



The New Defense in Baseball: Predicting Alignments in the post-shift era

Johnny Vecchio

Outline

- Overview of Alignments, New Rule, and its Effect
- Question and Data
- Prediction tool
- Results
- Practical Scenario
- Limitations
- Future work

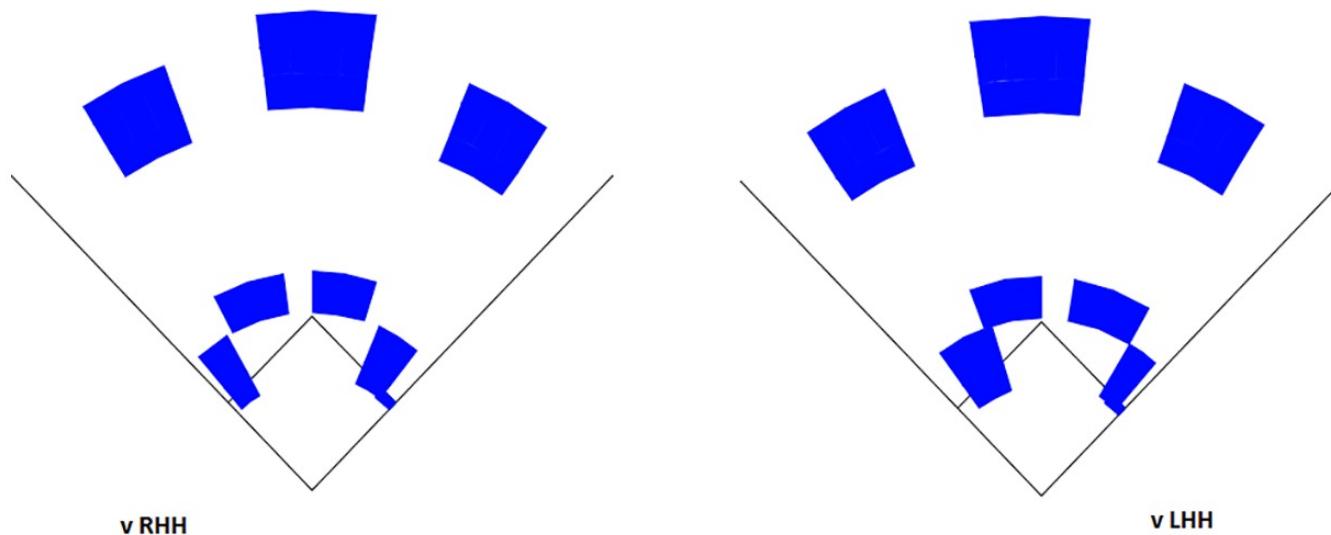
An aerial photograph of a baseball stadium during a game. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. The dirt infield is a warm tan color. Numerous players in white and dark uniforms are scattered across the field, some near the bases and others in the outfield. The stadium's seating area is packed with spectators, their colorful clothing creating a dense pattern. The overall scene is a dynamic representation of a live baseball match.

