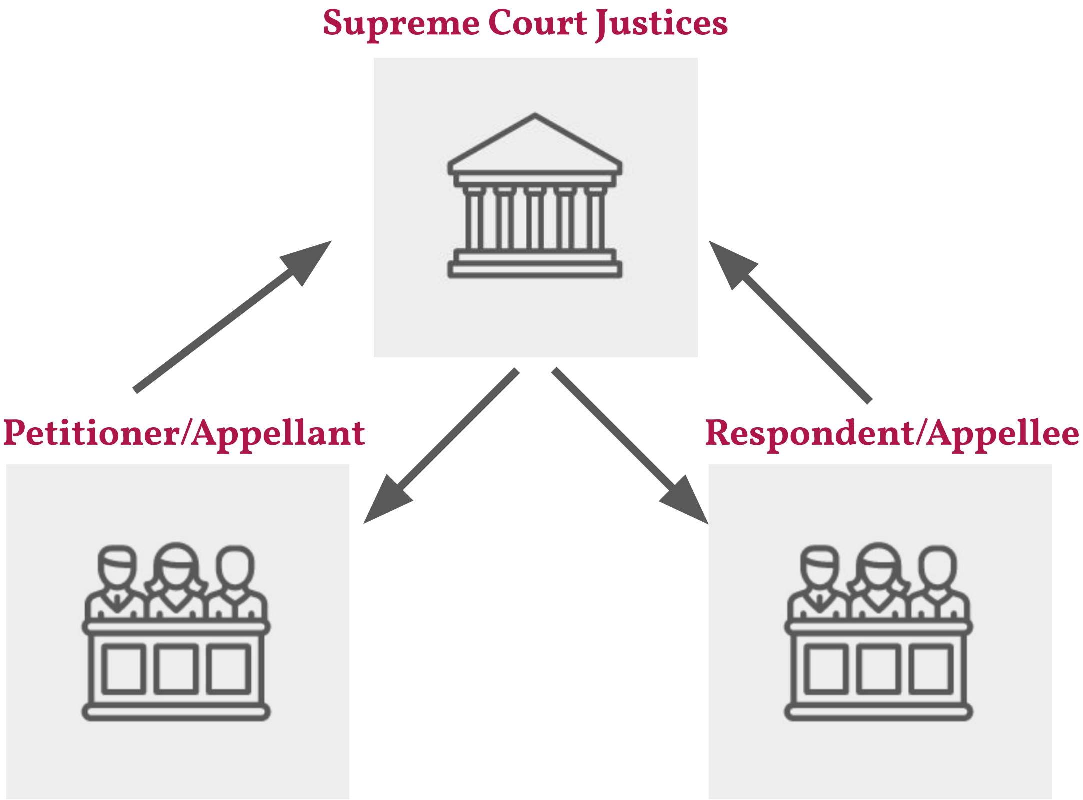
Using Oral Arguments to Predict Supreme Court Case Outcomes

Jaime DyBuncio, June 2020
github.com/jdybuncio
Galvanize g119 cohort

The U.S. Supreme Court in 1 Slide| Parties Involved



Creating a Dataset | Parsing PDFs

PDF of Oral Argument

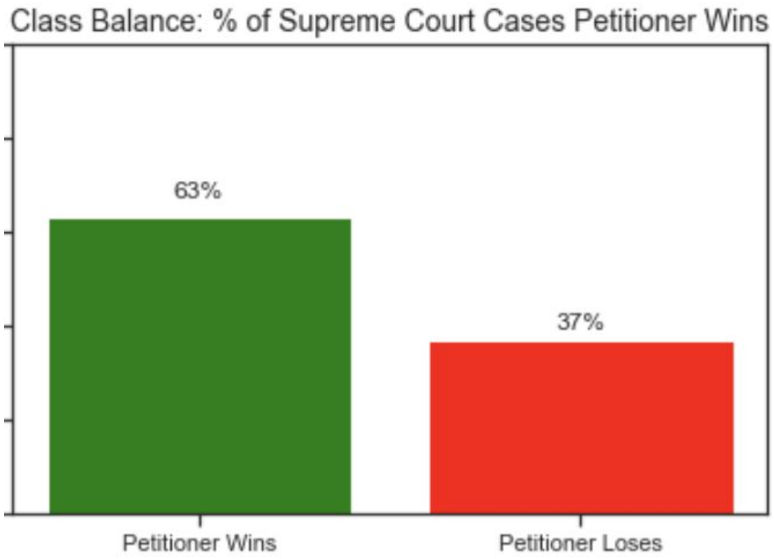
Oyez.org

My Parsing Script & Functions

- **Supplement w/ Case Data**
  - Identifies 10 types of speaker directions
    - Exs:
      - Petitioner to Justice
      - Justice to Petitioner
      - Respondent to Justice
      - Justice to Respondent
      - Other 6 have to do w/ an Amicus which can be neutral or on behalf of one of the parties
  - Add Label if Petitioner Wins

Party	Petitioner -to-Justice	Justice-to-Pe titioner
(1) Words	61	3
(2) Time	13.9 secs	1.23 secs
(3) Interruptions	1	1
(4) Questions	0	0
(5) Document	"I think it is very hard to imagine.."	"Well, that's another"
(6) Petitioner Wins	**1**	**0**

