MLB Player Projection Analysis

Joseph Resis

1 . Data

Observed statistics from 3173 hitters and 1298 pitchers from 2008-2020.

- ➔ Predicting player value, encompassed in a statistic called Wins Above Replacement (WAR).
- ➔ WAR scale (one full season): 2=average, 5=all-star, 8=MVP

Key Questions

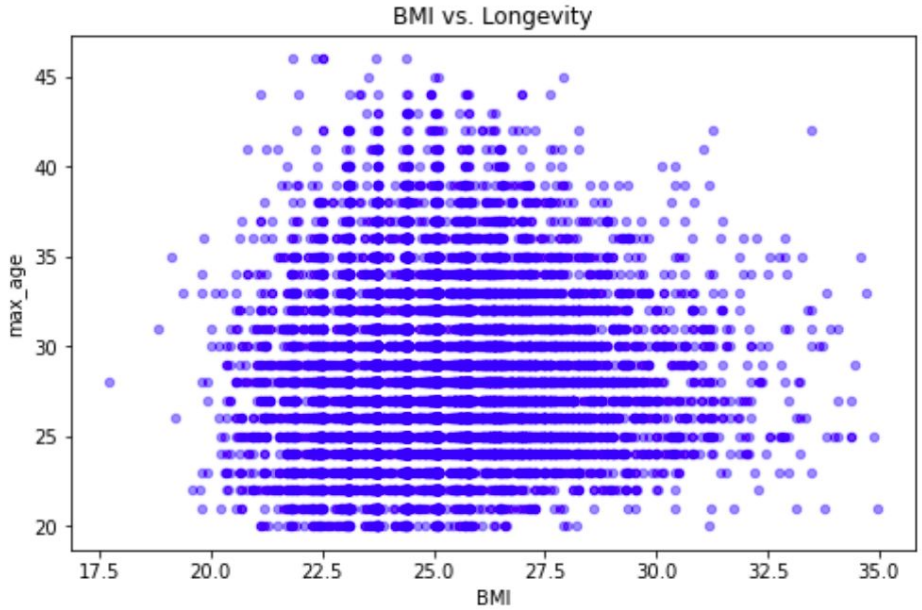
1. Which types of players tend to last longer in the league?

2. Which factors are most influential in WAR calculation?

3. Which statistics get too little/too much attention among today's front offices?

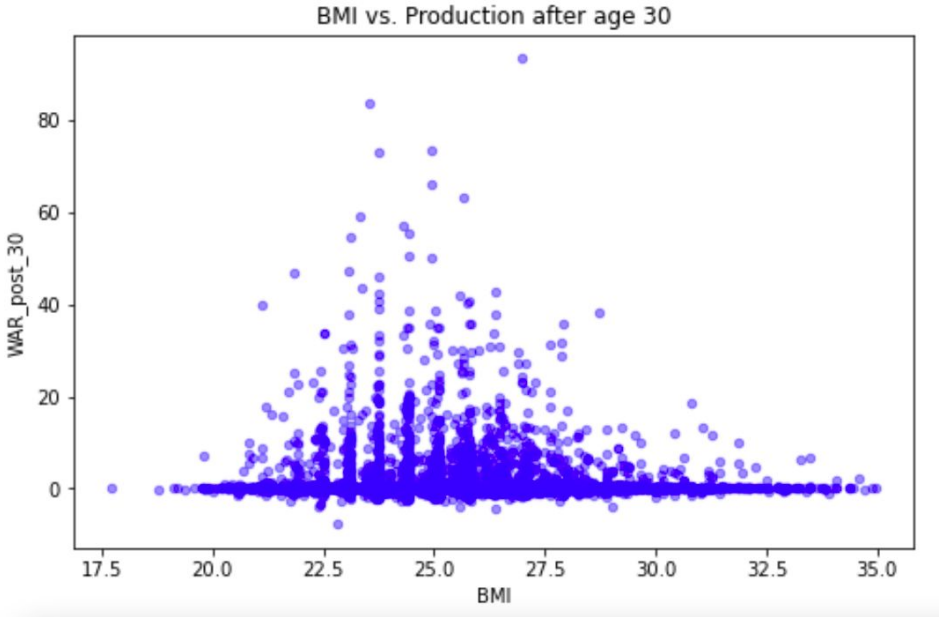
2. Pre-model EDA

- BMI vs. longevity (max age)



