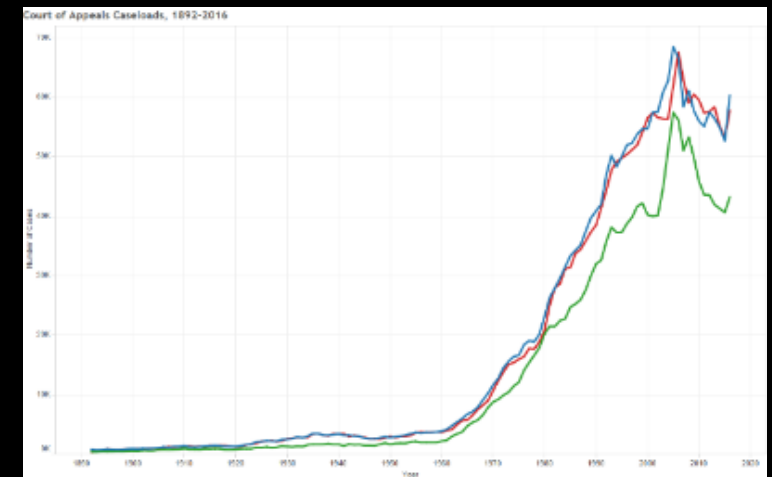# A data driven approach to optimizing the US Court system
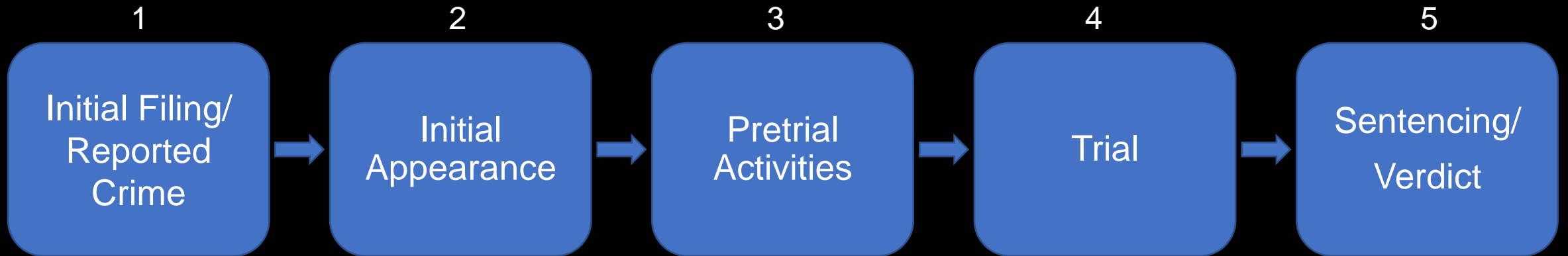
Dhiraj Kanneganti

Judicial Crisis

- The American court system is becoming increasingly overwhelmed, projected 9 million cases currently being delayed

- Lost the government upwards of 1.4 billion dollars over the past decade

- Research shows that nearly 80% of delayed cases have some level of bias present in the verdict