EDA | Class Balance

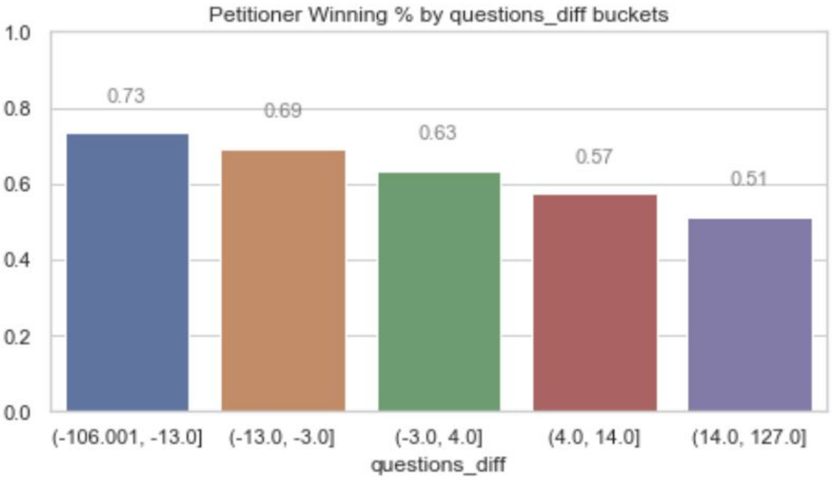


EDA | Interesting Feature - Difference in Questions b/w Petitioner & Respondent

The New York Times

When the Justices Ask Questions, Be Prepared to Lose the Case

By Adam Liptak



Note: Petitioner Win Rate is always above 50%

Baseline Model : Petitioner always Wins

	F1	Recall	Precision
Baseline_Train	0.77	1.00	0.63

Modeling | Numerical w/o using Text

Random Forest Parameters used in Grid Search

Parameter	Grid Search Values
n_estimators	[100,200,500,1000,1200]
max_features	['sqrt', 0.25,.50,None]
min_samples_split	[3, 5, 10, 50]
min_samples_leaf	[5, 10, 50]
max_depth	[3, 5, 10, None]

(best model's params are underlined)

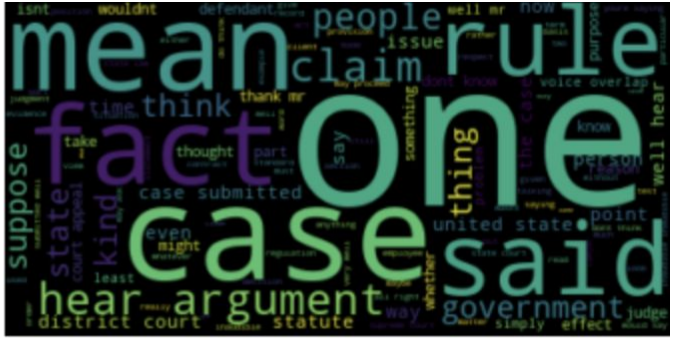
	F1	Recall	Precision
Baseline_Train	0.77	1.00	0.63
Model_Train	0.78	0.98	0.65

- **GridSearched** over 1,400 models
- **Tested:** Logistic Regression, Random Forest, Gradient Boost Classifiers
- **Top 3 Features:** Petitioner - Respondent Difference in: Judge Talk Time, Questions, & Interruptions

Modeling | Numerical w/ Tf-idf Features

Petitioner-to-Justice most frequently used words

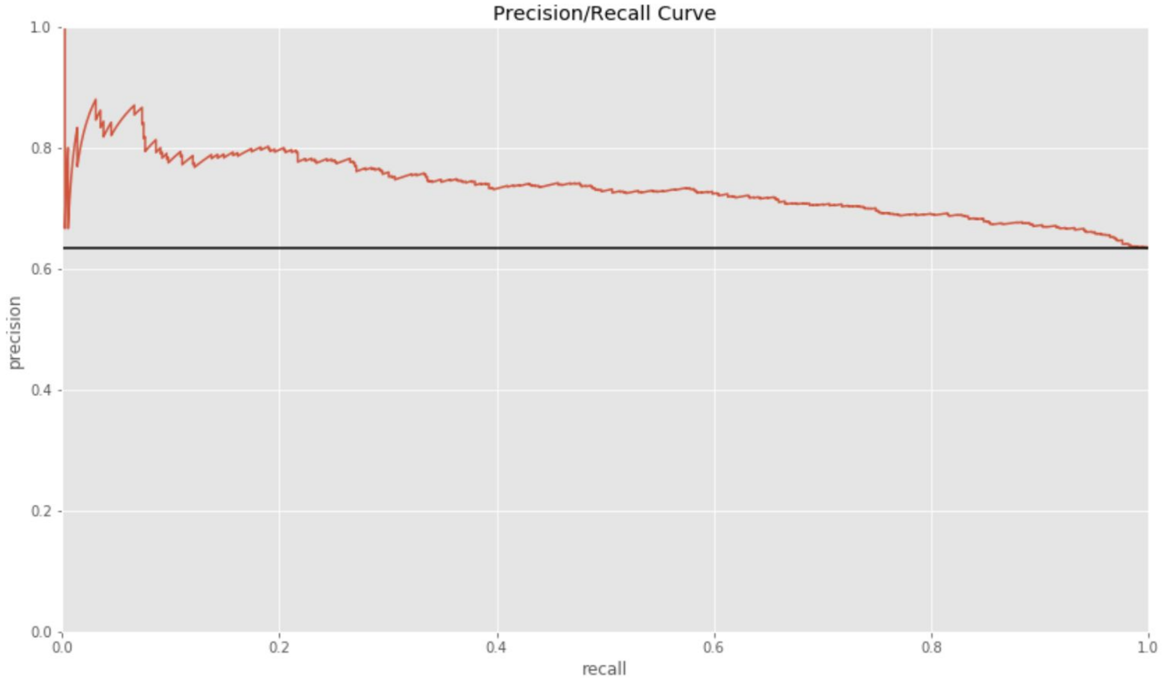
Justice-to-Petitioner most frequently used words



