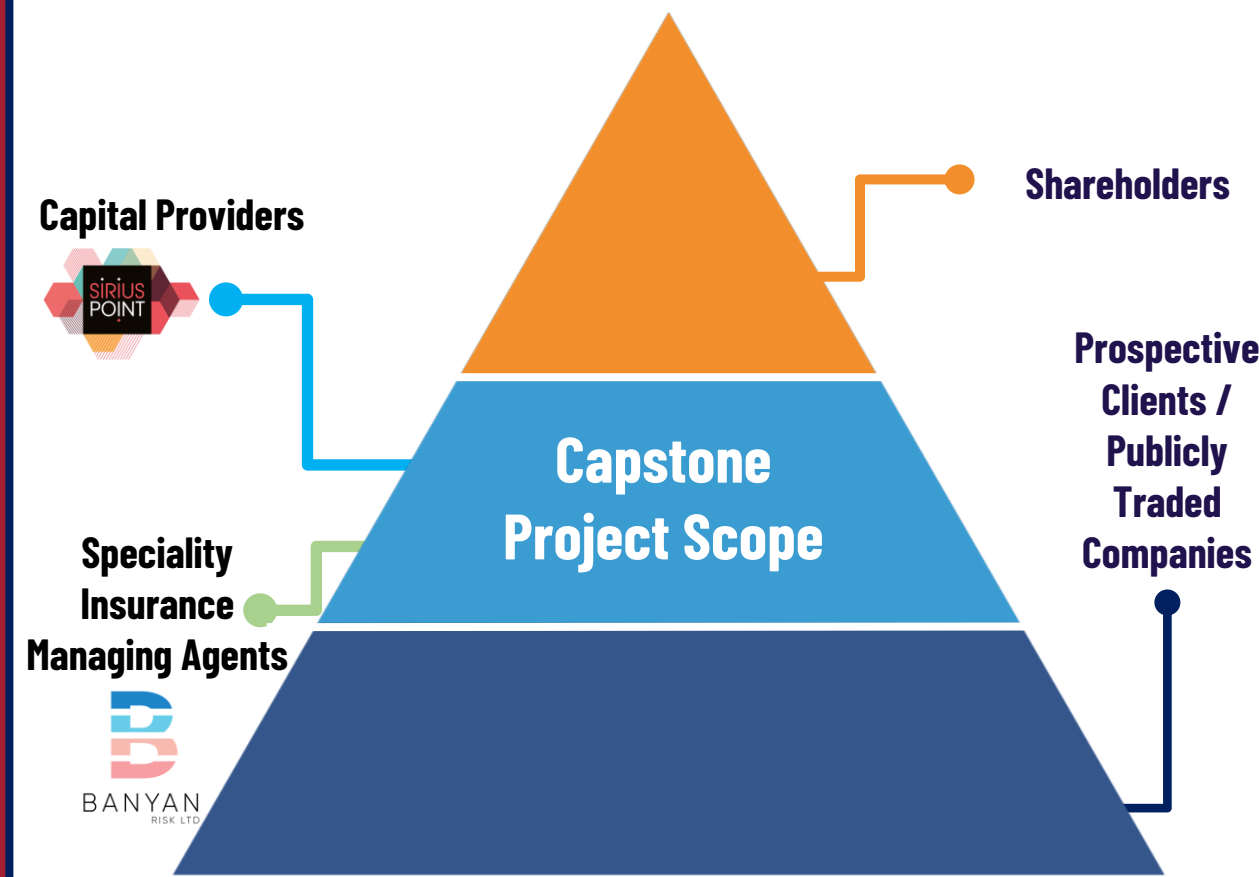
Note: The middle part of the pyramid measures, prices, and (in the event of a securities class action settlement) compensates for risk. Softening the blow to the bottom segment.