Alignments

Standard Alignment

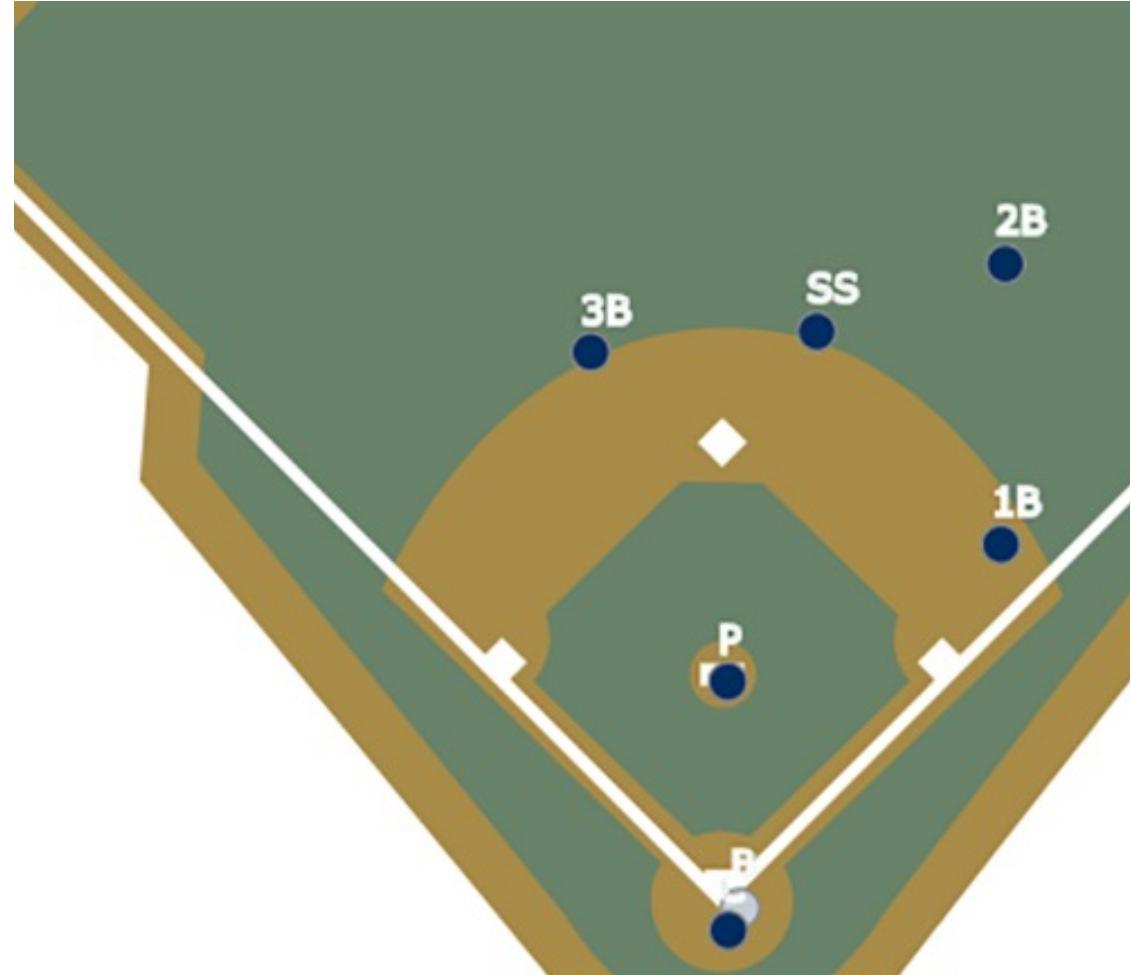
All four infielders are positioned in a basic/typical/traditional spot on the field

Standard Fielding Alignments



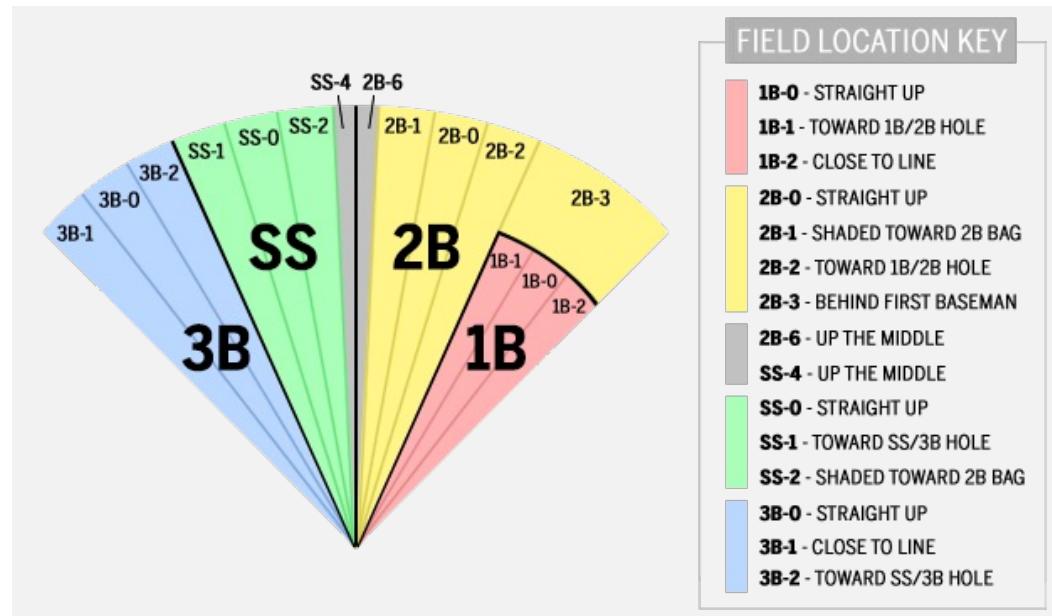
Infield Shift

Three or more infielders on
the same side of second base
(only applies to pre-2023)