- **Leverage Tf-idf** to try to identify if there are words which are important in certain cases with respect to the corpus
- **Corpi tested:** Petitioner Words to the Judges, Judges words to the Petitioner, and both combined
- Judgment words like: Correct, Good / Incorrect, Wrong do not appear frequently

	F1	Recall	Precision
Baseline_Train	0.77	1.00	0.63
Model_Train	0.79	0.98	0.65

Modeling | Final Results vs. Test Data



Confusion Matrix at Threshold which maximizes F1-Score

Predicting 1,114 Supreme Court Cases

Threshold	TP	FP	TN	FN	Recall	Precision	F1
Best Model (Threshold = 0.51)	680	357	52	25	0.96	0.66	0.78
Baseline (Petitioner Wins 100%)	705	409	0	0	1.00	0.63	0.78

Predictions for the 10 Cases heard in May, 2020

case_name	Argued	Petitioner	Respondent	Talk_time_judges_diff	Questions_diff	Interruptions_diff	Prediction_Proba
U.S. Patent and Trademark Office v. Booking.com B.V.	2020-05-04	United States Patent and Trademark Office	Booking.com B.V.	-0.03	-0.26	0.67	0.64
United States Agency for International Development v. A	2020-05-05	United States Agency for International Develop	Alliance for Open Society International, Inc., et al.	-0.21	-0.06	0.24	0.66
Barr v. American Association of Political Consultants Inc	2020-05-06	William P. Barr, Attorney General; Federal Con	American Association of Political Consultants, Inc., e	-0.34	-0.26	-0.05	0.69
Little Sisters of the Poor Saints Peter and Paul Home v.	2020-05-06	The Little Sisters of the Poor Saints Peter and	Commonweath of Pennsylvania and State of New Je	-1.78	0.20	-0.56	0.68
Our Lady of Guadalupe School v. Morrissey-Berru	2020-05-11	Our Lady of Guadalupe School	Agnes Morrissey-Berru	-0.26	0.05	0.38	0.67
McGirt v. Oklahoma	2020-05-11	Jimcy McGirt	Oklahoma	0.29	0.41	0.96	0.54
Trump v. Vance	2020-05-12	Donald J. Trump	Cyrus R. Vance, Jr., in His Official Capacity as Distric	-1.14	-1.24	0.31	0.73
Trump v. Mazars USA, LLP	2020-05-12	Donald J. Trump, et al.	Mazars USA, LLP, et al.	-1.93	-1.50	-0.27	0.76
Colorado Department of State v. Baca	2020-05-13	Colorado Department of State	Micheal Baca, et al.	0.31	-0.01	-0.48	0.60
Chiafalo v. Washington	2020-05-13	Peter Bret Chiafalo, Levi Jennet Guerra, and E	State of Washington	0.55	0.97	0.89	**0.47**

^Features are Centered & Scaled

- Models cited this space use Accuracy to optimize their models (source) and are able to get around 70% Accuracy
- When I optimized my models for Accuracy, I was able to get 67%...and I also did this to try to predict each current Justice’s vote

Current Justices |

% of Cases Justice Votes for Petitioner to Win & Accuracy on Test Data from Prediction Model

	%_Petitioner_Vote	Model_Accuracy
clarence_thomas	0.56	0.62
john_g_roberts_jr	0.61	0.68
samuel_a_alito_jr	0.56	0.6
neil_gorsuch	0.55	0.61
brett_m_kavanaugh	0.57	0.58
ruth_bader_ginsburg	0.56	0.65
stephen_g_breyer	0.59	0.67
sonia_sotomayor	0.55	0.65
elena_kagan	0.53	0.58

Interesting! | Similarity in voting amongst the Judges

	clarence_thomas	john_g_roberts_jr	samuel_a_alito_jr	neil_gorsuch	brett_m_kavanaugh	ruth_bader_ginsburg	stephen_g_breyer	sonia_sotomayor	elena_kagan
clarence_thomas	1	0.83	0.84	0.78	0.78	0.61	0.61	0.62	0.63
john_g_roberts_jr	0.83	1	0.86	0.69	0.88	0.67	0.72	0.7	0.7
samuel_a_alito_jr	0.84	0.86	1	0.7	0.9	0.62	0.66	0.62	0.64
neil_gorsuch	0.78	0.69	0.7	1	0.69	0.58	0.55	0.55	0.63
brett_m_kavanaugh	0.78	0.88	0.9	0.69	1	0.63	0.67	0.61	0.67
ruth_bader_ginsburg	0.61	0.67	0.62	0.58	0.63	1	0.85	0.89	0.87
stephen_g_breyer	0.61	0.72	0.66	0.55	0.67	0.85	1	0.85	0.84
sonia_sotomayor	0.62	0.7	0.62	0.55	0.61	0.89	0.85	1	0.86
elena_kagan	0.63	0.7	0.64	0.63	0.67	0.87	0.84	0.86	1

Two other projects during Galvanize

- Project on Predicting Student Outcomes in Virtual Learning environments
- Observing Overall Mortality Rates in the context of Covid-19

Thanks!