Insurance Risk Education and S-1 Tool

In order to develop my understanding of the intuition surrounding risk-determinants within the Directors and Officer's insurance space, I begun my project by creating a tool that locates, parses, and exports data from SEC S-1 filings.
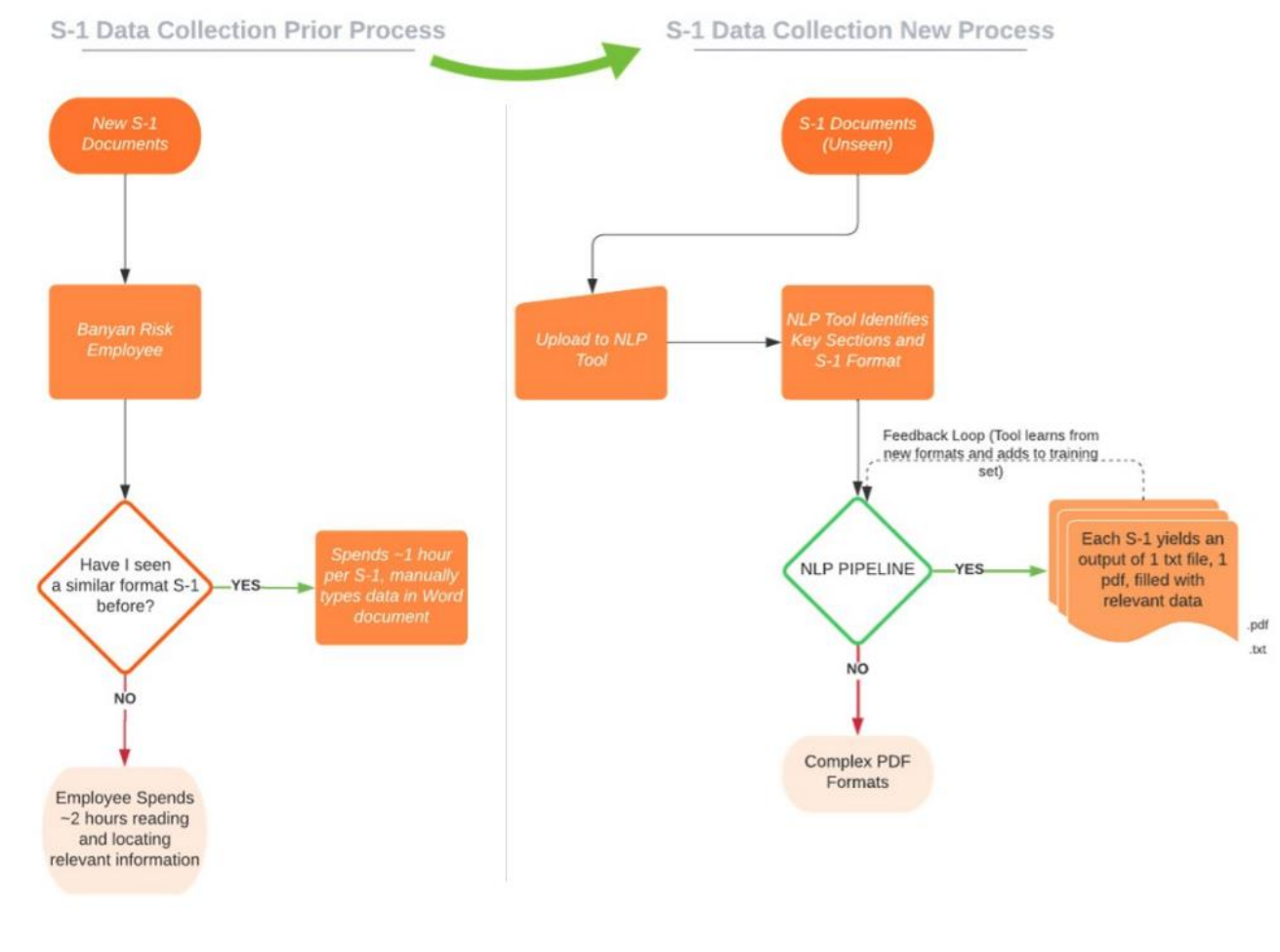
S-1 Tool Impact, Performance

- **85% Accuracy** in Collection and Reporting of Key S-1 Risk-Indicating Fields
- **2.2 minute average runtime** (relative to 90+ minute manual process)
- **Learning vs Fatigue**: With each iteration, tool learns new formats as S-1 reporting changes.
- **New Fields**: Ease of integration for new datapoints, inclusion of risk phrase counts, reporting sentiment.