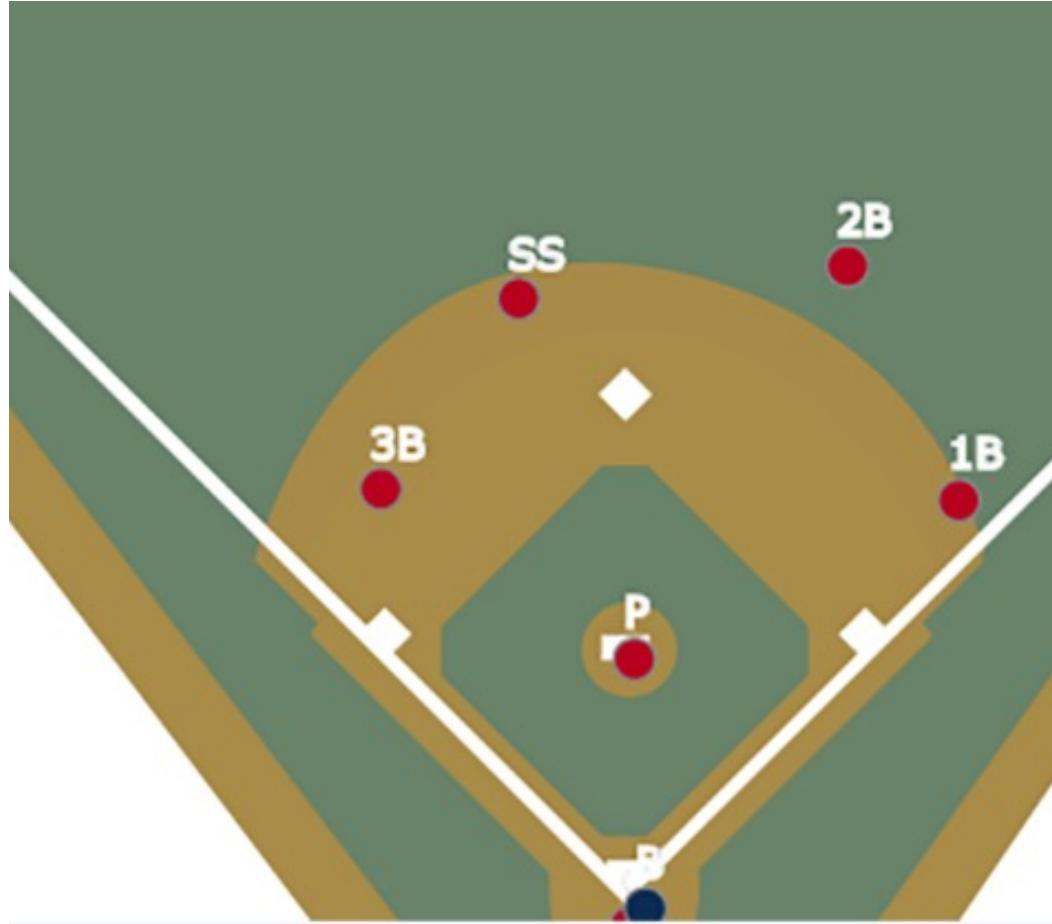
Infield Shade

At least one fielder is positioned outside the area of the field they're typically responsible for.
(Example: The SS is "shaded" when playing up the middle against a LHB or toward the 3B zone against a RHB)