Jaime DyBuncio, June 2020
github.com/jdybuncio
Galvanize g119 cohort

Thanks!

Jaime DyBuncio, June 2020
github.com/jdybuncio
Galvanize g119 cohort

Interesting Stats from Oral Arguments

Speaking Direction	Avg Mins	Avg Words	Avg Unique Words
Petitioner to Justice	24	1,910	675
Justice to Petitioner	8	640	300
Respondent to Justice	21	1705	620
Justice to Respondent	7.5	648	300

APPENDIX

Petitioner Always Win's Strategy

- Recall - the percentage of Petitioner wins that I accurately predict, to be **1,00**.
- Precision - the percentage of Petitioner Win predictions are actually cases where they Win, to be **0.63**

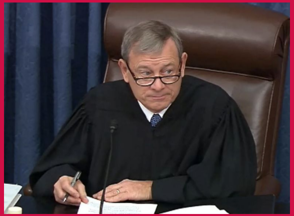
or

Model @ 67% Accuracy

- Recall - the percentage of Petitioner wins that I accurately predict, to be **0.87**.
- Precision - the percentage of Petitioner Win predictions are actually cases where they Win, to be **0.68**

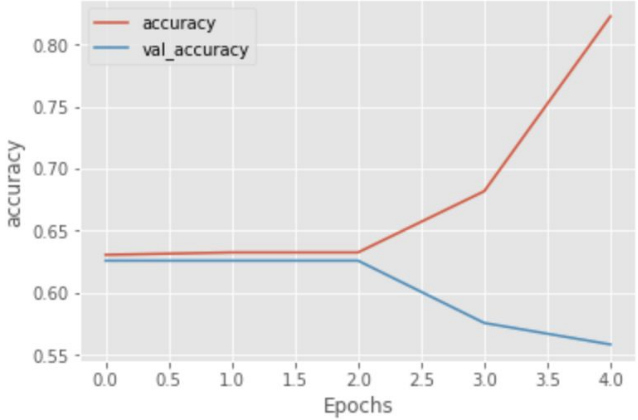
“The secret to successful advocacy, is simply to get the court to ask your opponent more questions.”

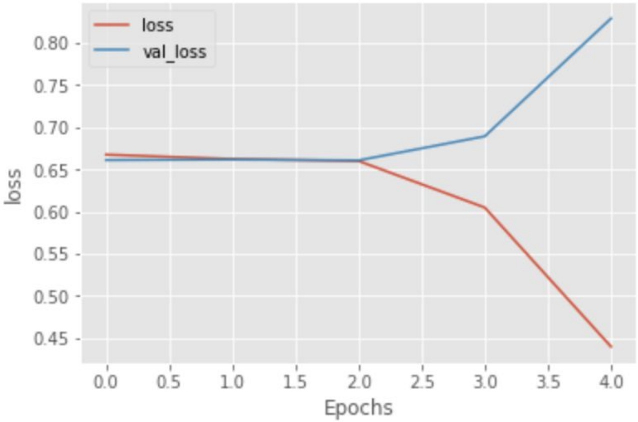
Chief Justice Roberts



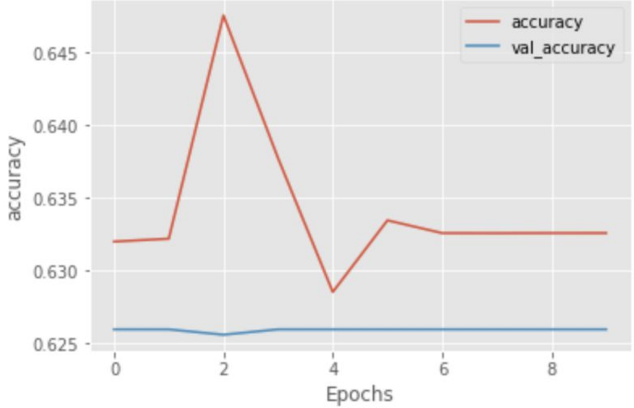
Modeling | LSTM using entire oral argument document to preserve dialogue

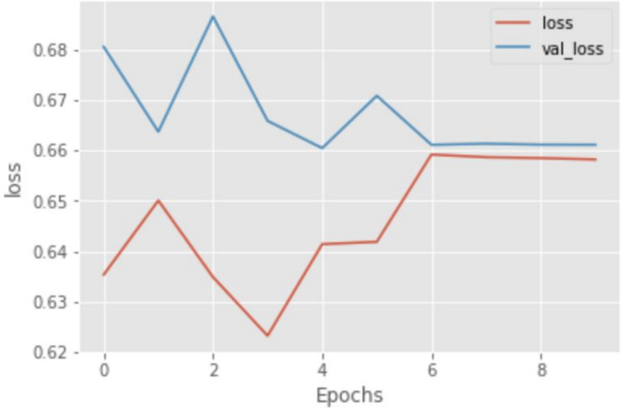
Overfitting





No Fitting





The U.S. Supreme Court in One Slide | Timeline of a Supreme Court Case

Granted

4 Votes needed

≈160 Days
Written Briefing

Argued

Oral Arguments
(timed session)

