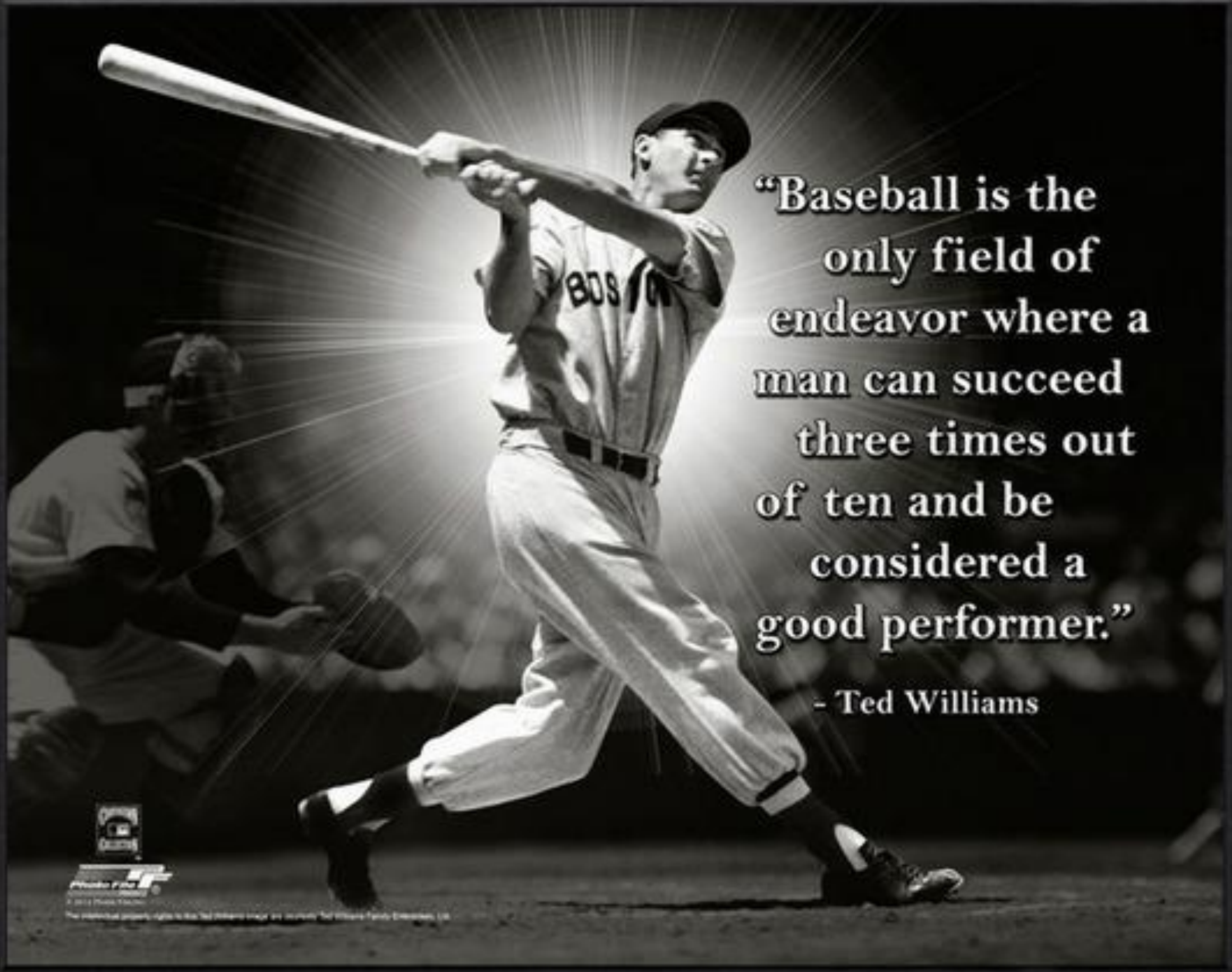
Predicting the Next Pitch

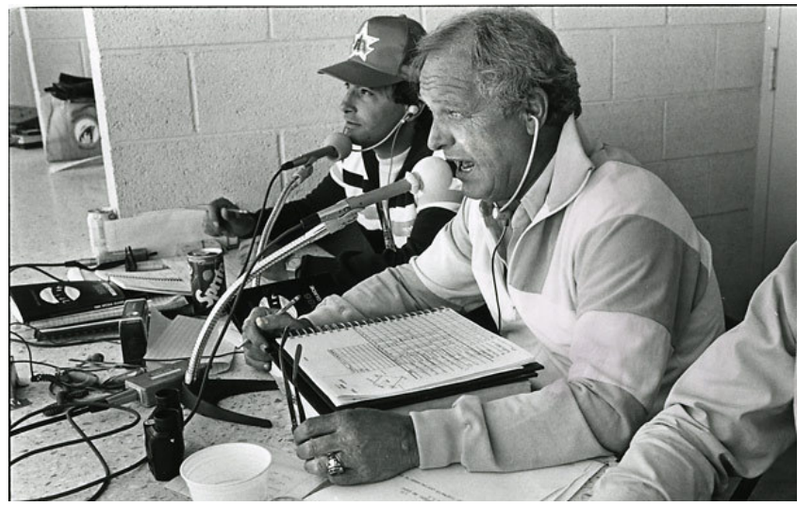
Alex Bell
Metis Data Science


"Baseball is the only field of endeavor where a man can succeed three times out of ten and be considered a good performer."
- Ted Williams