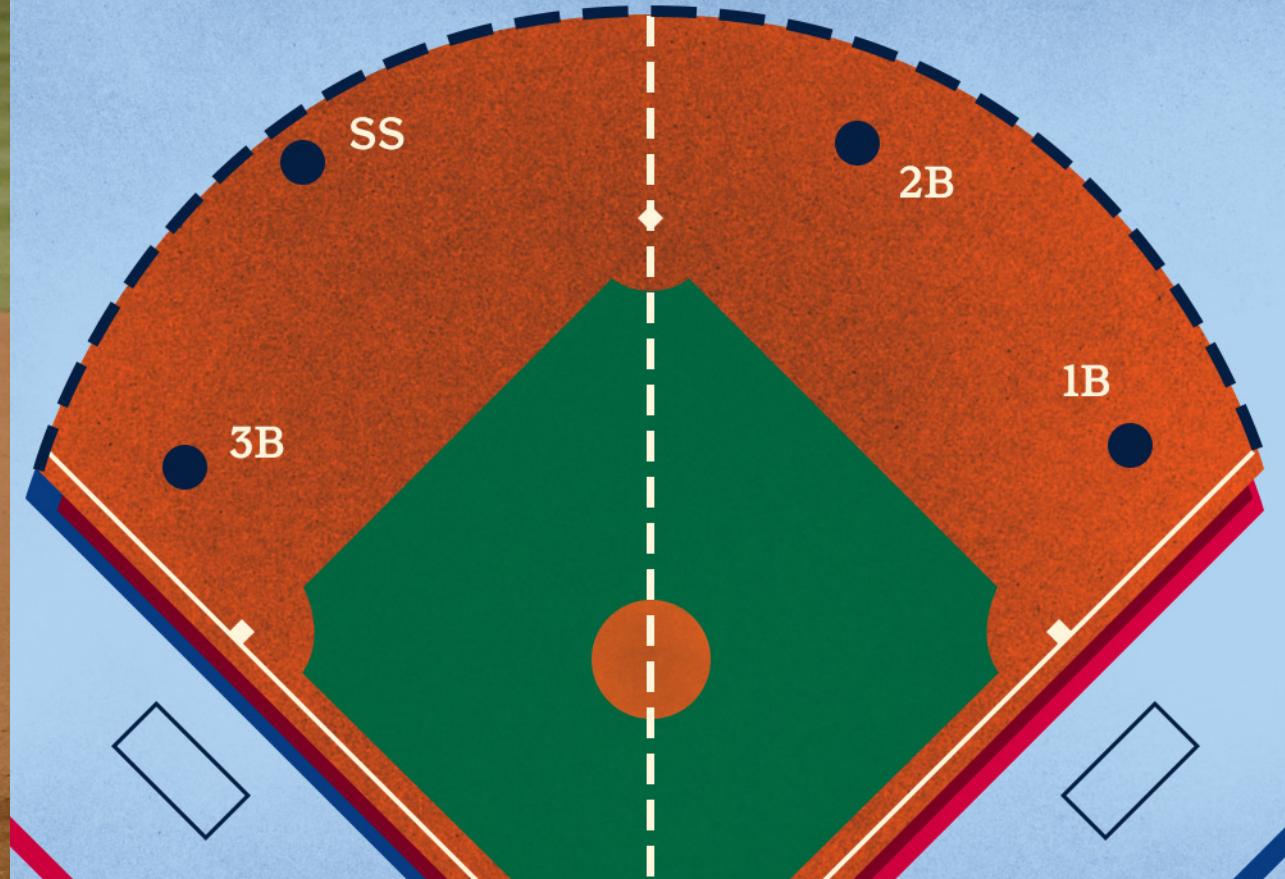
Strategic

Any other tactical alignment.
(Examples: an infielder playing in or holding the line, an outfield rover, a fifth infielder, etc.)



An aerial photograph of a baseball stadium during a game. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. The dirt infield is a rich tan color. Numerous players in white uniforms are scattered across the field, some near the bases and others in the outfield. The stadium's seating area is packed with spectators, their colorful clothing creating a dense, textured pattern. The overall scene captures the energy and scale of a live baseball match.

New Rule and Effects



New Defense Rules

- Shift Banned in 2023 Season
- Must have two players on each side of the dividing line

Effects of Ban on Batters at a glace

- Right-handed batters slightly improved
- Left-Handed batters saw greatest increase in stats
 - Batting Avg increased from .239 in 2022 to .249 in 2023
 - OPS increased from .706 to .746
 - 738 more singles hit
 - Over 1,000 more RBIs
 - 1,463 more total hits than in 2022

**Can the new positionings
be predicted?**

An aerial photograph of a baseball stadium filled with spectators. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. Several players in white uniforms are scattered across the field, some near the bases and others in the outfield. The dirt infield shows signs of wear and dust. The surrounding stadium lights are visible against the night sky.

Data and Prediction Tool

Data

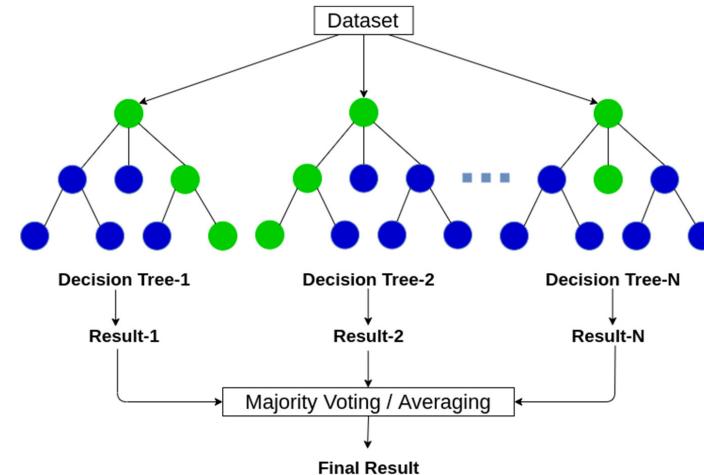


- Utilized Statcast Search from Baseball Savant
 - Every in-play batted ball event for the 2023 season
 - Over 120,000 events
- Dropped Useless/Missing Columns
- Dropped any NA values
- Changed some columns names for readability
- Factored certain columns to prepare for Prediction tool
- 56 Independent Variables from 2023 season in play event data including things like batters, fielders, batter handedness, hit coordinates, launch angle, etc.
- Split Dataset into Training and Testing

