A Machine Learning Approach to MLB Catcher Framing

Nathan Hemenway and Matthew Boyd

12/10/2021

Introduction

- “Catcher framing is the art of a catcher receiving a pitch in a way that makes it more likely for an umpire to call it a strike – whether that’s turning a borderline ball into a strike, or not losing a strike to a ball due to poor framing.” - MLB.com Glossary

Motivation

- Baseball catchers can influence the call of a ball or strike on how they catch the ball
- Some catchers are better than others at this skill
- Baseball teams are aware of this and are acquiring players good at this skill to win more games
- We want to quantify the best catcher's at framing for the 2021 season
- There are several factors that influence whether a pitch will be a strike or ball
- Catchers getting more strikes translates to more outs, and fewer runs for the opposing team
- https://baseballsavant.mlb.com/sporty-videos?playId=273b22a4-f522-4009-87d1-14176772182d

Data

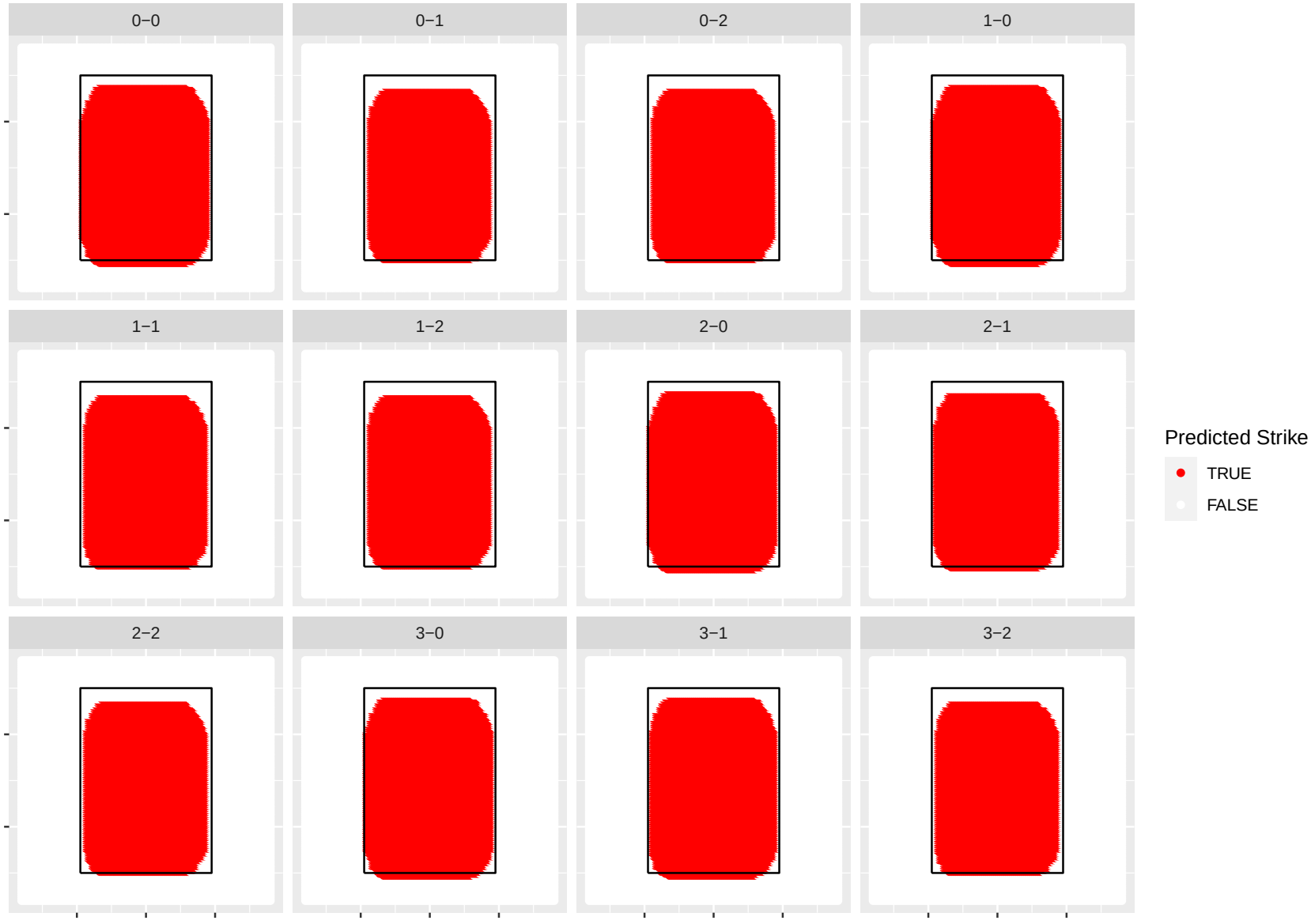
- 2021 pitch data scraped from Baseball Savant through baseballr package
- ~700,000 rows (each for a single pitch)
- Wanted to look at pitches that were not swung at by the batter (called strike or ball)
- ~350,000 rows remain

