PITCH PERFECT:

Determining the next pitch through situational data in Major League Baseball

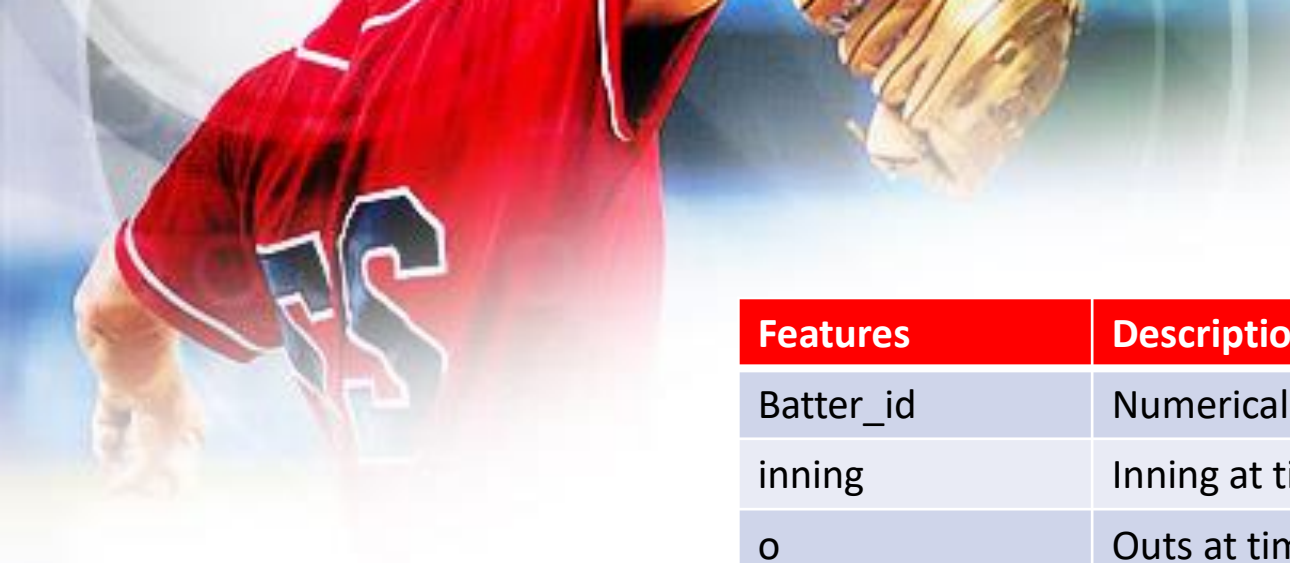
https://poweredtemplate.com/01142/0/index.html

THE PROJECT

It is becoming more apparent in MLB that data driven information not only helps managers in making decisions, it has improved the mediated consumption of the game. Instantaneous visual information on pitch location, ball trajectory and scoring probability has changed the game for the better for the numbers-hungry fan.

This project aims to add to mediated consumption of the game by attempting to predict the next pitch based on a given situation in a game. Being able to present this type of information before-hand will add significantly to the most watched portion of the game – the battle between pitcher and batter.

https://poweredtemplate.com/01142/0/index.html



THE DATA

Data for this project was gathered from a Kaggle dataset named MLB Pitch Data 2015-2018. Two files, pitches and at bats were sed for this project. The overall observations totaled over 2.8 million. Due to processing constraints and ensuring the most recent data was used, the 2018 year was isolated from the rest of the data. This data has over 700,000 observations and included the following features and target:

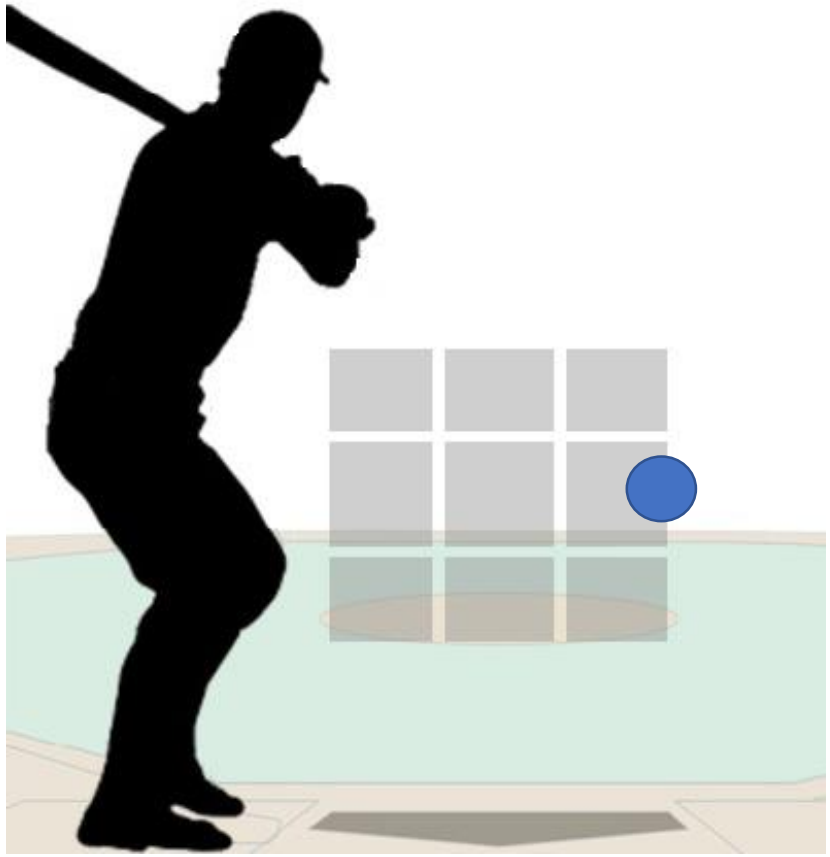
Features	Description
Batter_id	Numerical ID of the batter
inning	Inning at time of pitch
o	Outs at time of pitch
P_score	Score of pitcher's team at time of pitch
P_throws	Pitcher throws right/left
Pitcher_id	Numerical ID of the pitcher
stand	Batter hits right/left
top	Determines if top of the inning
B_score	Score of batter's team at time of pitch
B_count	Ball count at tme of pitch
S_count	Strike count at time of pitch
On_1b	Is there a runner on 1st base
On_2b	Is there a runner on 2nd base
On_3b	Is there a runner on 3rd base