“Guessing what the pitcher is going to throw is 80 percent of being a successful hitter. The other 20 percent is just execution.”
Hank Aaron

- **Predictions can help improve hitter performance**
- **Predictions can help pitchers decrease their predictability**
- **Predictions can help teams increase their win totals, win championships, and rake in the cash!**

The Data:



- Characteristics of 720,000+ Major League Baseball Pitches 2016 - 2018 from baseballsavant.com

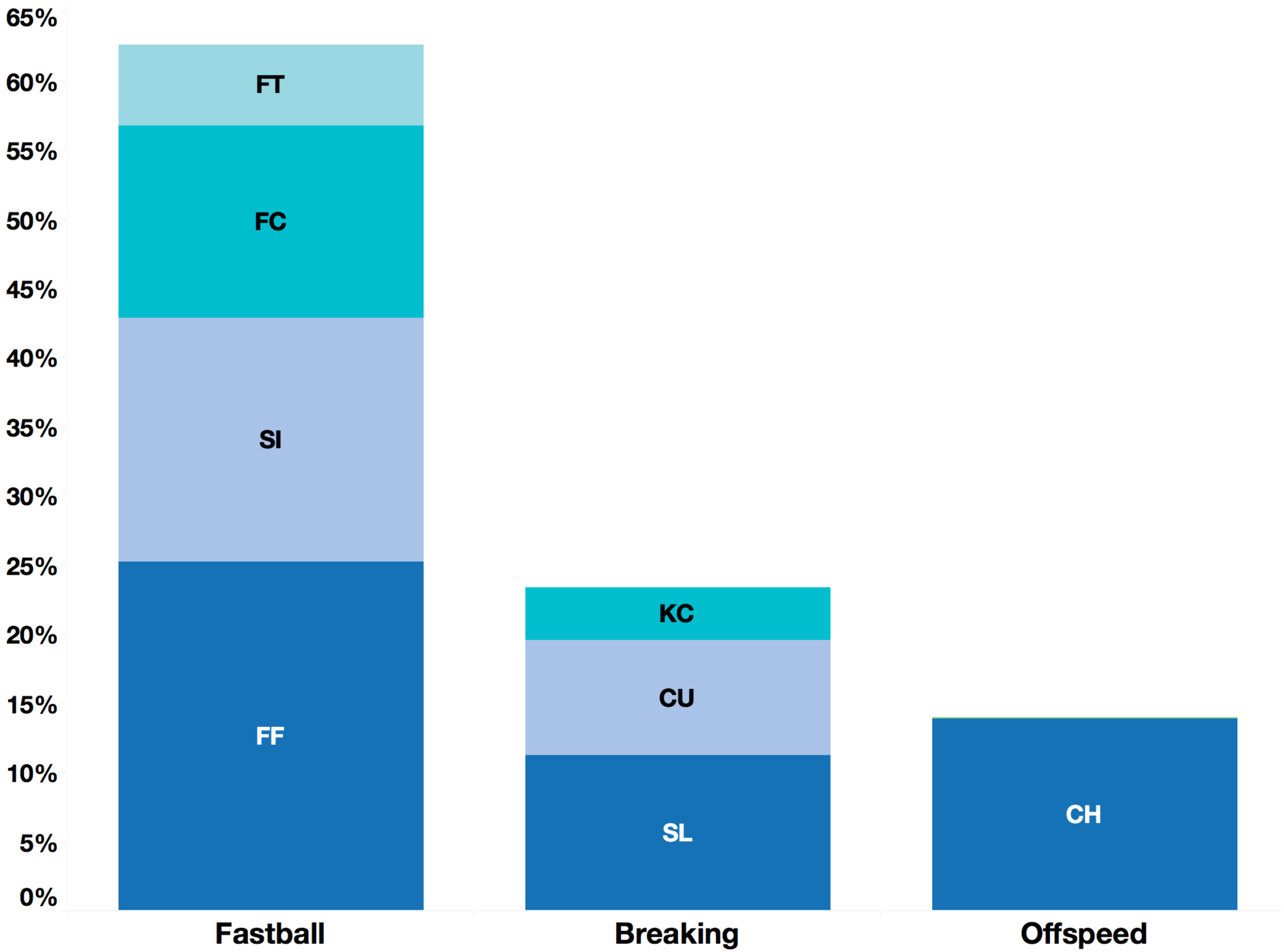
- Generated Data Tables for Individual Pitchers