Variables

- We used:
- Pitch type, pitch release speed, position, and spin rate
- Whether or not the pitcher and batter are right or left handed
- Count, number of outs during the at-bat, and inning number
- Where the pitch landed
- Whether the game was played home or away
- How tall the batter is

Example

- Strike Probability by Location and Count



Logistic Regression Model

```
##           true
## pred           0      1
##   ball    112696  52821
##   strike    5057   5372

## [1] 0.6710468
```

- The variables with significant results were:
- Pitch type, release speed, z release position
- Whether the batter is a lefty/righty
- Count
- Where the pitch landed
- Most of the innings
- Batter height
- Whether it was a home game

Ridge Regression

```
##      true
## pred      0      1
##    0 114268  53642
##    1   3769   4267
```

```
## [1] 0.673701
```

LASSO

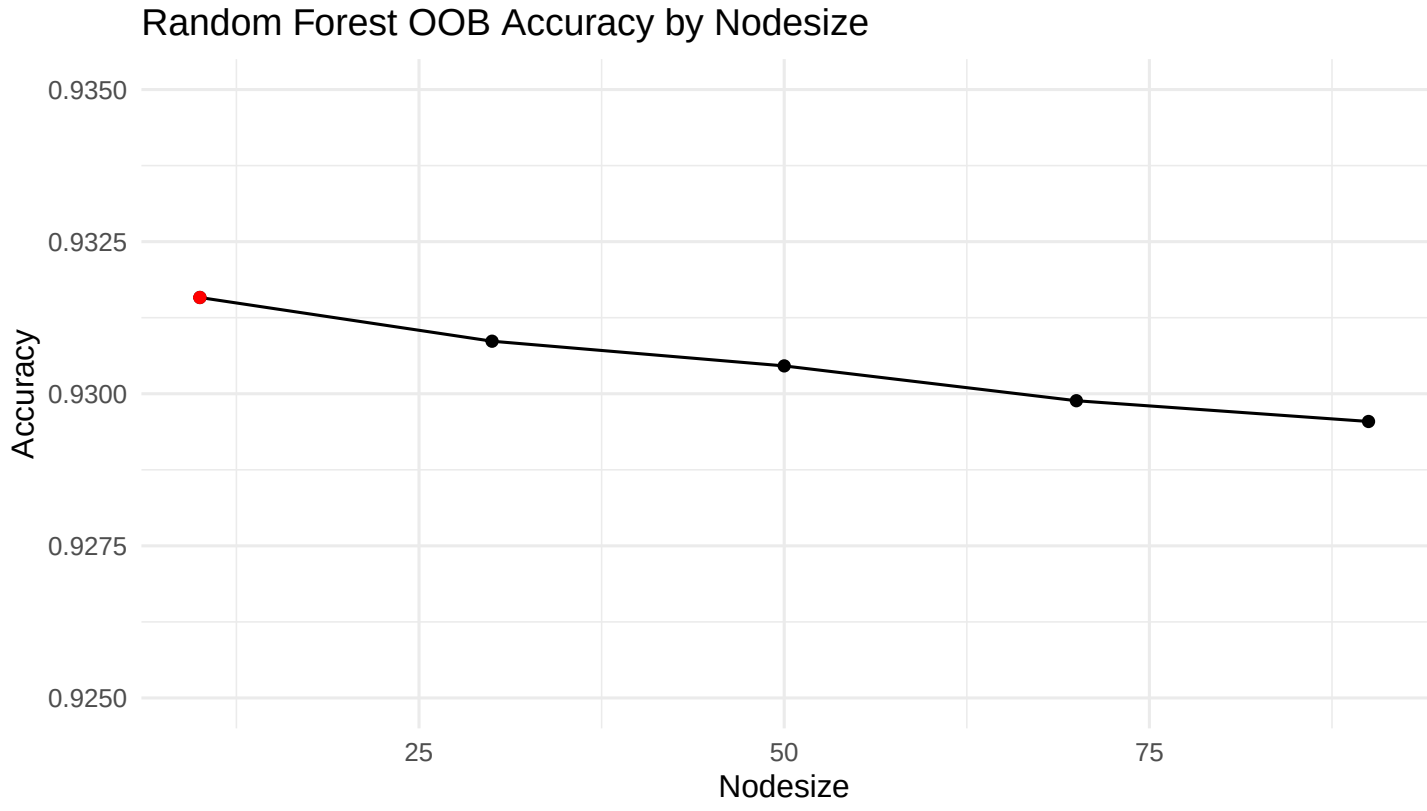
```
##      true
## pred      0      1
##    0 115220  54523
##    1   2817   3386
```

```
## [1] 0.6741046
```

