



MLB Statcast Data

Casey McGonigle
Mickey Piekarski
Noor Gill
Jaypal Bhatia

Our Team



Casey



Mickey

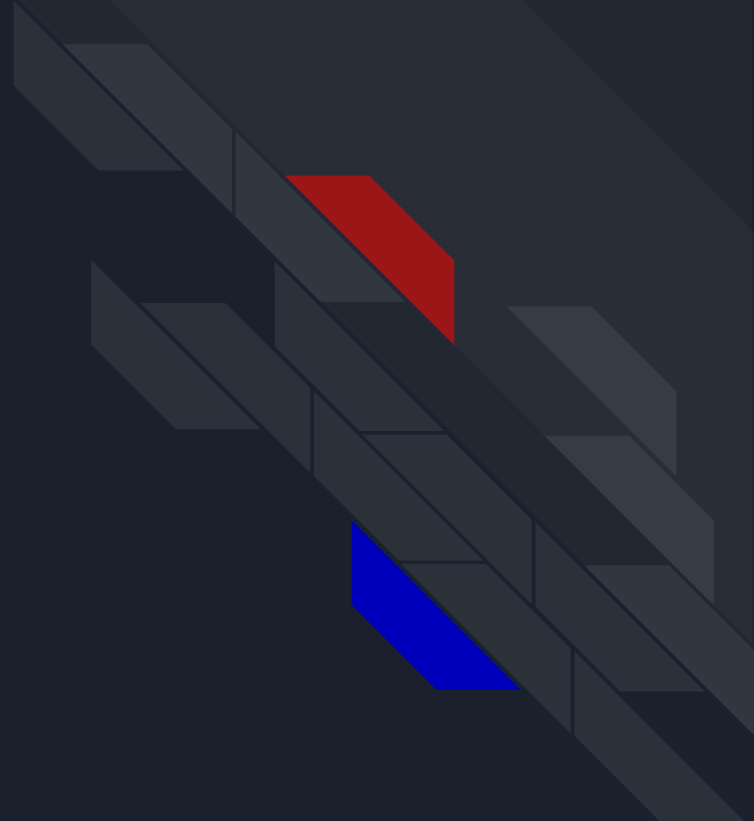


Jay



Noor

Understanding The Game of Baseball for our Model





1. OHTANI 0-3
 **2**

CISNERO P-5
 **1**

▲8  0-1

  STRAW: 1-4 2 WALKS RBI 2 RUNS SB (21)



Our Problem:

**Predicting the Outcome of any MLB Plate
Appearance**

(Hitter vs Pitcher Matchup Outcomes)



Data

MLB Statcast Data - [Baseball Savant](#)



Result of the Play: Double



Potential Next Steps

- Add more data on individual player stats / change clusters
- Add data on players' production in the past week
- Change format for outcome prediction / be able to compare to test set
- Make it update with different pitches/counts
- Instead, try to predict players' average in 2021 given 2020, 2019, 2018 stats
- predict era, k/9, whip, etc. off the spin rate and average velo of each pitch by pitcher; same for offensive stats and exit velo, launch angle, etc.
- Data Viz Idea: Every player — how they do against every pitch (and how often they see it) in 1 chart that is easy to understand
- Something like this: figure out which pitches pitchers should throw more often or which pitches hitters should sit on



EDA



Clustering (Unsupervised Learning)



Modeling (Supervised Learning)



Results & Interpretation



Further Applications & Development

Thank you for listening!
Any Questions?