≈90 Days
Weekly Conference

Decided

Opinions Read

Central Parties Involved: Petitioner, Respondent, & Supreme Court Justices

Project Goal & Application

Goal

Using Oral Arguments, create a model to predict whether a Petitioner will win a case or not

Application

1. Influence Petitioner Strategy during oral arguments
2. Craft strategy during waiting period b/w Argument & Decision (i.e. Op-eds)

Creating a Dataset | Features

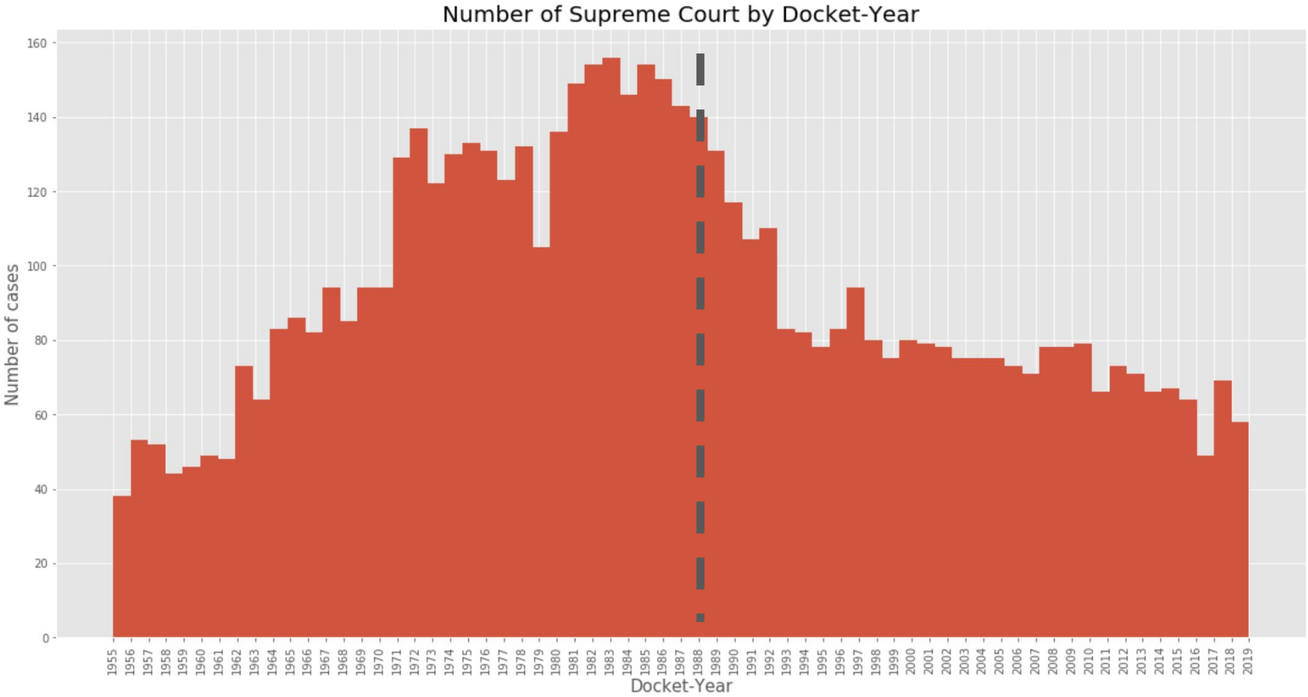
To-Justices Features

- Total Words
- Time spent talking
- Number of times interrupted
- Document of words
- *[Difference across parties]*

From-Justices Features

- Total Words
- Time spent talking
- Number of Questions Asked
- Document of words
- *[Difference across parties]*

The U.S. Supreme Court in 30 Seconds | Annual Caseload



<u>Sample</u>

- 6,000 Supreme Court Cases from 1955-Current w/ One Oral Argument
  - Most recently, Supreme Court takes <u>**60-70 cases per Year**</u>

Modeling | Workflow

5,567 Supreme Court Cases

Train Data

4,453 Cases

Test Data

1,114 Cases

5 - fold Cross Validation

Train

Validation

- Center & Scaled Features
- Used Supervised Learning Algorithms to Train Models
- Gridsearch parameters to optimize models for **F1-Score**

- Using model w/ **highest avg F1-Score on Validation Set:**
  - Predict outcome of Test Data and compare to actuals to evaluate final model performance