- Experts project that if a solution is not found to this problem immediately, the number of delayed cases could double soon
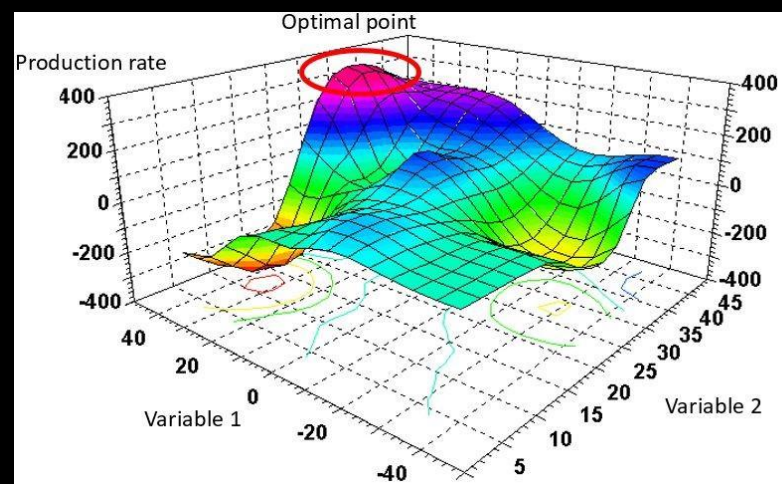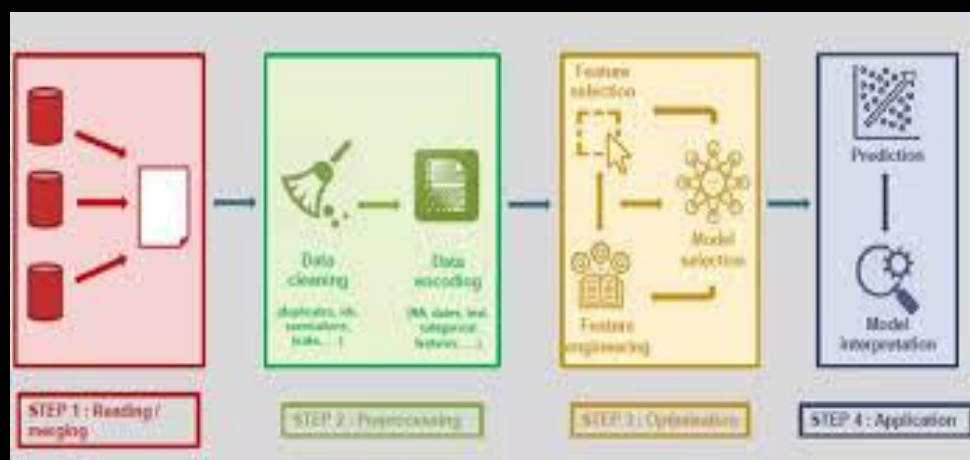
Path of a Case

- The path of a case is a process flow issue

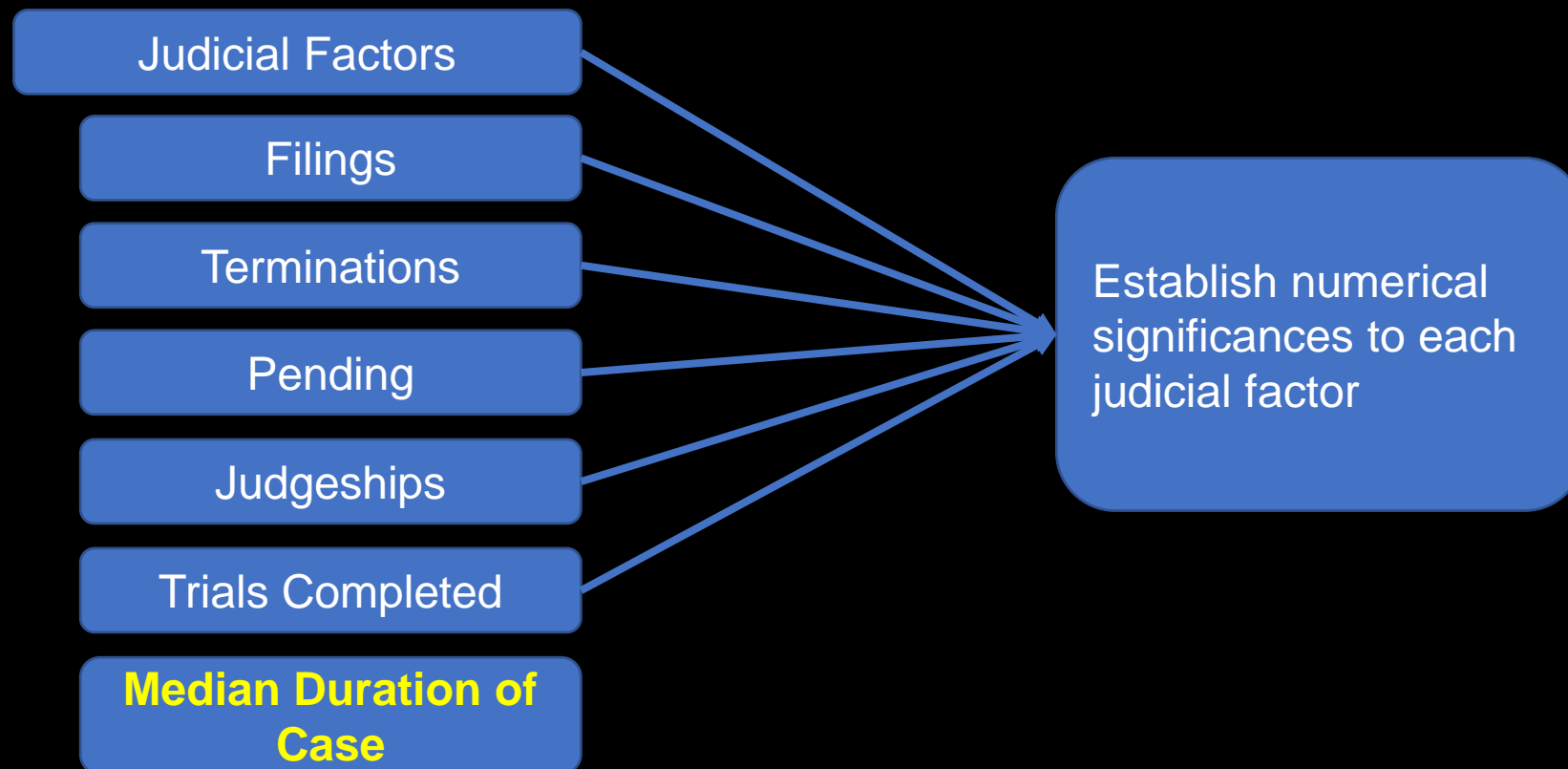| 1	2	3	4	5
Initial Filing/ Reported Crime	→ Initial Appearance	→ Pretrial Activities	→ Trial	→ Sentencing/ Verdict