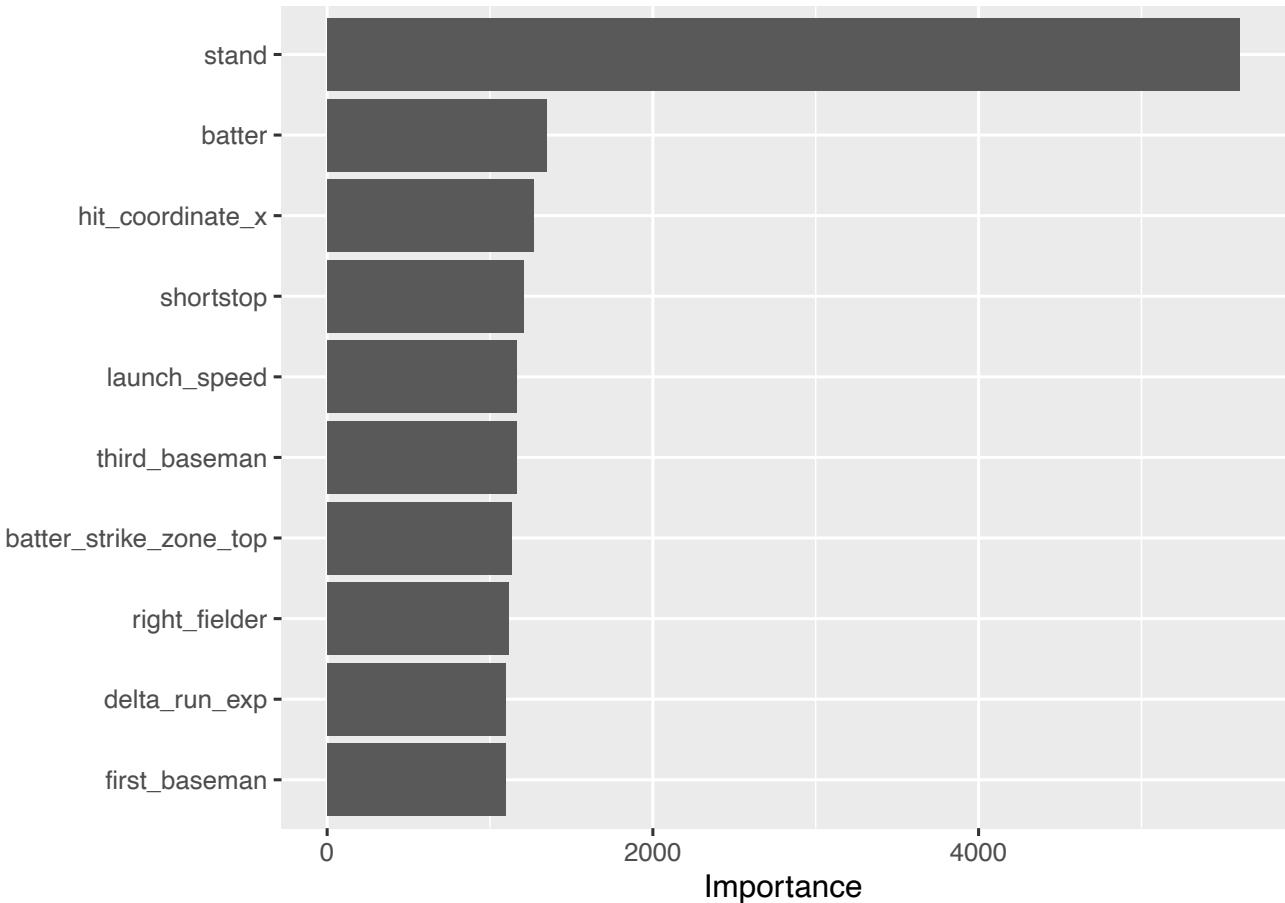
Prediction Tool

- Random Forest Model (ranger for high dimensionality)
- Ensemble learning method that combines multiple decision trees to make predictions or classifications.
- Each tree in the forest operates on a random subset of data and features
- The final prediction is determined by aggregating the individual predictions from all the trees.