- Some of the non-zero coefficients include:
- Pitch type
- Count
- Outs and Inning
- Batter height
- Whether the game is home or away
- Where the pitch landed

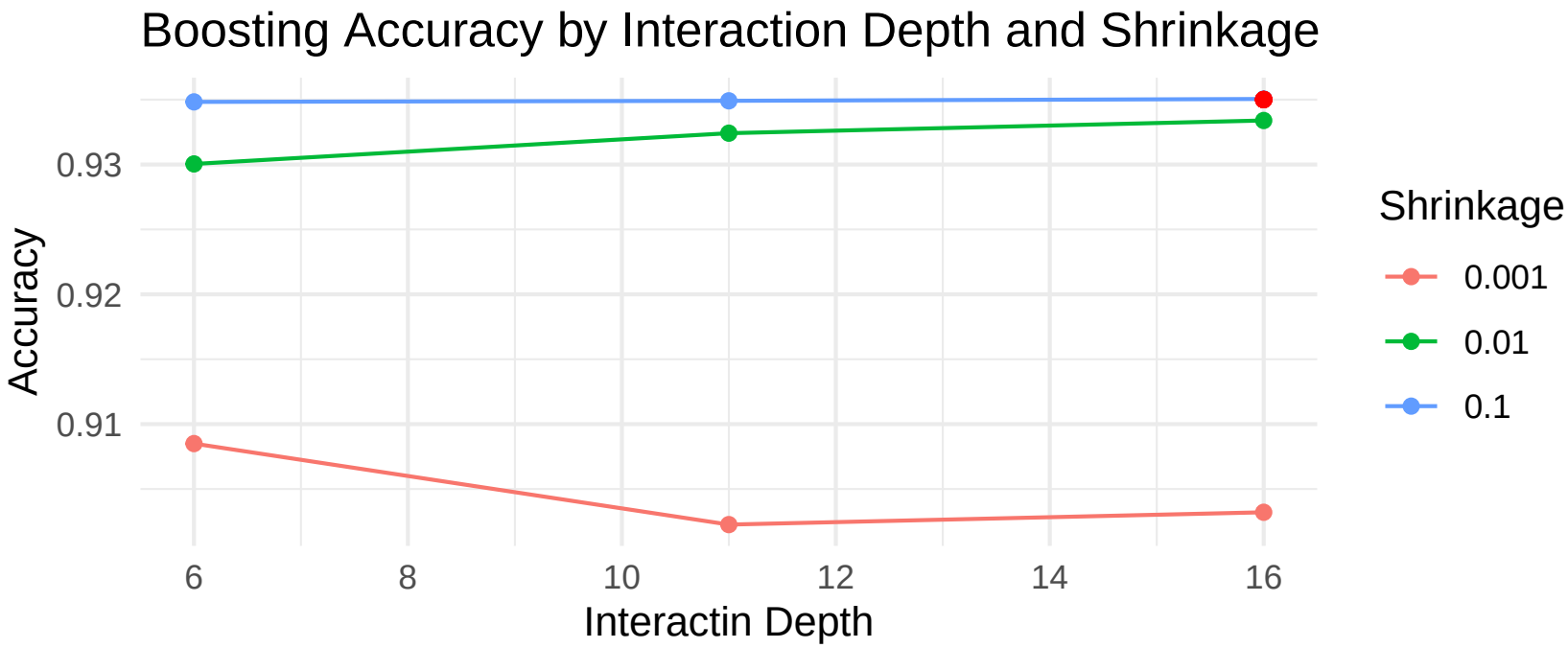
Random Forest

- Tested several Random Forest models
  - Nodesize: 10, 30, 50, 70, 90 | # of Trees: 500
  - Compared Out-Of-Bag Error