Workflow Optimization

- One of the primary use of ML in the status quo is to optimize production in the manufacturing industry
- The same process could be applied to the judicial system
- By targeting the most significant variables, we can identify what is the reason for the inefficiencies

Thought Process

- Reasoning behind why the experiment was formatted as such

```
Judicial Factors ──┐
Filings ───────────┤
Terminations ──────┤      Establish numerical
Pending ───────────┼──▶   significances to each
Judgeships ────────┤      judicial factor
Trials Completed ──┤
Median Duration of ┘
Case
```

Random Forest

- Type of unsupervised ML algorithm

Pros:
Able to evaluate data with multiple dimensions
Can handle outliers/uneven data
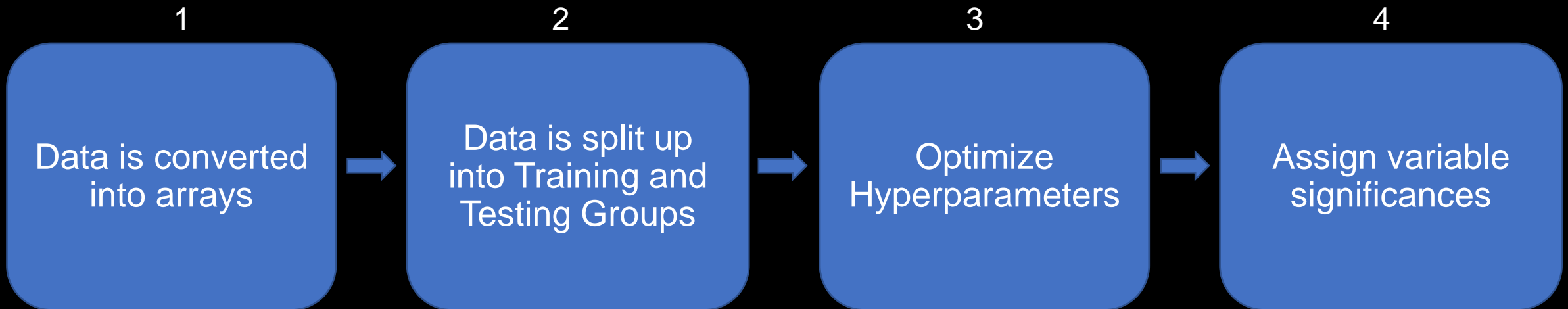Easily optimizable