Random Forest



Feature Importance



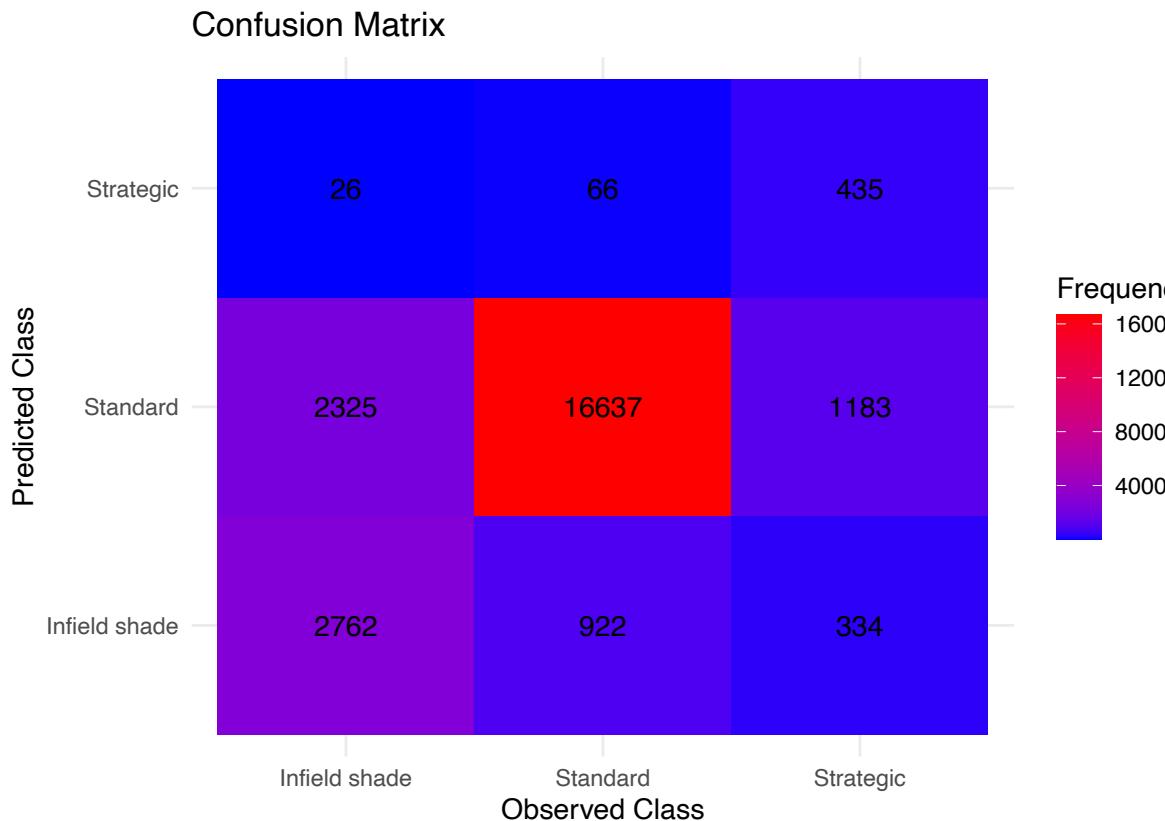
Confusion Matrix and Model Evaluation

Accuracy: 0.8033212

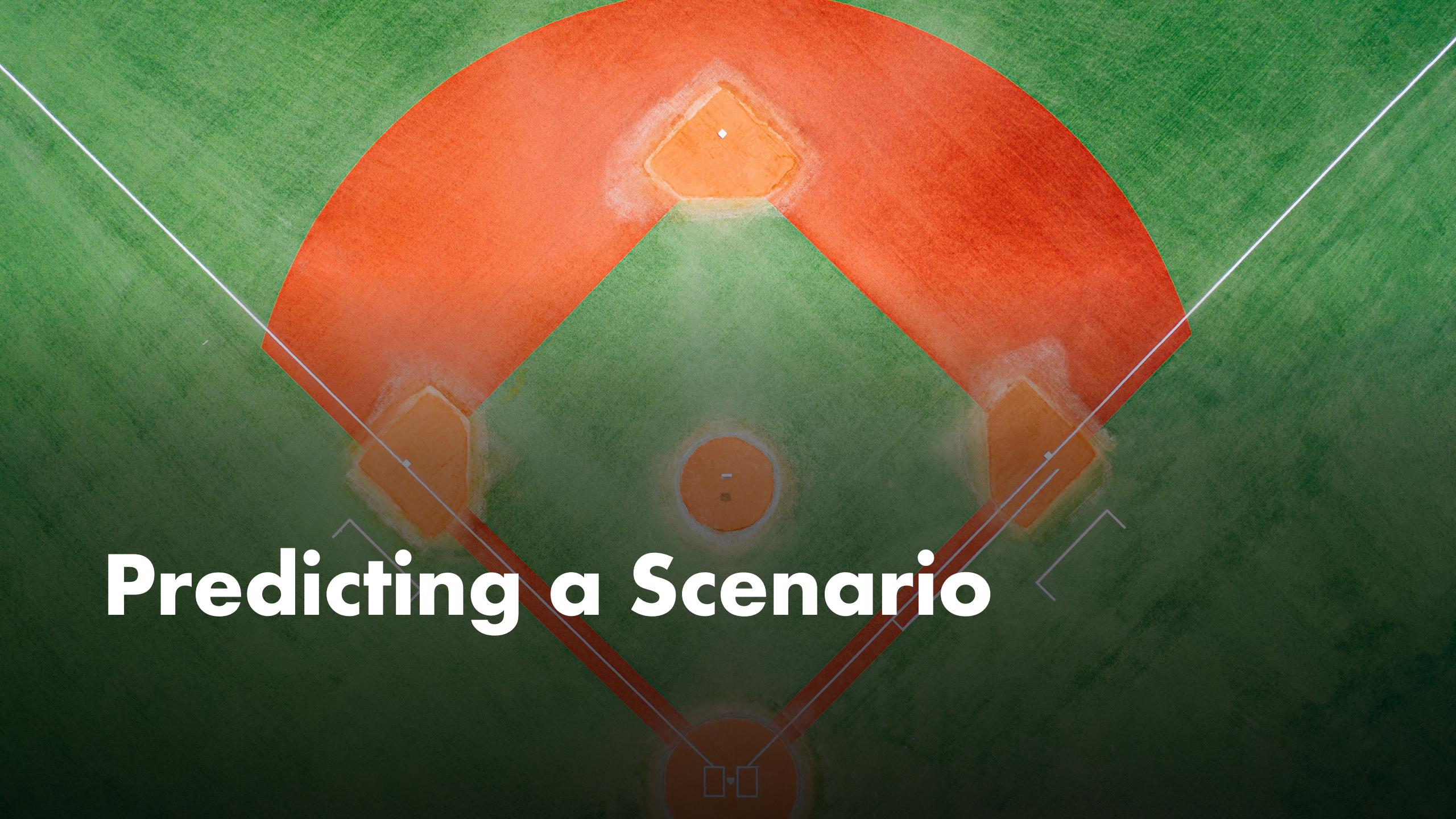
Overall Precision: 0.7795654

Overall Recall: 0.5689944

Overall F1 Score: 0.6122946



Predicting a Scenario





Matt Olson

First Baseman for Atlanta Braves

Saw Increased Stats in 2023

Led league in HR and RBI in 2023

Bats Lefty



Pitcher - Nathan Eovaldi

First Baseman - Nate Lowe

Second Baseman - Marcus Semien