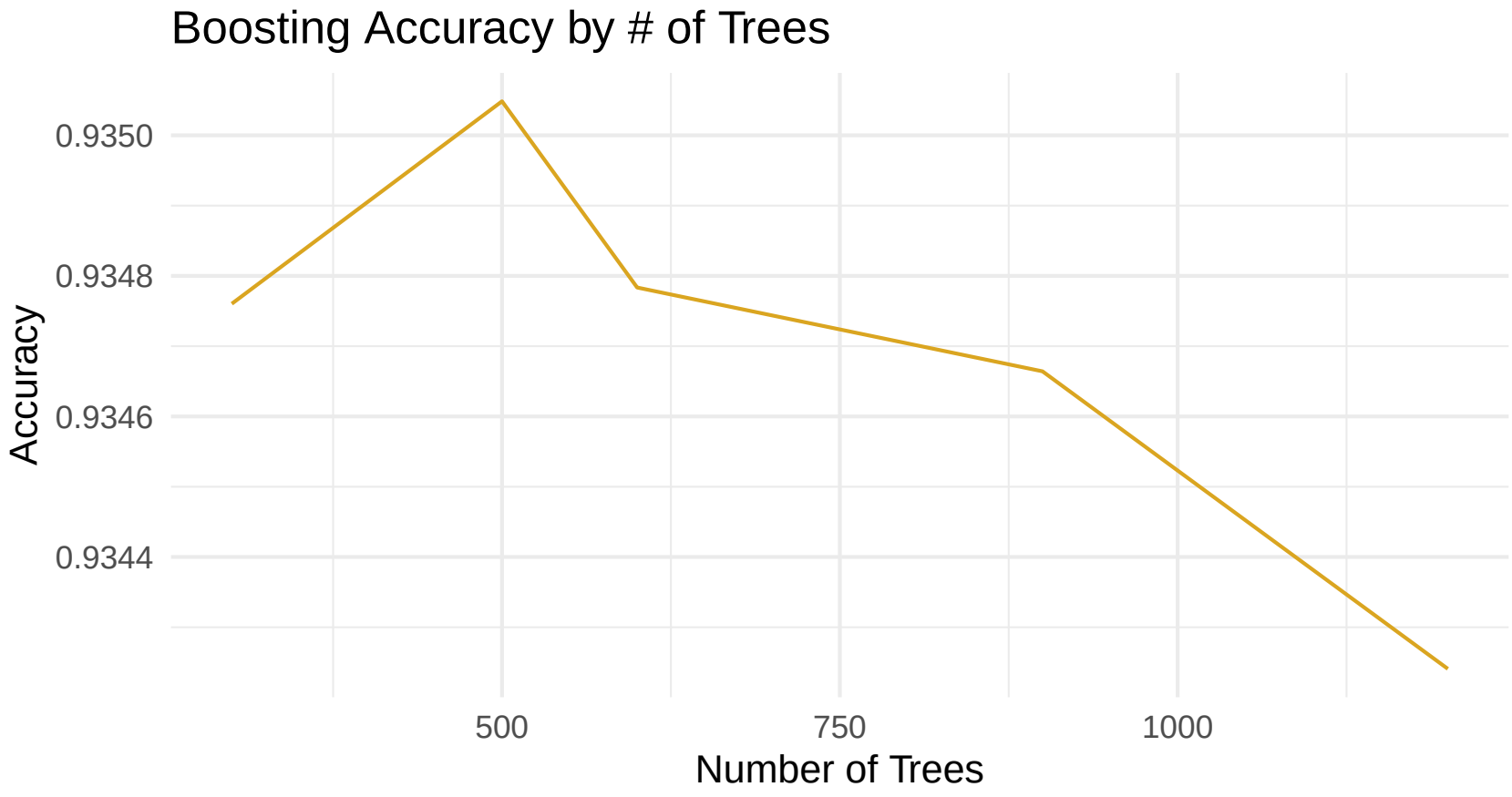
Boosting

- 3 Fold Cross Validation on Boosting
  - Interaction depth: 6, 11, 16
  - Shrinkage: 0.1, 0.01, 0.001
  - Number of Trees: 500
  - Compared CV accuracy
- 27 models took over 5 hours to run