97% Decrease in Minutes Spent Processing S-1 Documentation

- 87 Minutes Saved
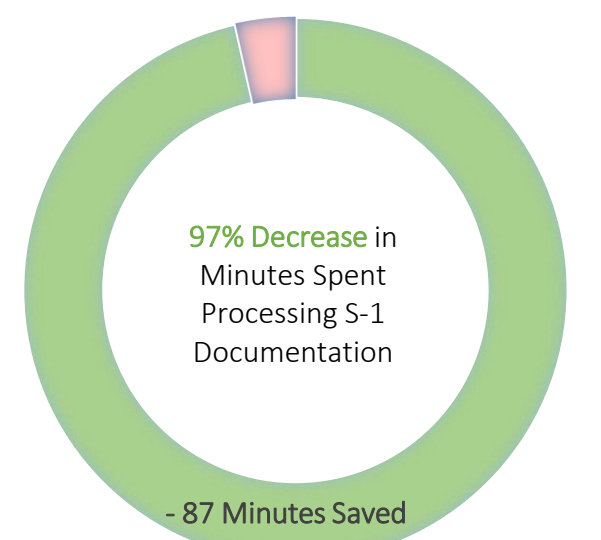
- **Employee Testimonial:**
"Ishaan's tool has helped me tremendously in my day-to-day at work. Instead of manually inputting cumbersome data from lengthy SEC filings (S-1), I have been able to access the targeted data upon the click of a button. With the help of Ishaan, I have been able to allocate many hours I would have spent reading and memorizing to other areas of my responsibilities." – Ben Lines, Banyan Risk

Methodology and Analytical Methods

Feature Overview and EDA

| S&P 500 Global Capital IQ (CIQ)	Institutional Shareholder Services (ISS)	Explanation
Firm Security/Ticker	CaseID	**Unique Identifier**
Average Volume	Filing Date	
Short Interest	Allegation Type	
Equity Beta	SIC Code (Sector)	
Market Capitalization	Headquarters Country	
Comparable Companies	Case Accountant	**Predictive Features**
P/E Ratio	Lead Plaintiff Lawfirm	
Credit Ratings	Lead Defendant Lawfirm	
Inside Ownership	Case Judge	
Firm Sector Performance	Case Underwriter	
Litigation	Case Outcome, Settlement Value	**Response Variables**

Table Summary
I used legal and financial data to inform my predictions. My key dependent variables that I aimed to predict were:
1. The occurrence of a Class Action Securities Case
2. The outcome of that Case (both status and magnitude)

Subproblem 1: Will the client be targeted by a Class Action Lawsuit?

As mentioned above, Banyan Risk had their own rater model, with a series of features that the industry believes predict class action lawsuits in public companies. These "red flags" are "consensus," and therefore have received little pushback from clients and industry leaders. Before assuming that each of these features were true, I performed analyses on each factor and explored its predictive power.

Price Volatility (Historically: Equity Beta)
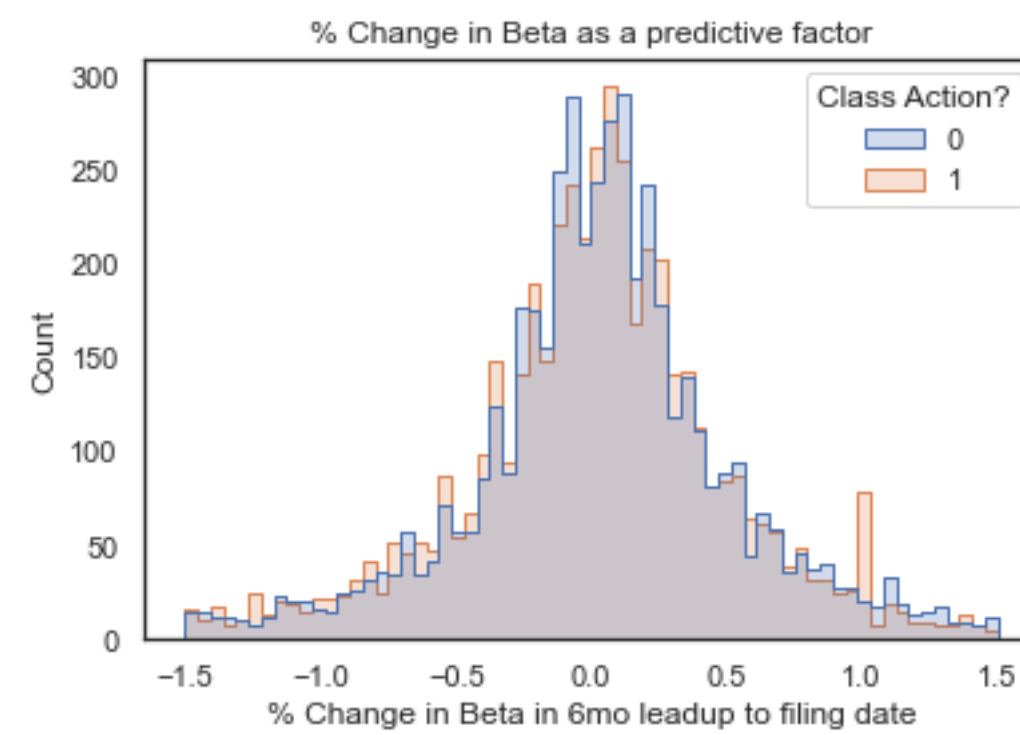
Subproblem 2: Given that the client is subject to a Class Action Lawsuit, what is the expected outcome?