Target	Description
CH	Changeup
CU	Curveball
FC	Cutter
FF	Fastball (4S)
FS	Splitter
FT	Fastball (2S)
KC	Knuckle curve
KN	Knuckleball
SI	Sinker
SL	Slider

THE RESULTS

CLASSIFICATION

BALL OR STRIKE?

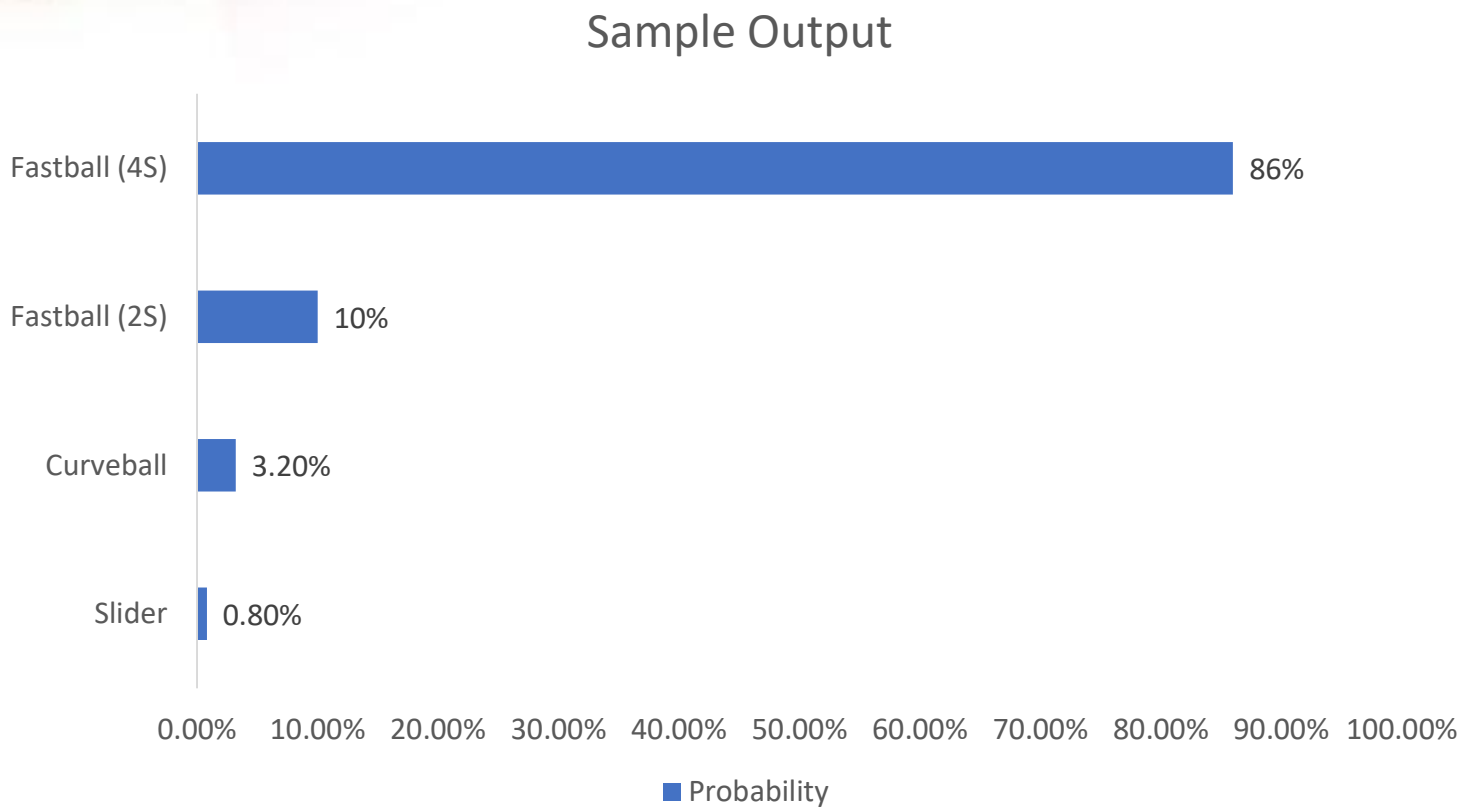


The first approach taken was to try to predict exactly which pitch would come next based on the situation. Out of 10 possible pitches that could possibly be thrown, our latest results would have selected the right answer about half (51%) of the time. Random guessing would be at 10%, but still, a 50/50 shot is not ideal for the goal.

https://poweredtemplate.com/01142/0/index.html

THE RESULTS

PROBABILITY



A better option would be to provide probability results. The sample output shown to the left represents what would appear when situational information is entered and a prediction is called.

https://poweredtemplate.com/01142/0/index.html

THE NEXT STEPS

- Continue with classification improvement
- Build-out for testing
- Alpha tests on model

https://poweredtemplate.com/01142/0/index.html