Boosting Part 2

- Only tuned shrinkage and interaction depth
- Now tune trees with interaction depth of 16 and shrinkage of 0.1
- Trees: 300, 500, 600, 900, 1200

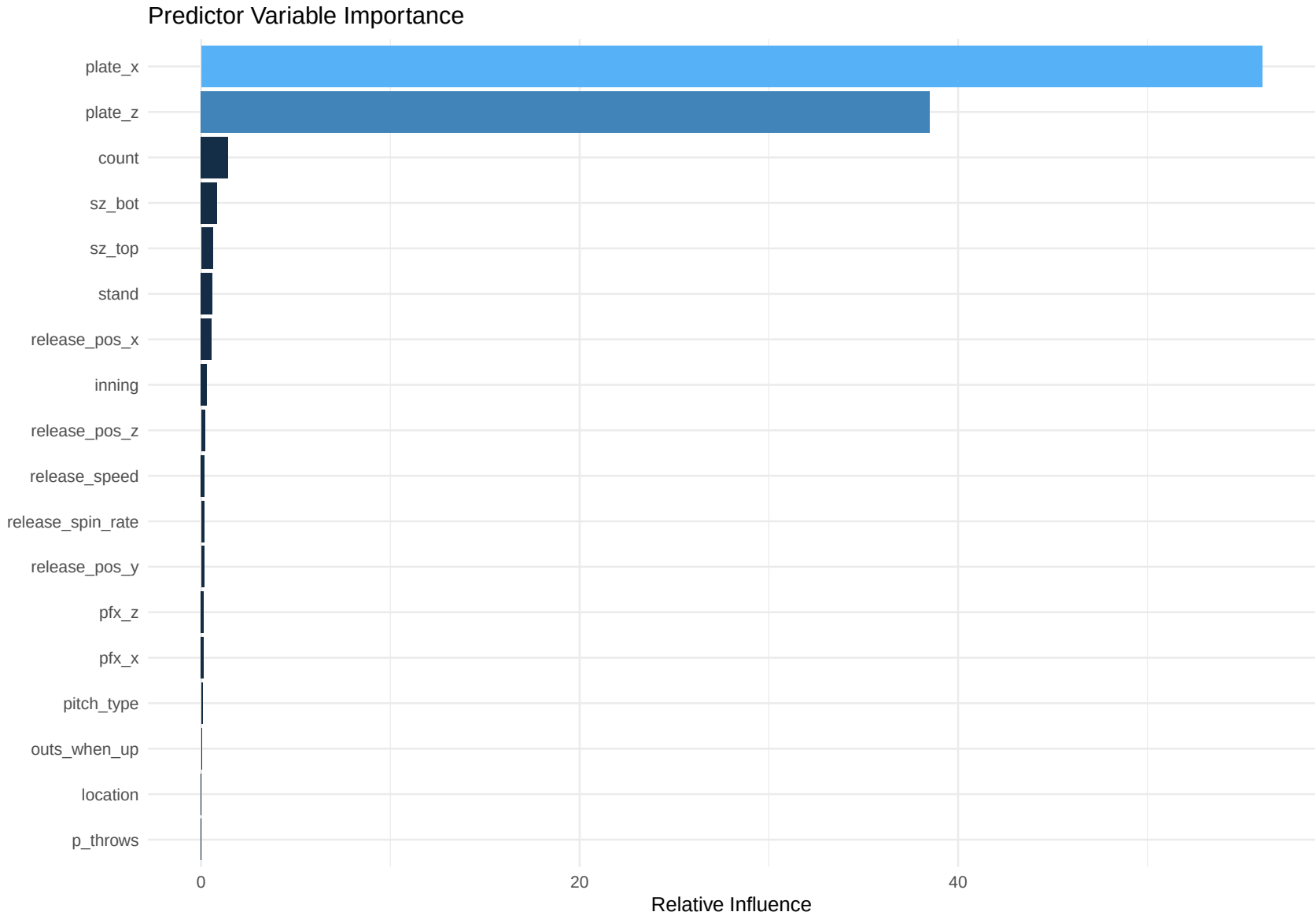


Model Comparison

- Boosting with interaction depth of 16, shrinkage of 0.1, and 500 trees has the highest accuracy