Choice: why I chose classification trees, feature selection relevance

Optimization Formulation and Induction:

Given ISS Securities Class Action training data $(\mathbf{x_i}, y_i)$, where:

- $\mathbf{x_i} \in \mathbb{R}^p$ ($\mathbf{x_i}$ represents ISS case features)
- $y_i \in \mathcal{K}$ (y_i represents ISS case outcomes)
- We denote by $\mathcal{K}$ the set of class labels), $i = 1, 2$, we want to find a tree $\mathcal{T}$ that solves the problem:

$$\min \; error(\mathcal{T}, \mathbf{X}, \mathbf{y}) + \alpha \cdot complexity(\mathcal{T})$$

Specify the following hyperparameters:

- D: Maximum depth: Limit on the number of layers in tree - selected using k-fold cross validation approach with grid-search, upper limit of **30** to ensure interpretability.
- N_{min}: Minbucket, i.e., minimum number of cases to assign to a leaf: **4** to ensure model doesn't overfit on specific class action earthquakes.
- α: Complexity parameter that controls the trade-off between accuracy and complexity of the tree - selected using k-fold cross validation approach with grid-search.
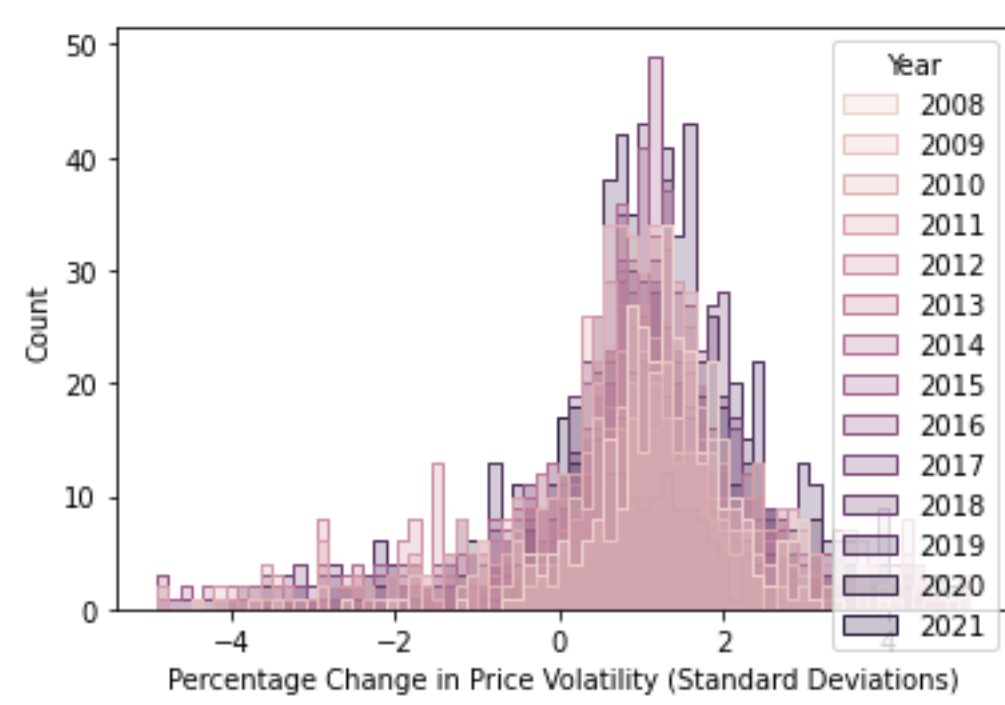
The two graphs displayed in the panel to the left exemplify the fallback of applying vague rules/buckets as "risk factors," without considering changes in time, industry, and progress through the business cycle. These factors were not significant at the global (all cases) scope.
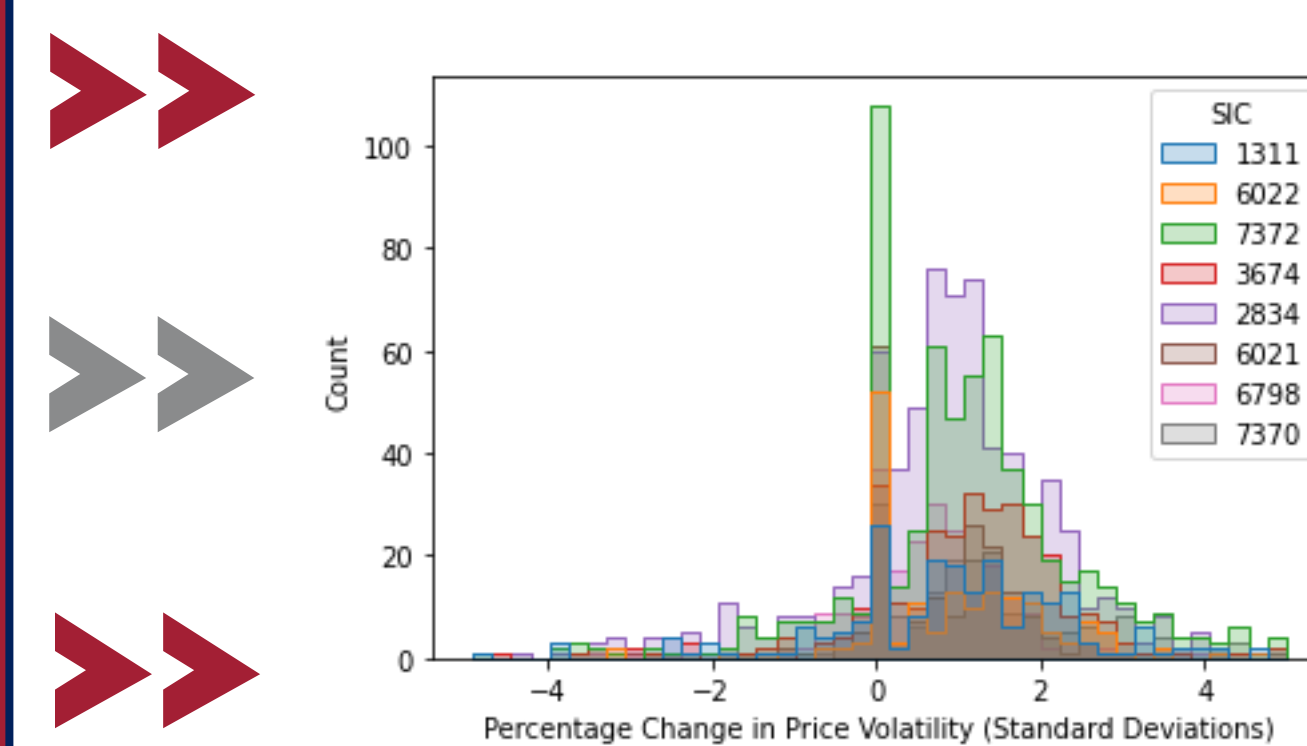
Benefit of Detailed Case-by-Case Segmentation