Modeling | LSTM using entire oral argument document to preserve dialogue

Data Preprocessing for LSTM: Integer Based Representation of Oral Argument

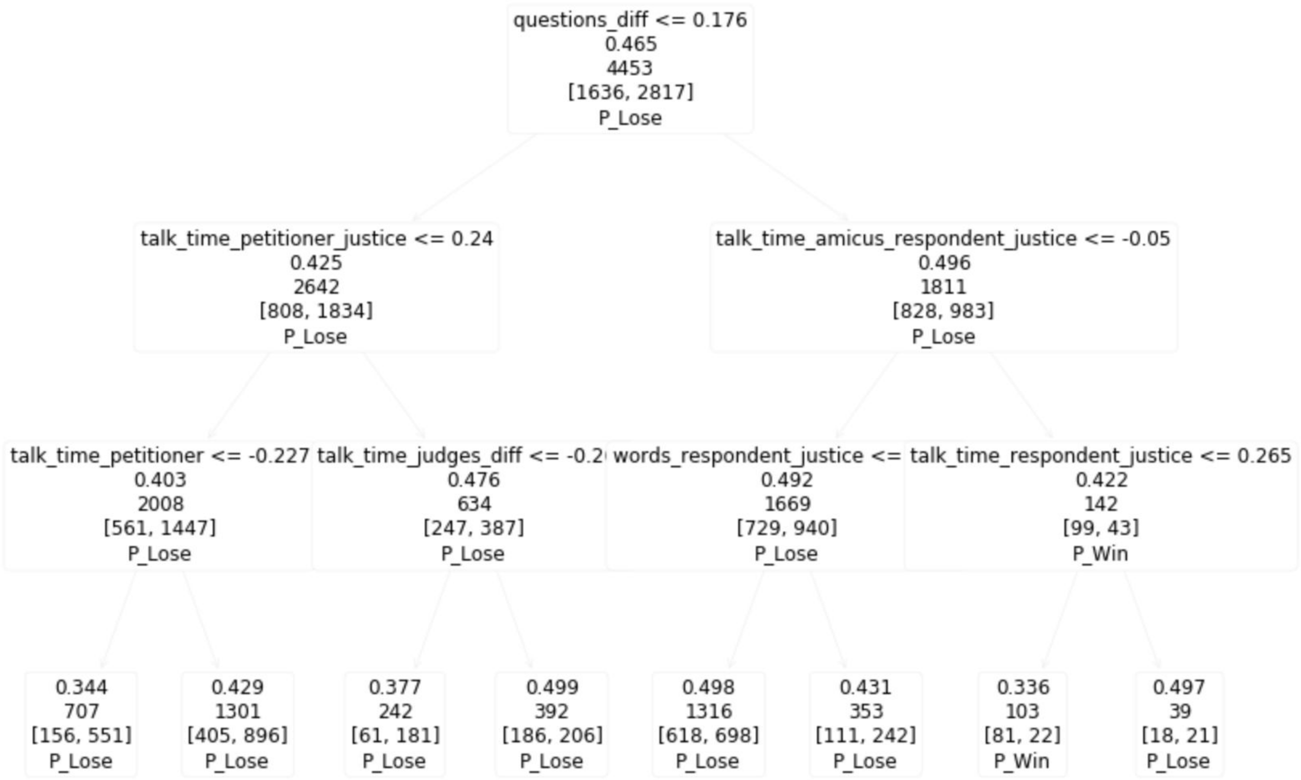
```
'number 21 docket united states america versus twin city power company et al
mr spritzer mr chief justice honors case involves issue compensation arising
eminent domain proceedings instituted united states 1947 western district so
uth carolina taking made necessary building dam savannah river forms border
south carolina georgia location known clarks  clark hill clark hill dam vita
l part comprehensive plan pictured map reproduced engineers engineers studie
s clark hill dam central part savannah river basin project located point rig
ht tip brown area represent reservoir basin project declared purposes flood
control navigation improvement hydroelectric development condemned lands who
se value question proceeding located miles upstream governments dam clark hi
ll flooded result backing river dam otherwise stated part reservoir basin ta
king fast lands adjacent river respondent course entitled compensation respo
ndent however awarded believe special increment value entitled whether right
view  issu'

...

[153, 2133, 4292, 86, 30, 3221, 234, 7696, 274, 135, 200, 2285, 3874, 9, 807
4, 9, 124, 19, 719, 4, 1065, 38, 611, 2219, 5035, 3461, 443, 3649, 86, 30, 3
047, 2151, 37, 1089, 928, 390, 36, 299, 1235, 3277, 9570, 896, 2079, 2660, 1
089, 928, 875, 2405, 825, 1, 3009, 3425, 3009, 3425, 3277, 3668, 95, 3049, 2
91, 1, 2672, 5526, 3166, 3166, 3151, 3009, 3425, 3277, 1269, 95, 9570, 896,
6377, 1718, 2010, 31, 14, 5112, 1593, 240, 1196, 6292, 6377, 1718, 2088, 27
6, 3263, 415, 3262, 4974, 8927, 1612, 4806, 872, 1244, 484, 13, 272, 2010, 1
505]
```