Cons:
Often overfits data
Not interpretable

Random Forest Simplified
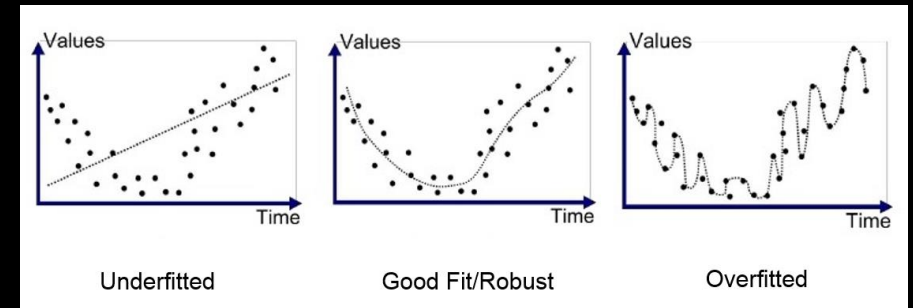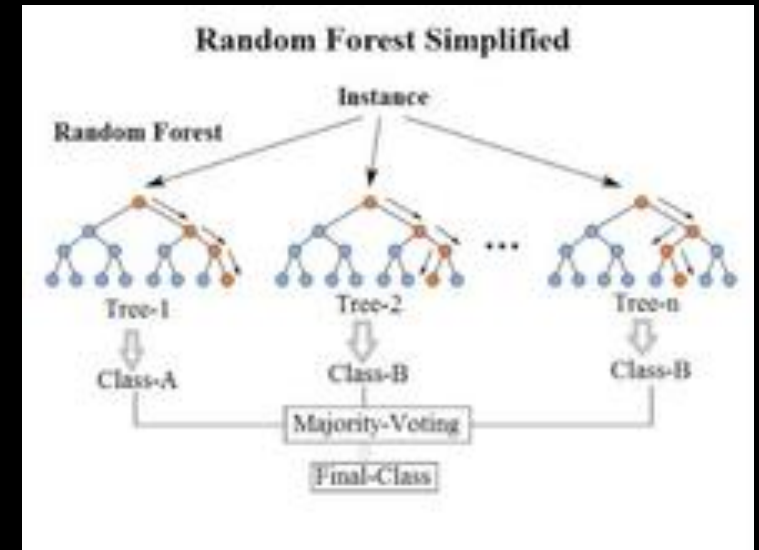
Is it possible to design and construct a machine learning algorithm to successfully identify factors that cause the bottlenecks of the court system?

Methodology

- Process of using judicial data to identify potential inefficiencies in the pipeline

1	2	3	4
Data is converted into arrays	Data is split up into Training and Testing Groups	Optimize Hyperparameters	Assign variable significances

Step 1: Data Conversion

- Data was taken from National Judicial Database
- Data was converted into vector arrays

Scalar
24

Vector
[2 -8 7]

[2
  -8
  7]

Matrix
[6 4 24
  1 -9 8]

U.S. District Court — Judicial Caseload Profile

MASSACHUSETTS		Jun 30 2015	Jun 30 2016	Jun 30 2017	Jun 30 2018	Jun 30 2019	Jun 30 2020
	Filings [1]	6,006	4,290	3,697	3,964	4,099	3,621
Overall Caseload Statistics	Terminations	3,905	3,894	4,706	5,037	6,004	3,705
	Pending	8,444	8,824	7,813	6,725	4,853	4,731
	Percent Change in Total Filings Current Year Over Earlier Year	-39.7	-15.6	-2.1	-8.7	-11.7	
	Number of Judgeships	13	13	13	13	13	13
	Vacant Judgeship Months [2]	13.0	12.0	0.0	17.9	24.0	24.0

This piece of data shows that in Massachusetts District Court in 2019, there were 6,004 cases terminated.

