MLB Pitch Outcome Classification Capstone

Eric Wehmueller

September 9, 2021

Project Summary

- Advanced metric collection on every pitch
- Illegal substances in 2021 -> spin rate increase -> huge advantage?
- What else makes a pitch effective?
- What metrics give us the best results?
- Coaching staff/Analyst for the St. Louis Cardinals

Division Rival- Cincinnati Reds

		Standard Stats											
Player	Season	PA	AB	H	2B	3B	HR	BB	SO	BA	OBP	SLG	WOBA
 Jesse Winker	2021	481	420	129	32	1	24	53	75	.307	.395	.560	.406
 Nick Castellanos	2021	499	457	145	35	1	27	33	107	.317	.371	.575	.397
 Joey Votto	2021	459	391	105	20	1	29	60	115	.269	.368	.547	.383

- wOBA= Weighted On-Base Average
  (Offensive Contributions per Plate Appearance)
- Plan = Model pitching AGAINST Jesse Winker to start to minimize damage to look for advantages

Data Info & Features

- Pitch Type (Fastball vs Offspeed)
- Pitcher Handedness (R or L)
- Velocity
- Exact Pitch Location at Plate
- Horizontal Pitch Movement
- Vertical Pitch Movement
- Spin Rate
- Current Strike Count
- Current Ball Count

*Data from ~6,000 pitches against Jesse Winker

Classification Explanation

Pitcher Outcomes

Negative

- Any on-base result (single, error, double, HR, etc.)