Word - Number Map

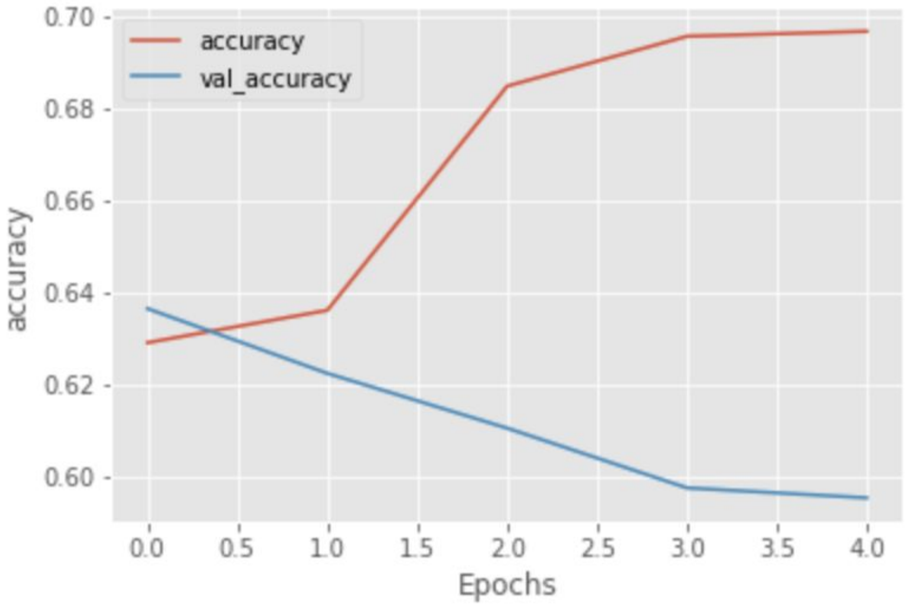
```
{'<OOV>': 1,
 'court': 2,
 'would': 3,
 'case': 4,
 'think': 5,
 'well': 6,
 'state': 7,
 'one': 8,
 'mr': 9,
 'say': 10,
 'thats': 11,
 'yes': 12,
 'question': 13,
 'right': 14,
 'could': 15,
 'honor': 16,
 'statute': 17,
 'dont': 18,
 'justice': 19,
 'may': 20}
```

*OOV stand for unseen words (catch all for non top 10k most frequent words)

APPENDIX


questions_diff <= 0.176
0.465
4453
[1636, 2817]
P_Lose
talk_time_petitioner_justice <= 0.24
0.425
2642
[808, 1834]
P_Lose
talk_time_amicus_respondent_justice <= -0.05
0.496
1811
[828, 983]
P_Lose
talk_time_petitioner <= -0.227
0.403
2008
[561, 1447]
P_Lose
talk_time_judges_diff <= -0.2
0.476
634
[247, 387]
P_Lose
words_respondent_justice <=
0.492
1669
[729, 940]
P_Lose
talk_time_respondent_justice <= 0.265
0.422
142
[99, 43]
P_Win
0.344
707
[156, 551]
P_Lose
0.429
1301
[405, 896]
P_Lose
0.377
242
[61, 181]
P_Lose
0.499
392
[186, 206]
P_Lose
0.498
1316
[618, 698]
P_Lose
0.431
353
[111, 242]
P_Lose
0.336
103
[81, 22]
P_Win
0.497
39
[18, 21]
P_Lose