- Tables Contain 88 Features for Classifying Pitch Types

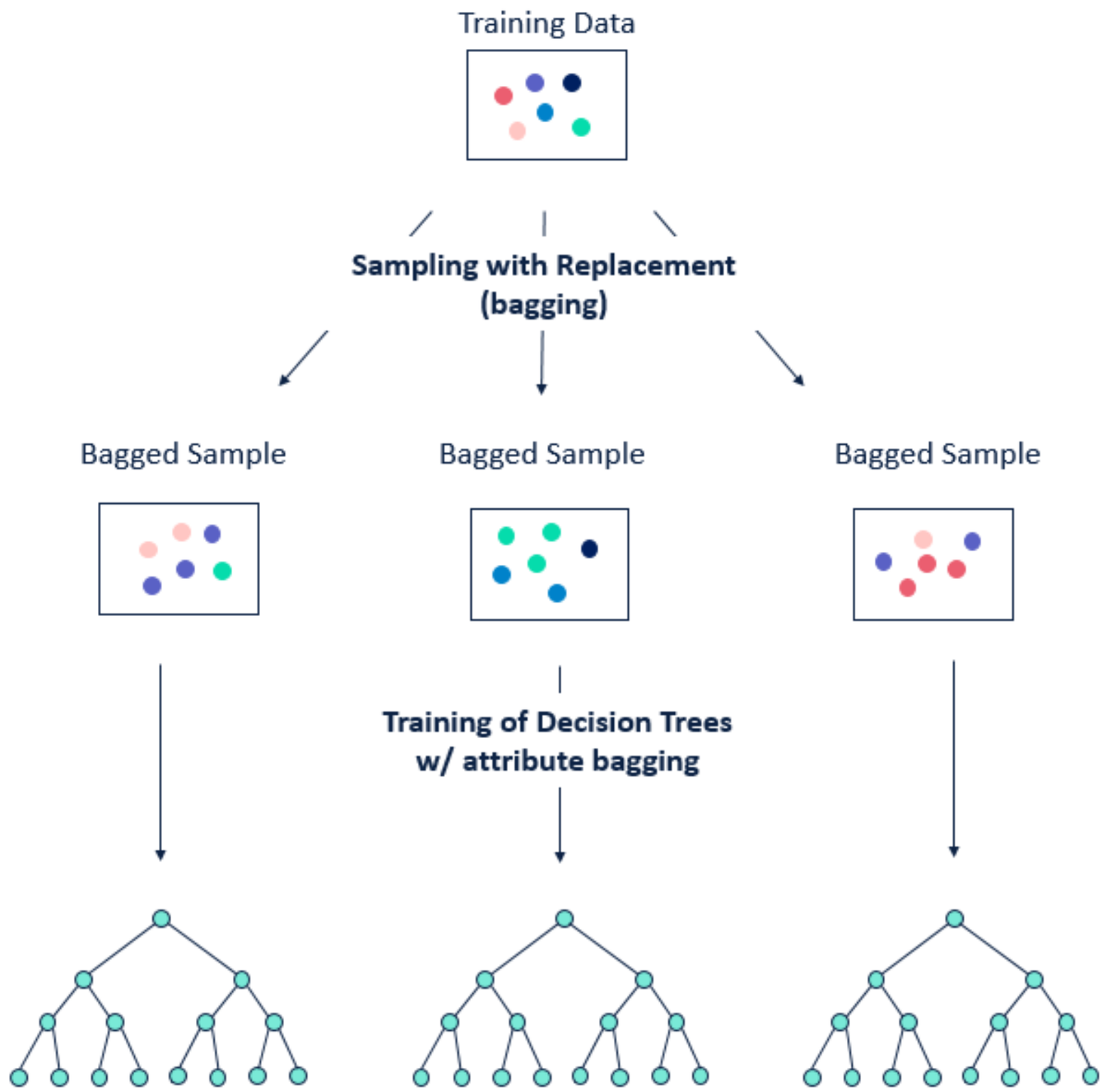
- Pitches Grouped into Multiple Classes:

  Fastball Breaking Ball Off-Speed

Distribution of Pitch Classes



Random Forest Classification Method

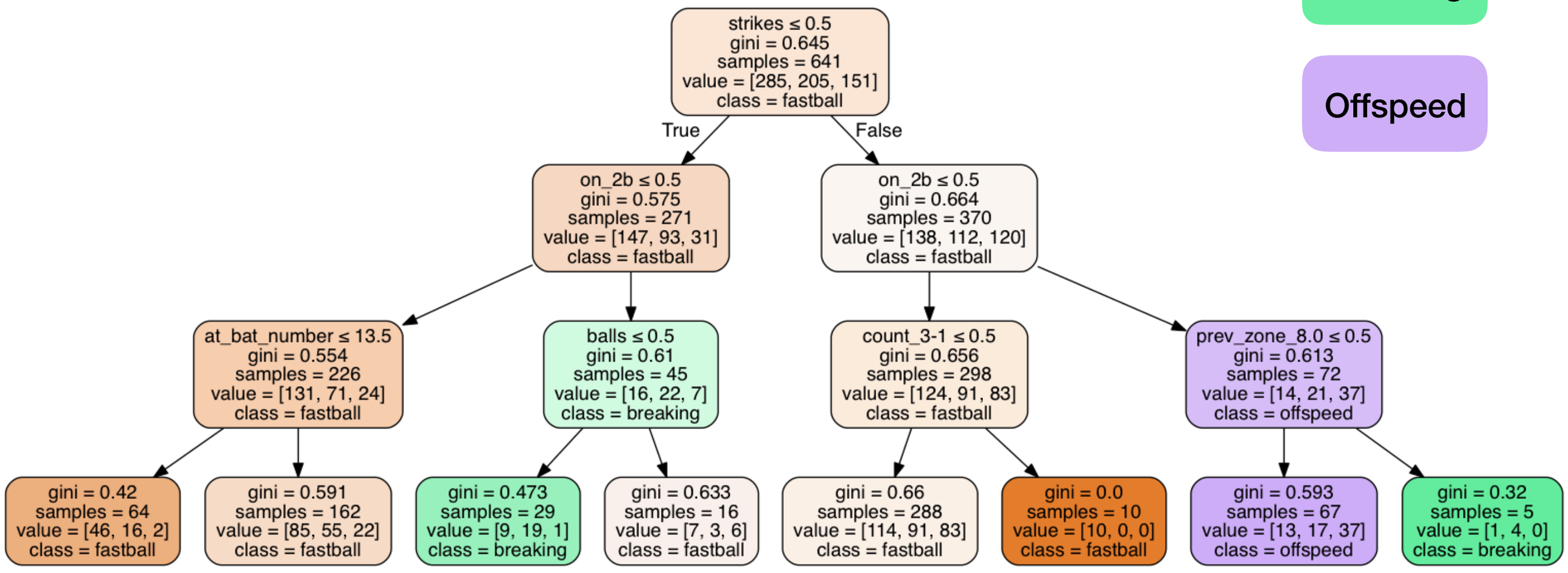


Sample Decision Tree From Random Forest Model

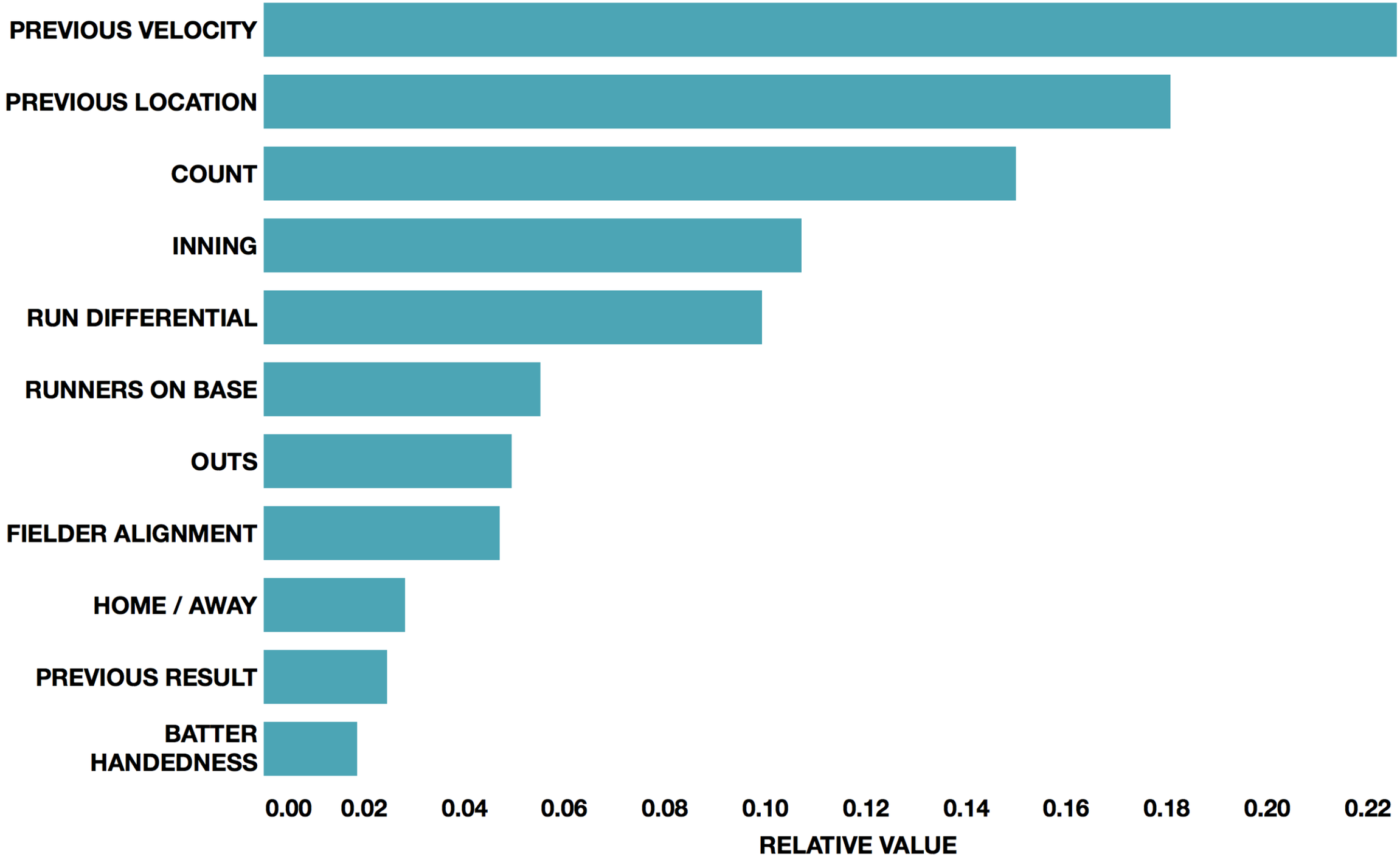
Fastball

Breaking

Offspeed



Feature Importance

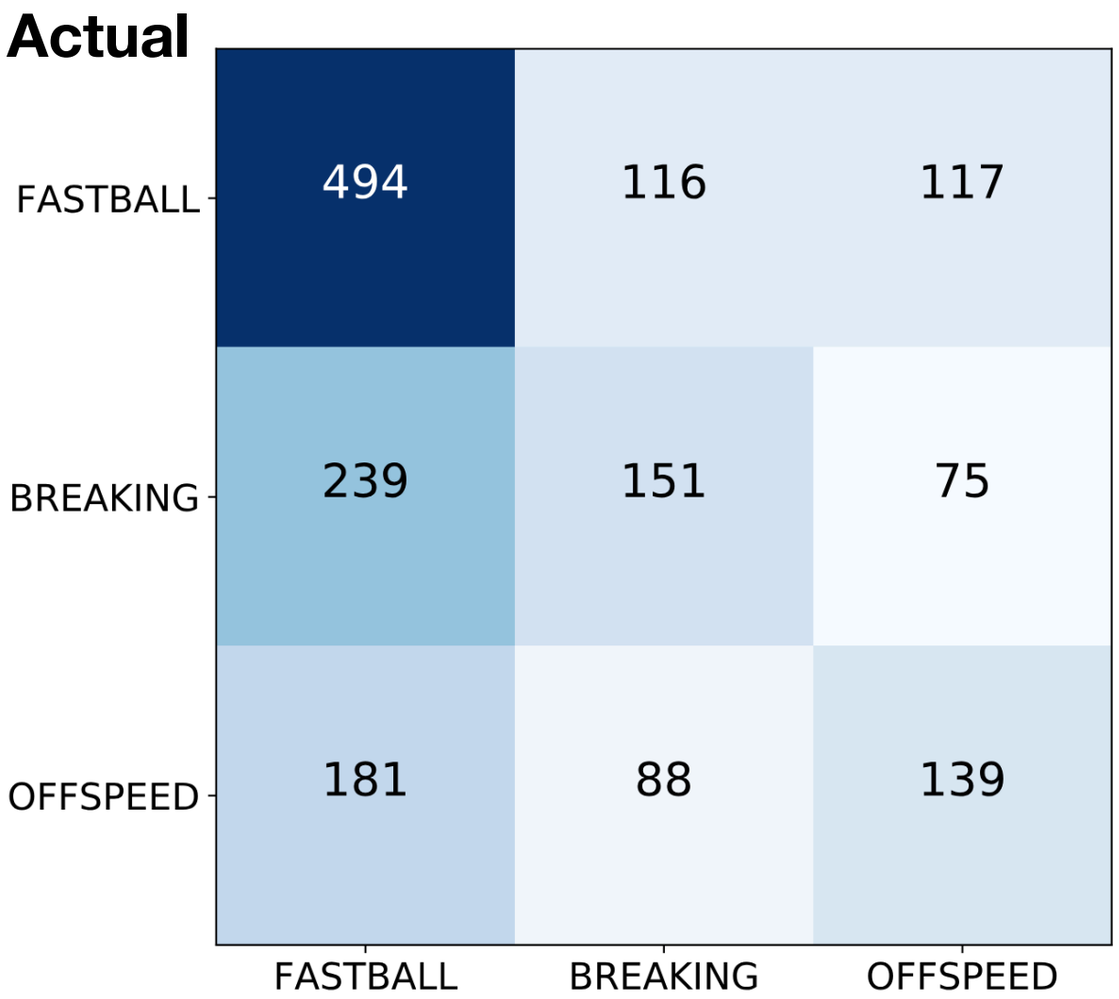


Felix Hernandez Pitch Types