Third Baseman - Josh Smith

Shortstop - Corey Seager

Left Fielder - Evan Carter

Center Fielder - Leody Taveras

Right Fielder - Adolis Garcia

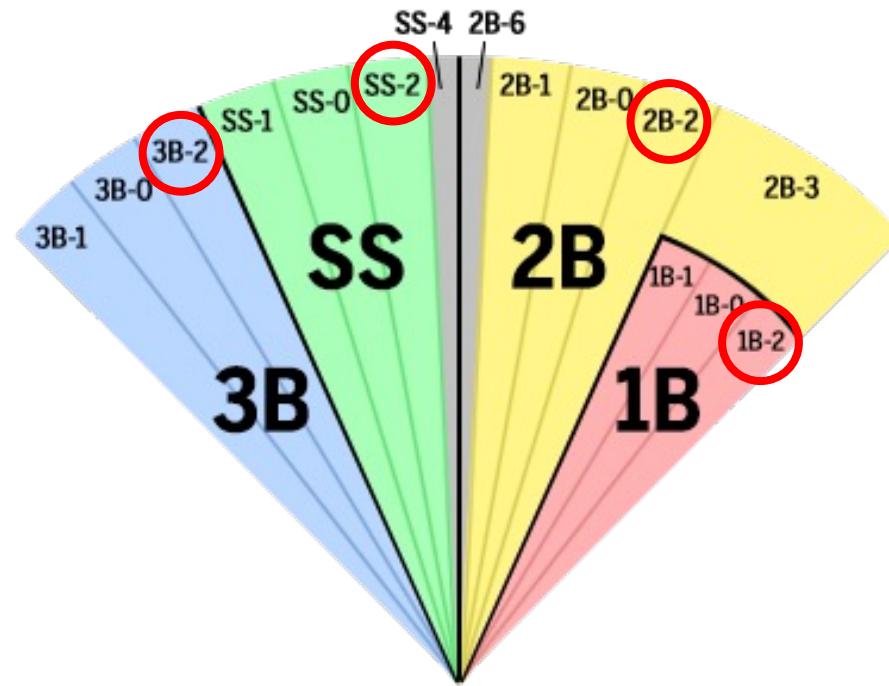
Batter - Matt Olson

Input Data

- Created Second Model including a handful of high importance features
 - Batter, Pitcher, Fielders 3-9, Batter Handiness, Hit Coordinate X, Hit Coordinate Y, Launch Speed
 - Chose Matt Olson's average X and Y hit coordinates and average Launch Speed for his in-play balls for 2023 season