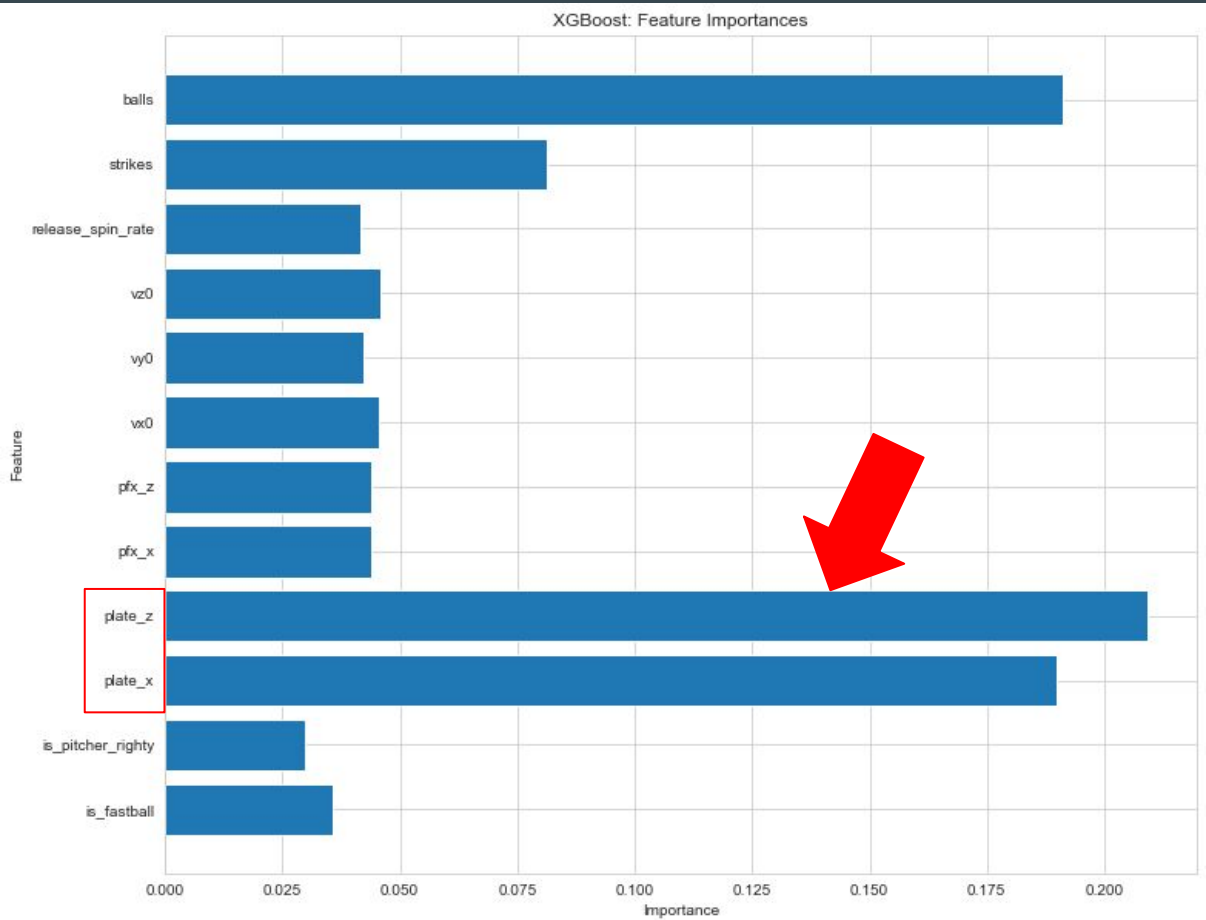
Neutral

- Called Ball

Positive

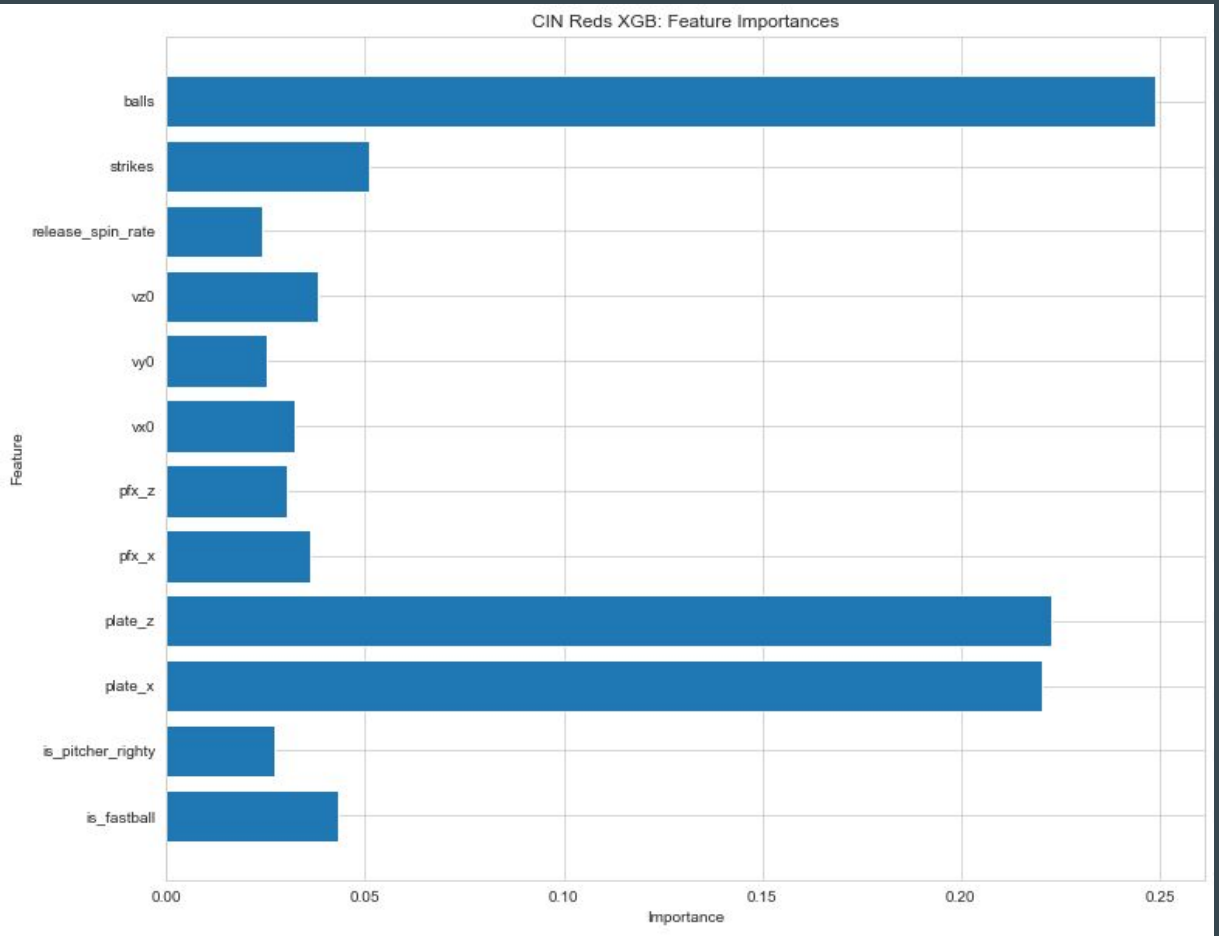
- Ground/Fly Out
- Results in called strike
- Swinging strike
- “Looking strike”
- Foul Ball

Jesse Winker Model - Feature Importances



- Plate_z + plate_x = Location, Location, Location
- # of balls = behind in the count-> heading towards negative pitcher outcome
- # of strikes = ahead in the count -> heading towards positive pitcher outcome
- Pitch Deception
- R vs L pitcher = Jesse is great against both
- Spin Rate vs. Jesse in bottom 3 for predicting pitcher outcomes
- Some overfitting issues + low recall scores w/negative pitcher outcomes
- **81.4%** accuracy model

Entire Reds Team Model



- Location Still Important
- # of balls = the team tends to generally do better when ahead in the count
- R vs L pitcher = we have a mix of hitters, not relevant here
- Spin Rate vs Reds is the worst predicting factor for pitch outcome
- No more overfitting issues
- Continued low recall scores for negative pitcher outcomes
- **80.3%** accuracy model

Conclusions and Results

What are the most important metrics that go into a pitch against the Reds?

- **The current ball count + the pitch location**

What is the least important metric that goes into a pitch against the Reds?

- **Spin Rate (surprisingly!!)**

Future Work

- More models against entire MLB starters/teams
- Spin Rate just irrelevant vs the Reds, is it a myth/overhyped statistic in 2021?
- Tweaked Models per player for a game setting
- Isolated Models just for exact pitch type and location (beyond is_fastball)

Thank You!

Eric Wehmueller

Email: ericwehmueller@gmail.com
Github: @ewehmueller
Linkedin: https://www.linkedin.com/in/eric-wehmueller-58719780/