SIC Code Description
Management Investment Offices, Open-End
Engines And Turbines
Personal Services
Computer Programming, Data Processing, And Other Computer Related Services
Commercial Banks
Security Brokers, Dealers, and Flotation Companies
Electronic Connectors
Engineering, Accounting, Research, Management, and Related Services

Results and Impact

Impact / Explanation

93.5% Accuracy on Class Action Outcomes Classification Model
Using only historical class action legal data, model predicts correct outcome of 12.1k/13k class actions in the last 30 years.

End-to-end data collection, cleaning, analysis takes 25 minutes
Scalability is crucial for Banyan. My tools placed emphasis on runtime, and integrate with monthly ISS dataset updates seamlessly.

**0.82% Accuracy for Optimal Impute Procedure**
Banyan has very little control over the quality of data from its data providers. My capstone project leverages optimal impute to accurately estimate missing key financial data.

Sparse Feature Selection Summarizes Data in just 20 features
Banyan has very little control over the quality of data from its data providers. My capstone project leverages optimal impute to accurately estimate missing key financial data.

Exemplar Optimal Case Outcome Classification Tree Output
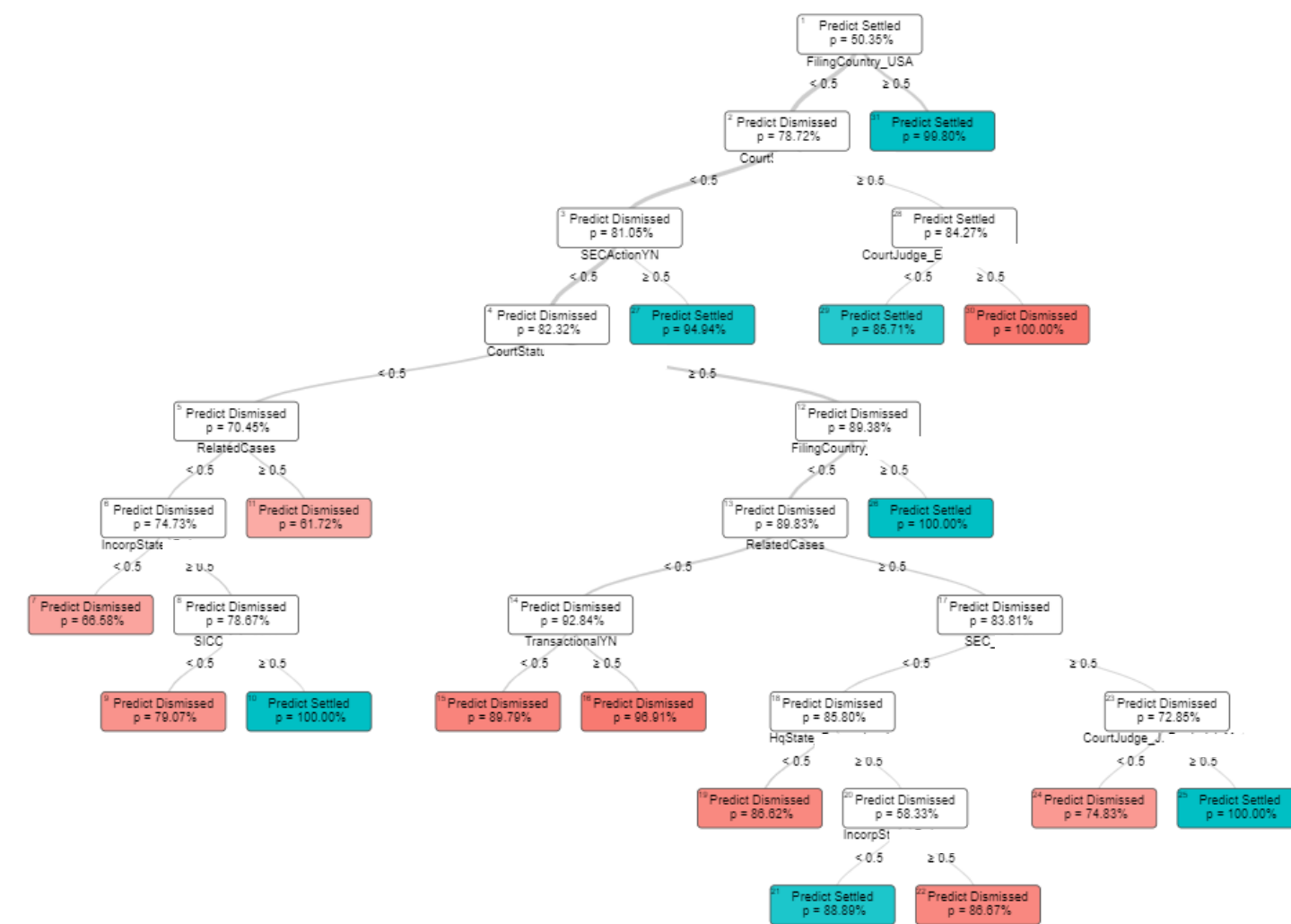
What this really means: To interpret the above graphic, imagine we have all the case details for an ongoing class action lawsuit. Follow the tree at each step. We note that the first step is "if filing country is the US, predict Settled." This is accurate 98% of the time, largely as my dataset was missing a value for Filing country 95% of the time.

We note that this tree provides valuable insights, outlining the key judges, industries, or even states of incorporation that may lead to higher probability of a settlement. Specific Judge Names, Firms, Countries, and Industries have been removed for confidentiality purposes.

Next Steps

Forecasting and Budgeting
The new rater model framework allows for Banyan Risk to project their expected cash outflows, to their capital provider. And construct an optimal portfolio to manage their exposures more accurately.

Competing Rater Models
If Banyan Risk would like to outperform incumbents, my project's new rater model will allow them to watch as signals gain/lose strength. Their rules are different to competitors, as unorthodox metrics were selected to proxy for risk-driving financial characteristics.

Dataset and Tool Scope

13k+ Class-Action Firms

4k+ Case Underwriter(s) and Accountant(s)

5k+ Judges, 1k+ Plaintiff, Defendant Lawfirms

30 Years of Case Data