2. Pre-model EDA

- BMI vs. WAR after age 30



3. Linear regression (batters)

- R-Squared: .30
- Limits of stolen bases
- Ability to make contact

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.2045	2.170	1.016	0.310	-2.057	6.466
BMI	0.1302	0.077	1.689	0.092	-0.021	0.282
batting_score	0.0206	0.002	12.188	0.000	0.017	0.024
strikeout_rate	-17.2422	2.491	-6.921	0.000	-22.135	-12.349
fielding_score	0.0184	0.004	4.166	0.000	0.010	0.027
infield_score	0.0834	0.038	2.175	0.030	0.008	0.159
catching_score	0.0497	0.015	3.347	0.001	0.021	0.079
SB	0.0153	0.007	2.135	0.033	0.001	0.029
CS	-0.0575	0.026	-2.197	0.028	-0.109	-0.006

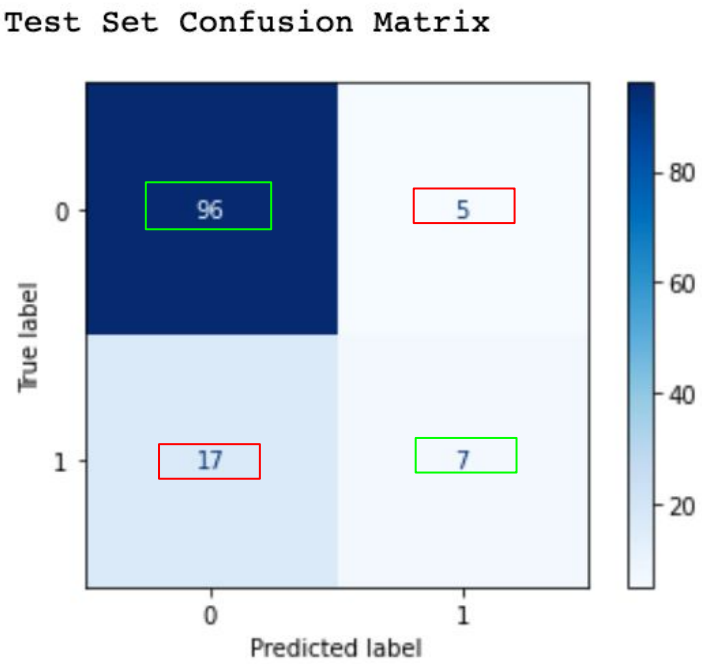
3. Linear regression (pitchers)

- R-Squared: .30
- Quality of contact allowed
- Value of strikeouts, cost of walks

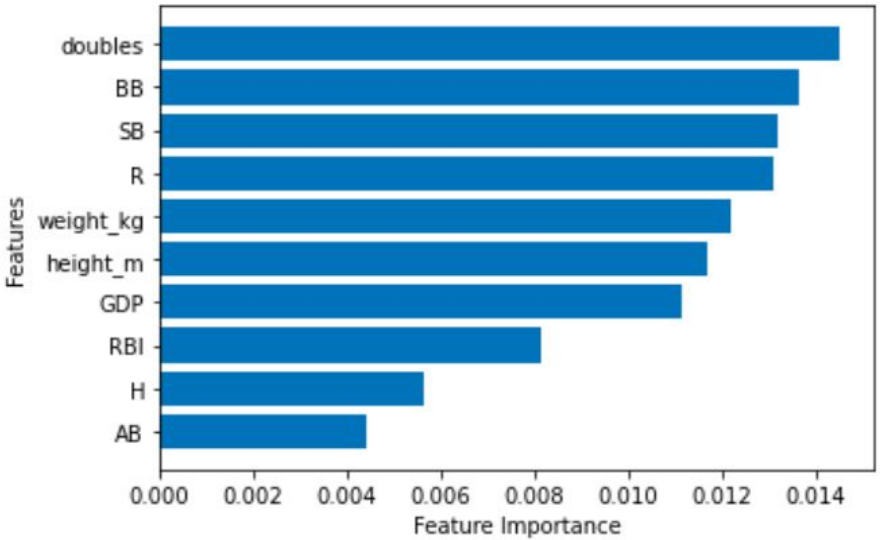
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	7.0497	5.869	1.201	0.230	-4.477	18.576
runs_per_9	-1.8337	0.362	-5.059	0.000	-2.546	-1.122
ER	0.0080	0.001	8.590	0.000	0.006	0.010
GB_div_FB_ratio	8.1586	6.479	1.259	0.208	-4.568	20.885
line_drive_rate	-26.6198	7.139	-3.729	0.000	-40.642	-12.598
popup_rate	37.2821	18.320	2.035	0.042	1.298	73.266
SO9	-0.3940	0.141	-2.793	0.005	-0.671	-0.117
SO_div_BB	1.3374	0.268	4.993	0.000	0.811	1.864