Step 2: Training and Testing Groups

- Data was split into Training and Testing Groups (70/30 split)
- Avoids overfitting but also makes sure algorithm has enough data to make a prediction model
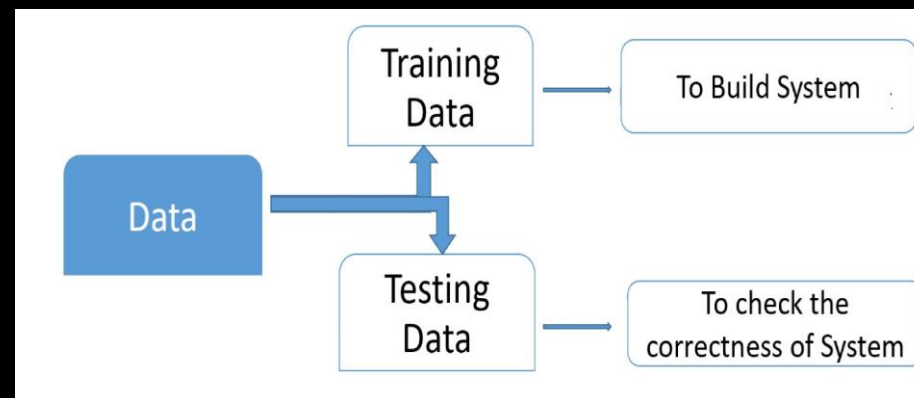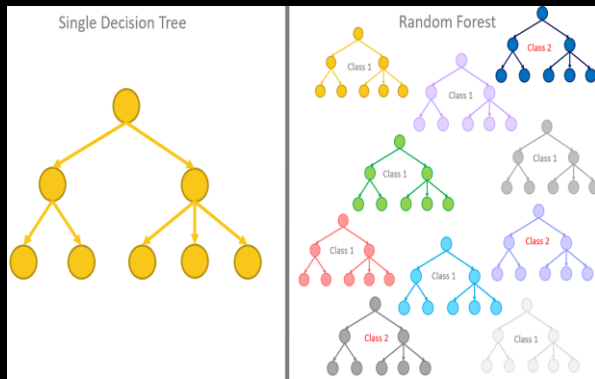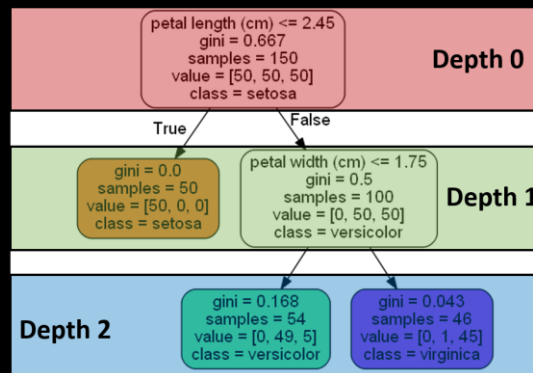
Step 3: Optimization

A hyperparameter is a parameter that has a default value that can be changed to increase the accuracy of the model
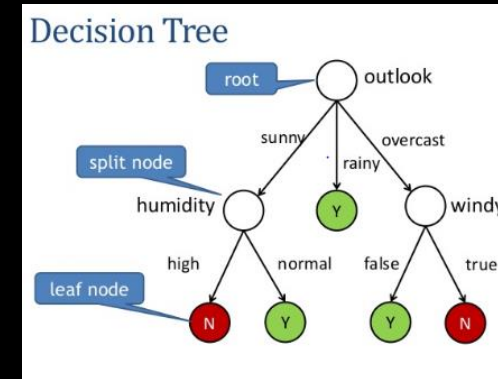
Hyperparameters
:

Number of Decision Trees

Depth of Decision Trees

Nodes per Split

Step 3: Optimization Cont.

- Implications of changing the hyperparameters

Hyperparameters	Advantages	Disadvantages
Decision Trees	Forces algorithm to take all judicial factors into account	Leads to Overfitting
Depth of Decision Trees	Helps algorithm filter out noise/outliers	Leads to Overfitting
Nodes per Split	Algorithm is more specific when splitting branches	Increases Processing Time

Step 4: Assign Variable Significances

The chart below shows the numerical significances of the Judicial Variables from 2016 to 2020.