Modeling | LSTM

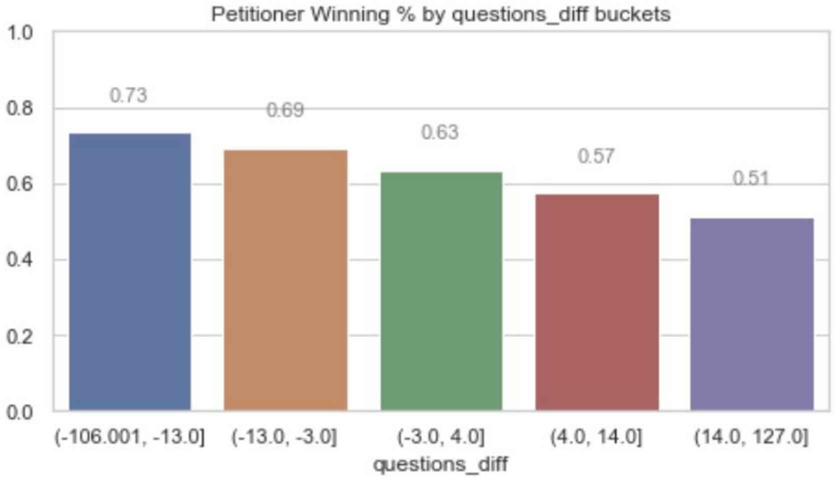


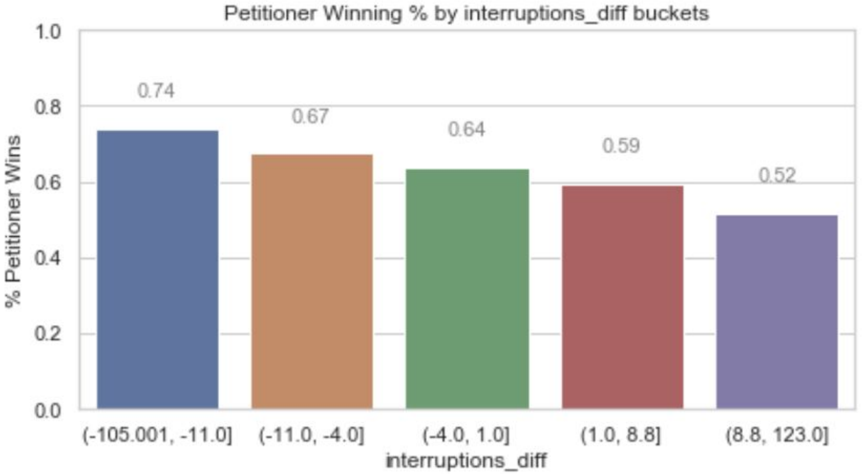

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_3 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
lstm_6 (LSTM)                (None, None, 50)          23000
_________________________________________________________________
lstm_7 (LSTM)                (None, 10)                2440
_________________________________________________________________
dropout_1 (Dropout)          (None, 10)                0
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 11
=================================================================
Total params: 665,451
Trainable params: 665,451
Non-trainable params: 0
_________________________________________________________________
```

“Liptak and Dahlia Lithwick, Slate’s Supreme Court writer, both emphasized the importance of attending oral arguments rather than just parsing transcripts. **Crossed arms, rolled eyes and tone of voice can be telling.** And the computer is ignorant of all of that”

- Oliver Roedger, Fivethirtyeight

EDA | Interesting Features

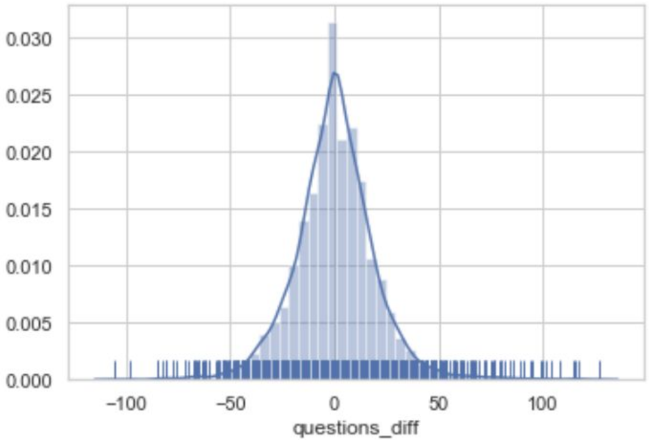




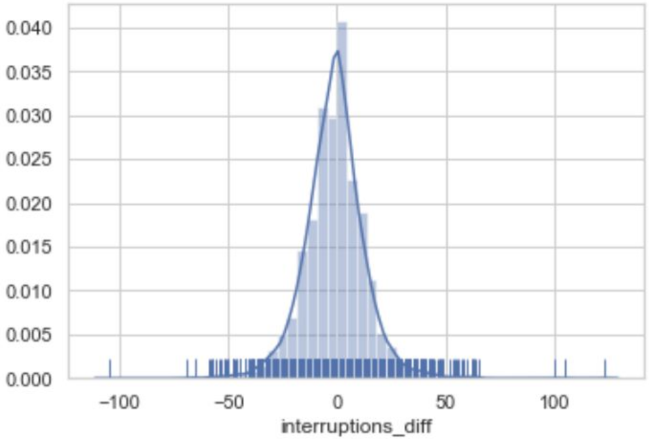
Note: Petitioner Win Rate is always above 50%

APPENDIX

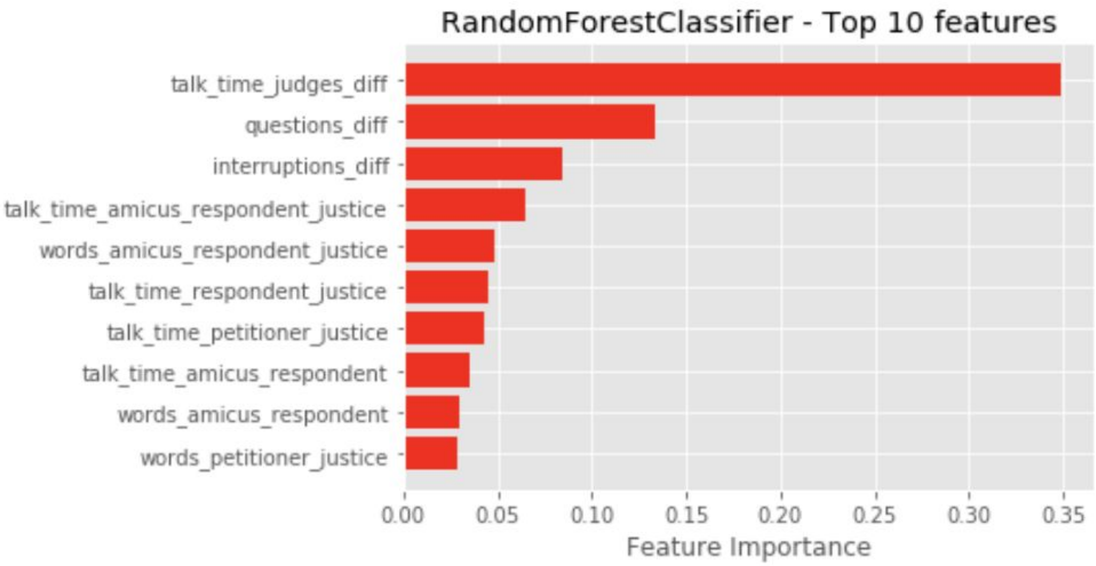
```
sns.distplot(df.questions_diff, kde=True, rug=True);
```




```
sns.distplot(df.interruptions_diff, kde=True, rug=True);
```



APPENDIX



1. **Tested: Logistic, RF, GBoost models - GridSearch over 1,400 models**
2. **Criterion** = Gini
3. **Trees:** 1,000
4. **Max Features** = 0.25
5. **Max Depth** = 3
6. **Min Samples Split** = 3
7. **Min Samples Leaf** = 10
8. **Features:** Difference in Judge Talk Time, Difference in Questions, Difference in Interruptions

APPENDIX

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

True Positive + False Positive = Total Predicted Positive

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

True Positive + False Negative = Actual Positive

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$= \frac{True\ Positive}{Total\ Predicted\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$= \frac{True\ Positive}{Total\ Actual\ Positive}$$

Precision - % of Predicted Petitioner Wins actually are cases they Win

Recall - % of Actual Petitioner Wins model is able to Project