4. Random forest (batters)

- Accuracy: .82
- Precision: .58
- Cross-validation: .84



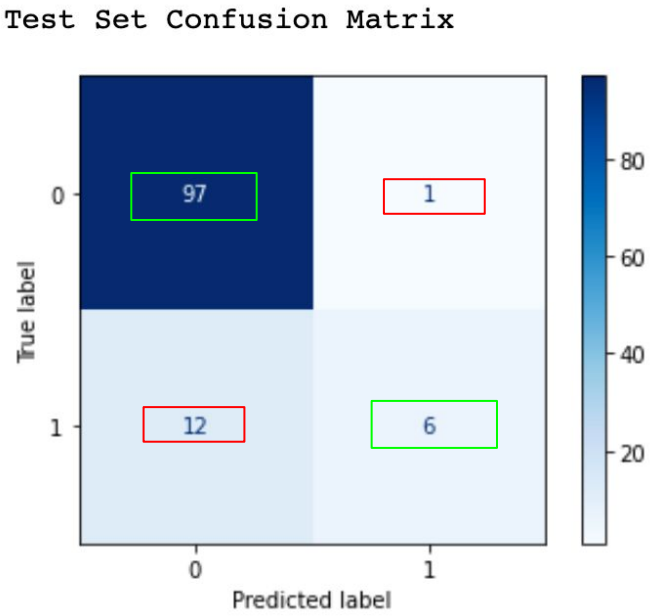
4. Random forest (batters)



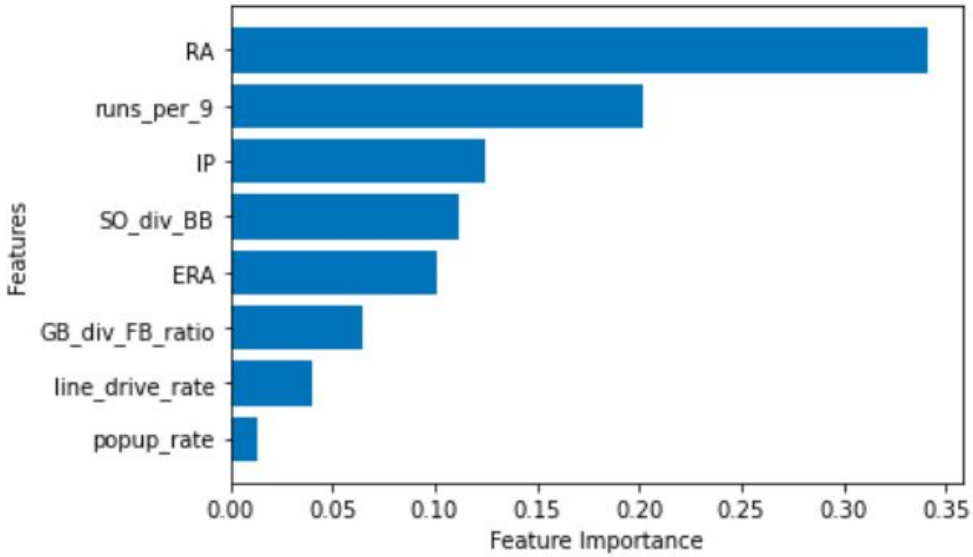
- Easiest to accumulate value while batting
- Position value is significant

4. Random forest (pitchers)

- Accuracy: .89
- Precision: .86
- Cross-validation: .88



4. Random forest (pitchers)



- Strikeout (SO) vs. walk (BB) ratio
- RA vs. ERA

5. Recommendations (batters)

- BB rate as a metric
- Batting value vs. other areas' value
- Speed-reliant players' limited longevity

5. Recommendations (pitchers)

- RA vs. ERA
- Strikeout/walk ratios
- BMI as a non-factor
- Ability to last long in games

6. Future Work

- Pitching durability
- Use of Statcast
  - Pitchers
    - Spin rate
  - Hitters
    - Exit velocity
    - Launch angle

Thank you for watching!

Joseph Resis
jresis10@gmail.com
https://github.com/jresis/mlb-free-agent-analysis