Average Significance from 2016-2020	
Judicial Factor	Average
Pending Cases	0.0846
Civil Time	0.0814
Average Present for Jury selection	0.078
Criminal Felony	0.0662
Percent of Cases over three years old	0.0622
Criminal Felony	0.06
Number of Cases over three years old	0.0574
Vacant Judge Months	0.0558
Percent Not Selected or Challenged	0.05
Filings	0.0454
Civil	0.0422
Supervised Release Hearings	0.0412
Terminations	0.0316
Trials Completed	0.0314
Weighted Filings	0.0292
Pending	0.0288
Terminations	0.0274
Total Filings	0.0246
Judges	0.0208
Average number of Felony Defendants per case	0.0188

This piece of data shows that from 2016 to 2020, Pending Cases had an average significance of 8.46%. This means that nearly a tenth of the inefficient output of the judicial system can be attributed to this variable.

Average Significance from 2016-2020
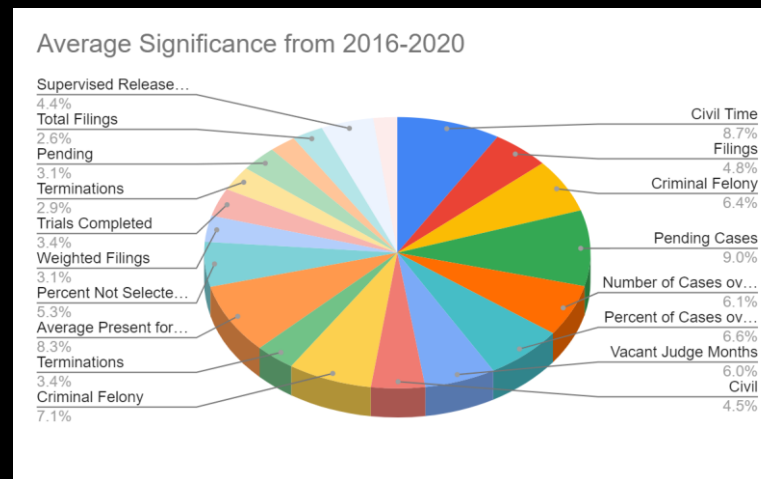
Yearly Analysis of Variable Significances