Fastball Accuracy 68%

Overall Accuracy 49%

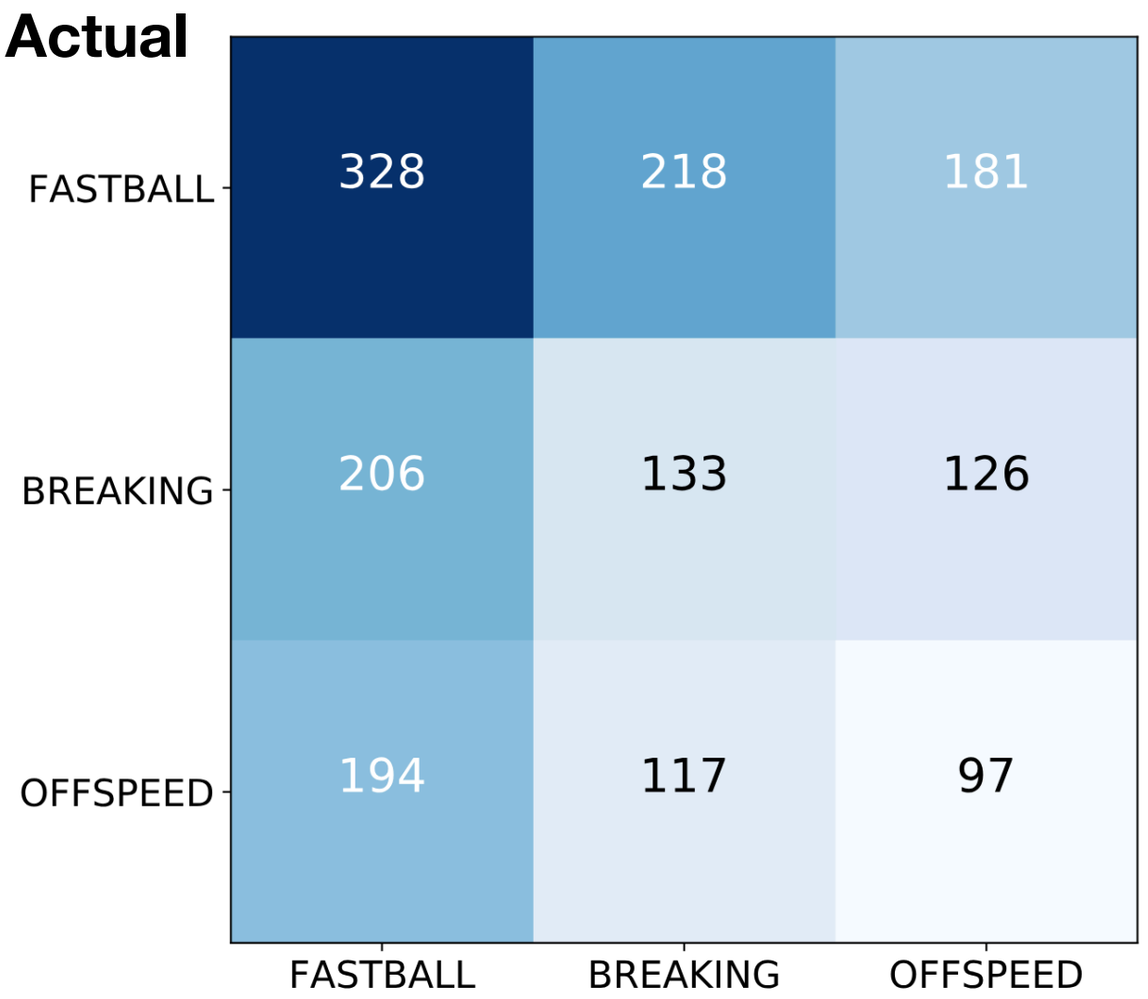
Random Forest Model



Fastball Accuracy 45%

Overall Accuracy 35%

Stratified Naive Model

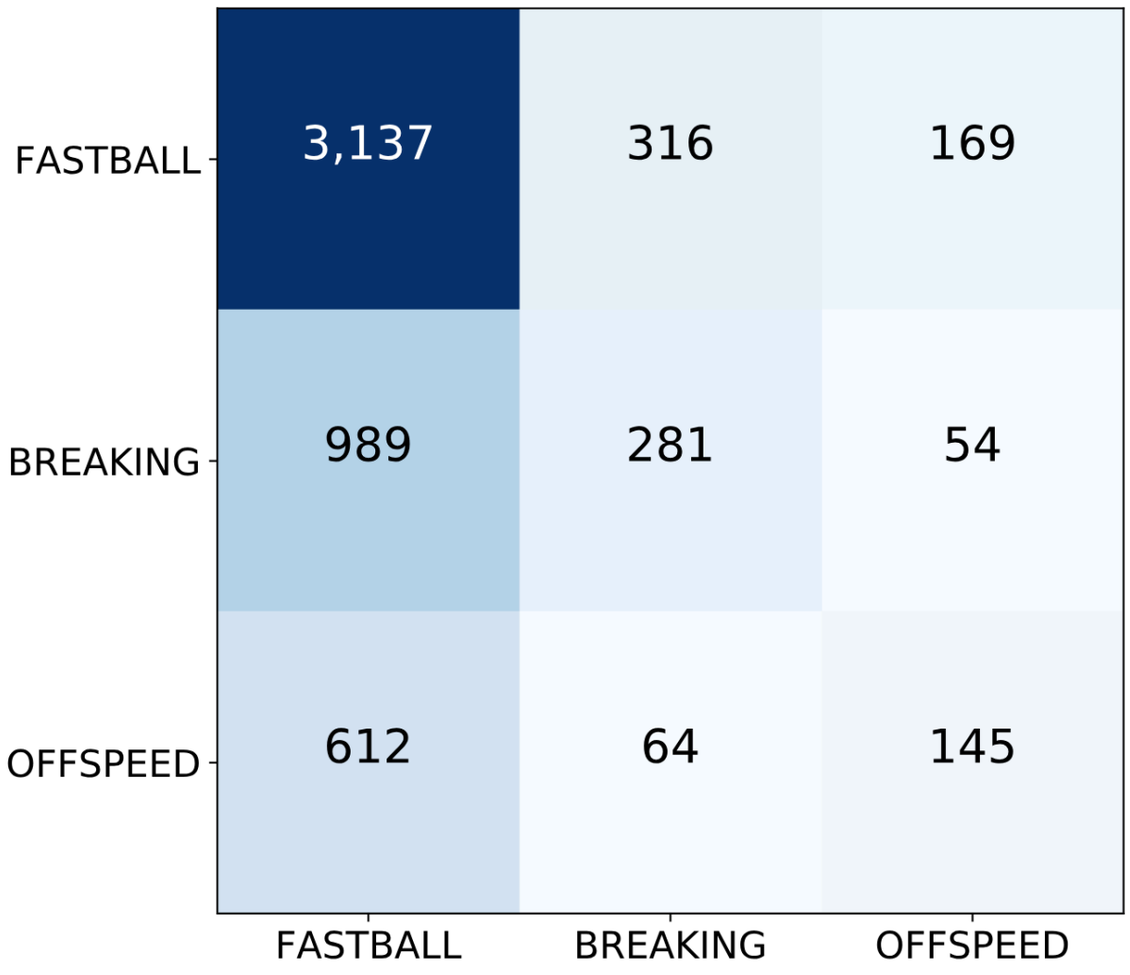


Seattle Mariners 2018

Fastball Accuracy 87%

Overall Accuracy 62%

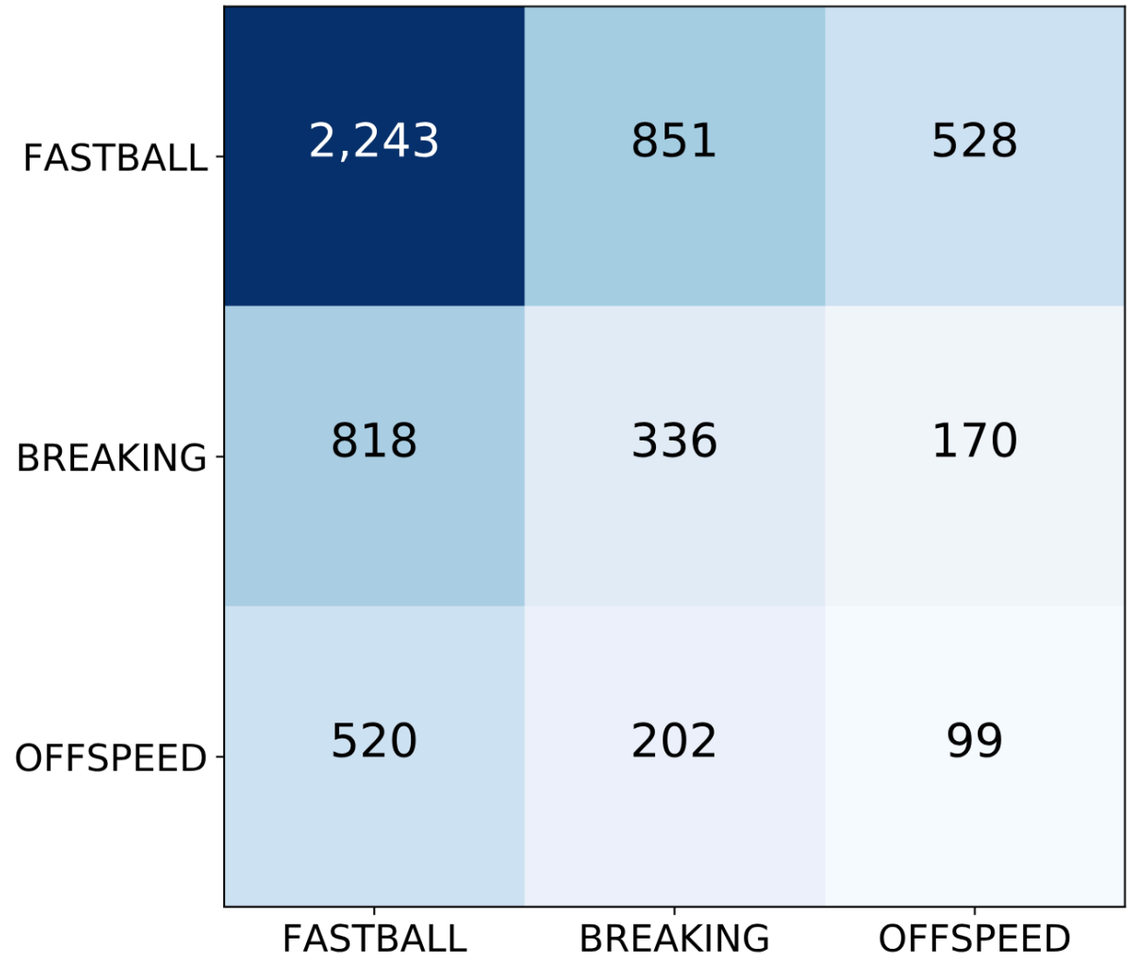
Random Forest Model