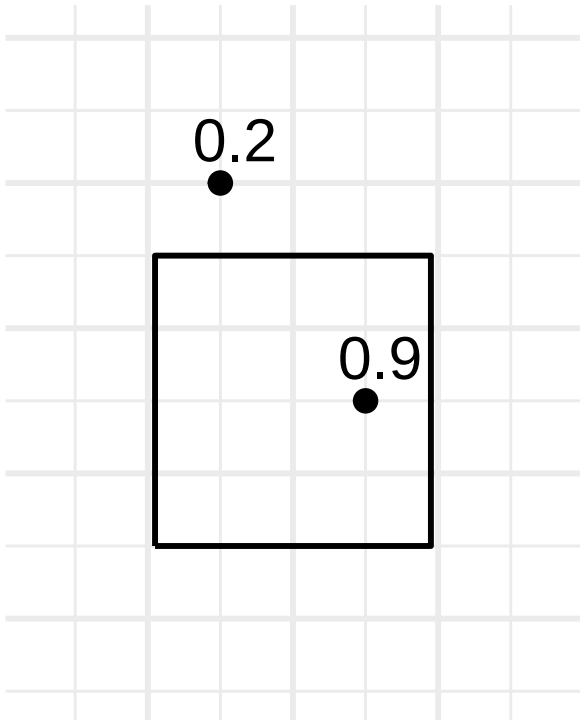
model	accuracy	parameter
boosting	0.9350484	500
boosting	0.9347835	600
boosting	0.9347604	300
boosting	0.9346642	900
boosting	0.9342408	1200
rf	0.9315823	10
rf	0.9308637	30
rf	0.9304577	50
rf	0.9298853	70
rf	0.9295443	90

Variable Importance



Results

- To give credit to each catcher, we use this equation for every pitch:
  - If strike: 1 - Strike Probability
  - If ball: Strike Probability * -1



- If strike: 1 - 0.9 = 0.1 | 1 - 0.2 = 0.8
- If ball: 0.9 * -1 = -0.9 | 0.2 * -1 = -0.2

Results

Catcher Name	Strikes Above Average	Pitches Caught	Strikes Above Average Per 100 Pitches Caught
Max Stassi	99.23274	6012	1.6505779
Jose Trevino	78.92431	5999	1.3156244
J. T. Realmuto	70.66656	8071	0.8755614
Jonah Heim	69.19894	5183	1.3351137
Mike Zunino	60.49536	6633	0.9120361
Austin Hedges	56.41083	6045	0.9331817
Kyle Higashioka	51.98333	4037	1.2876722
Jacob Stallings	51.74702	7468	0.6929167
Tucker Barnhart	51.55393	7475	0.6896847
Tomas Nido	48.82142	2975	1.6410561

Catcher Name	Strikes Above Average	Pitches Caught	Strikes Above Average Per 100 Pitches Caught
Salvador Perez	-102.59241	8791	-1.1670164
Pedro Severino	-97.25718	7648	-1.2716682
Zack Collins	-92.70735	4288	-2.1620183
Austin Wynns	-66.41618	3131	-2.1212449
Eric Haase	-44.84179	4534	-0.9890116
Riley Adams	-44.63059	2060	-2.1665335
William Contreras	-43.79019	3340	-1.3110834
Kurt Suzuki	-36.93331	4974	-0.7425274
Victor Caratini	-36.62842	6737	-0.5436904
Andrew Knizner	-36.19965	3420	-1.0584694

Video Example 1

- Strike (0.00108 strike probability)
  - https://baseballsavant.mlb.com/sporty-videos?playId=2a91c3cf-e5d9-41f9-86e4-88700536a6de
- Austin Nola gained his team 1 - 0.00108 = 0.9982 strikes from this pitch

Video Example 2

- Ball (0.99087 strike probability)
  - https://baseballsavant.mlb.com/sporty-videos?playId=3509cf12-7e2c-4beb-88af-d3b3cc736c02
- Kevan Smith lost his team -0.99087 strikes from this pitch