The bar graphs below depict the numerical significances on a yearly basis.
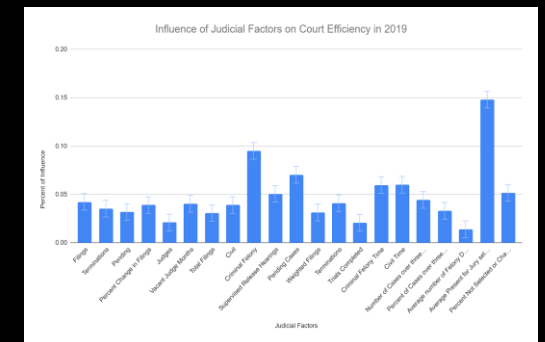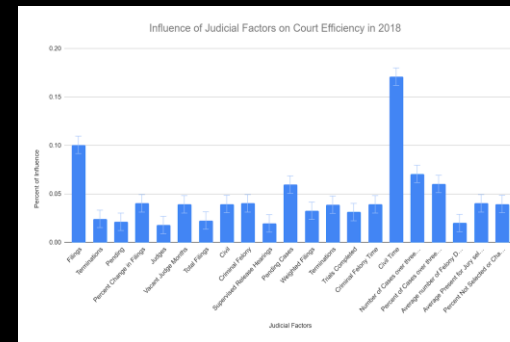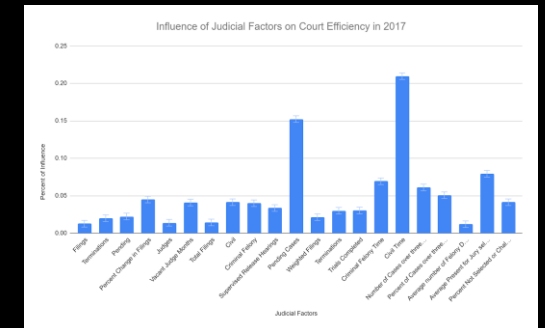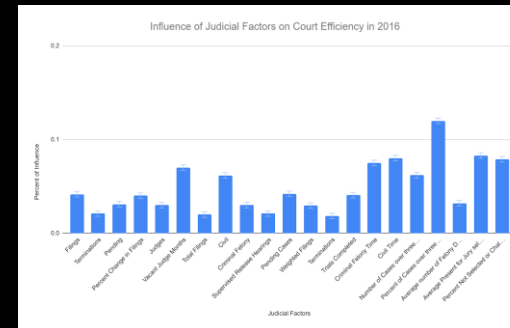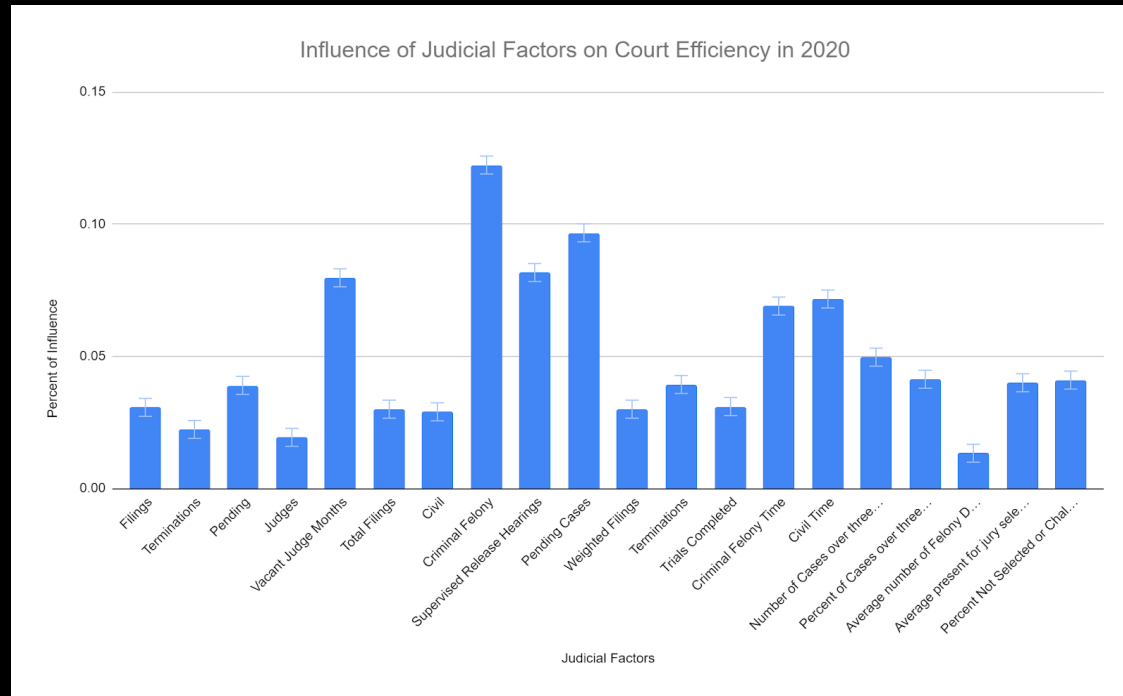
Conclusions

- The most significant variables were Number of Pending Cases, and Civil Time, Criminal Time

- There was one common trend between all the most significant variables: they all occur in between filing and the start of the trial

Path of a Case Revisited

- The results suggest that the duration in between the initial filing and the pretrial activities is a significant cause of the congestion in the legal system
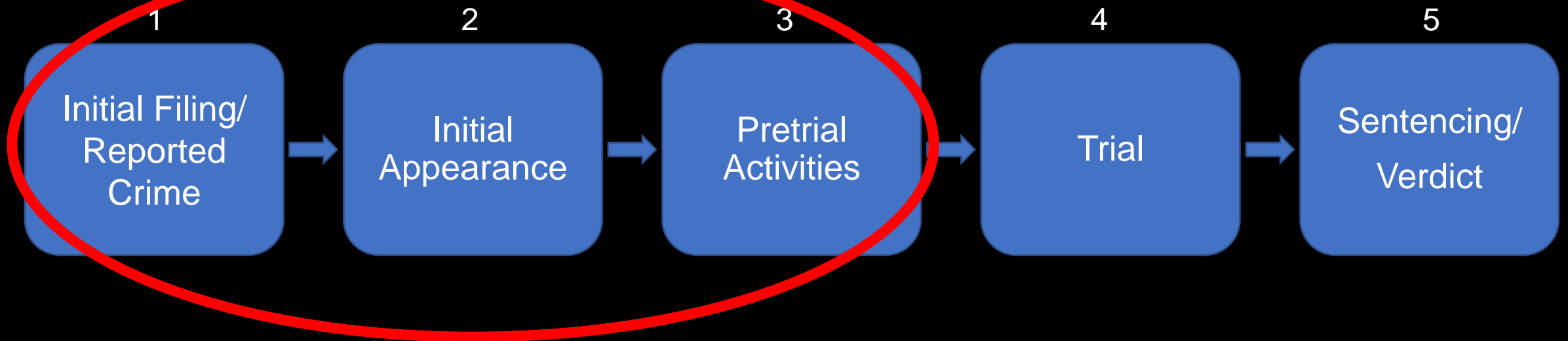
Potential Solutions