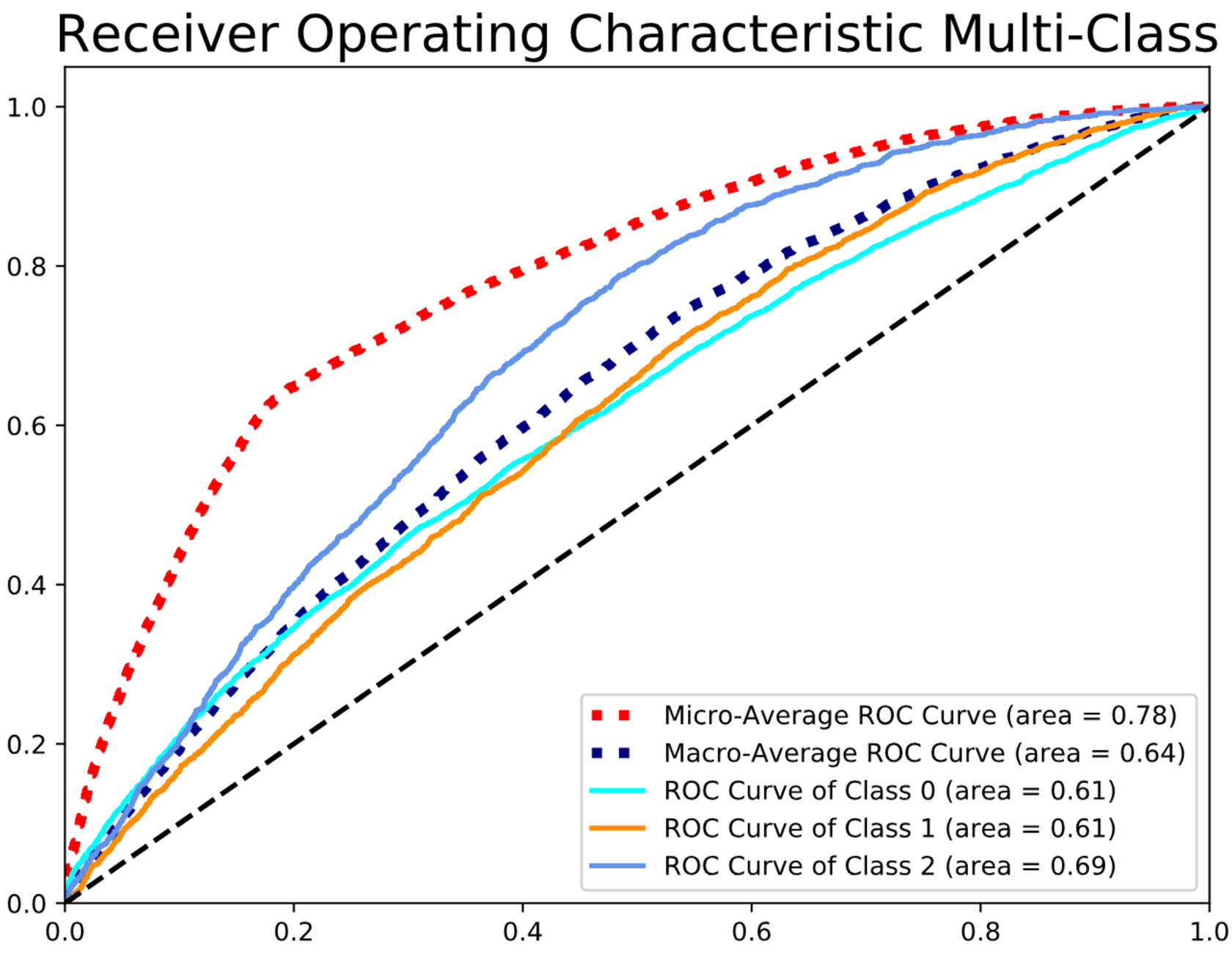
Fastball Accuracy 62%

Overall Accuracy 47%

Stratified Naive Model



One vs All Methodology



Next Steps

- Combine Data from Specific Batter-Pitcher Match-ups
- Expand Predictions to Pitch Location Zones
- Develop Application to Make Live In-Game Predictions

Thank You

Questions?

	precision	recall	f1-score	support
0	0.54	0.68	0.60	727
1	0.43	0.32	0.37	465
2	0.42	0.34	0.38	408
micro avg	0.49	0.49	0.49	1600
macro avg	0.46	0.45	0.45	1600
weighted avg	0.48	0.49	0.48	1600

	precision	recall	f1-score	support
0	0.45	0.45	0.45	727
1	0.33	0.33	0.33	465
2	0.24	0.23	0.23	408
micro avg	0.36	0.36	0.36	1600
macro avg	0.34	0.34	0.34	1600
weighted avg	0.36	0.36	0.36	1600

	precision	recall	f1-score	support
0	0.66	0.87	0.75	3622
1	0.43	0.21	0.28	1324
2	0.39	0.18	0.24	821
micro avg	0.62	0.62	0.62	5767
macro avg	0.49	0.42	0.43	5767
weighted avg	0.57	0.62	0.57	5767

	precision	recall	f1-score	support
0	0.62	0.62	0.62	3622
1	0.23	0.25	0.24	1324
2	0.15	0.14	0.14	821
micro avg	0.47	0.47	0.47	5767
macro avg	0.33	0.33	0.33	5767
weighted avg	0.47	0.47	0.47	5767