Alignment Prediction from Model

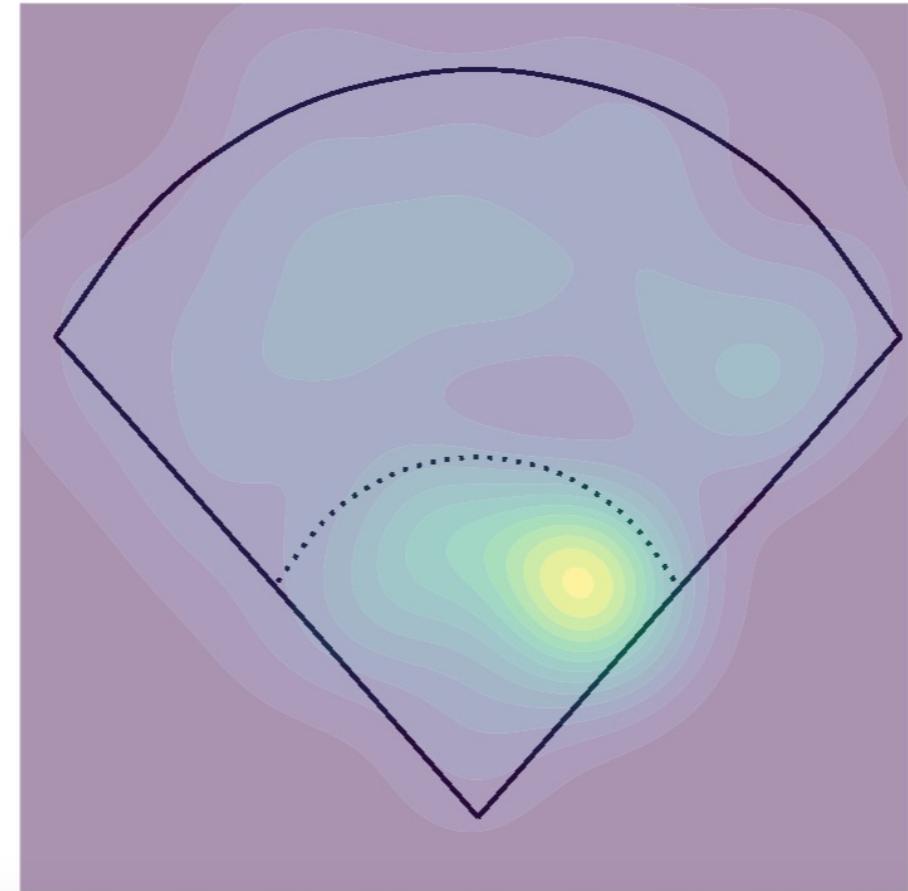
- **Infield Shade**
- Does this make sense for Matt Olson?





Matt Olson Spray Chart

Matt Olsen 2023 season



An aerial photograph of a baseball stadium during a game. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. The dirt infield is a warm tan color. Numerous players in white uniforms are scattered across the field, some near the bases and others in the center. The stadium's seating area is packed with spectators, their colorful clothing creating a dense, textured pattern. The overall scene is a dynamic representation of a live baseball match.

Limitations and Challenges

Limitations and Challenges

- Limited to just batted ball coordinate data
 - Data only had what type of alignment was used for each event
- Could not find a dataset that included defensive positioning play by play events
- Some features could be redundant
- Hard to implement

An aerial photograph of a baseball stadium during a game. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. The dirt infield is a warm tan color. Numerous players in white uniforms are scattered across the field, some near the bases and others in the center. The stands are packed with spectators, their colorful clothing creating a dense pattern. The overall scene is a dynamic representation of a live baseball match.

Future Work

Future Work

- Further refining and tuning the Model to improve accuracy and precision
- Possibly trying another Machine Learning technique
- Optimizing Defensive Positioning
 - Could not find coordinate based defensive data
 - Would need a little more linear algebra knowledge

References

- All Data from Baseball Savant
- https://baseballsavant.mlb.com/statcast_search
- <https://baseballsavant.mlb.com/csv-docs>
- <https://www.mlb.com/glossary/statcast/shifts>
- Packages used: baseballr, tidyverse, caret, ranger, vip, randomForest

An aerial photograph of a baseball stadium during a game. The field is a vibrant green with distinct white chalk lines marking the bases and foul lines. The infield dirt is a light tan color. Numerous players in white uniforms are scattered across the field, some near the bases and others in the outfield. The stadium's outer walls are visible on the left, and the stands are packed with spectators. The overall scene is a bright, sunny day.

Thank you