- Implementing healthy practices in the legal system to prevent gamification
- Use the algorithm to analyze individual courts
- Eliminate any other potential extraneous variables

Gamification

- Rebecca Kourlis, a former justice in Colorado, conducted a survey of judicial officials and lawyers and she reached a similar conclusion
- The gamification of the court system directly impacts the time in which it takes for a case to go from filing to trial
- Change needs to be made to the justice system to shift the problematic behavior

Judicial Survey about Gamification in the Court System
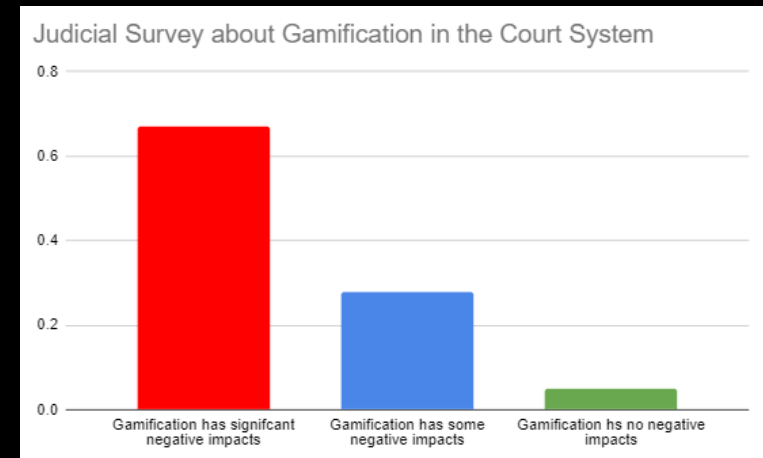
Individual Courts

- Looking at this workflow issue from a national perspective isn't enough
- Algorithm's model can be used to analyze individual court systems
- In-depth look could eliminate potential extraneous variables

	2010 pop'n (16+)	Alcohol/Drug Related Traffic Cases	Other Traffic Cases	Non-Traffic Ordinance Cases	Total Cases	Cases per Person	Fines	Court Fees	Fines Per Person
Byrnes Mill	2,155	111	3,269	508	3,888	1.80	$ 443,433	$ 32,853	$ 205.77
Hillsboro	2,131	27	2,618	94	2,739	1.29	$ 196,119	$ 22,051	$ 92.03
Festus	8,808	87	2,688	1,942	4,717	0.54	$ 583,019	$ 43,407	$ 66.19
Pevely	4,043	38	2,657	531	3,226	0.80	$ 244,519	$ 19,833	$ 60.48
Crystal City	3,759	110	1,880	1,197	3,187	0.85	$ 223,259	$ 21,921	$ 59.39
Arnold	16,529	163	11,307	1,262	12,732	0.77	$ 906,379	$ 129,957	$ 54.84
De Soto	5,000	77	726	475	1,278	0.26	$ 169,278	$ 12,942	$ 33.86
Herculaneum	2,718	3	658	92	753	0.28	$ 62,895	$ 5,605	$ 23.14
Jefferson County	170,082	614	22,008	2,965	25,587	0.15	$ 1,882,638	$ 167,809	$ 11.07
Kimmswick	157	0	0	0	0	0.00	$ -	$ -	$ -

Individual Court Analysis

- Proposed methodology when analyzing individual court systems

1	2	3	4
Data is converted into arrays	Data is split up into Training and Testing Groups	Data is entered into prediction model	Identify problem areas in workflow