Predicting Baseball

By Michael Tiernan

Disclaimer

This presentation is for educational purposes only. Findings and analysis are based on past results and are in no way a reliable indicator of future performance. Wagering real money based on the contents of this presentation is 100% discouraged. I hereby waiver any responsibility on my behalf if you choose to ignore this notice.

Table Of Contents

- Sports betting has become increasingly popular since 2018 when the US Supreme Court ruled in favor of legalizing the industry in New Jersey

- Billions of dollars a year are now being generated by this industry

Business Problem

1) Predict season total wins
2) Evaluate what the best predictors are to winning a single regular-season game

- The average bettor is losing $$$
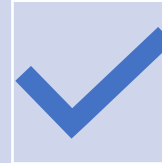
- Data science can help!

Methodology

Gathered data from reputable sources

Cleaned, explored, and analyzed the data
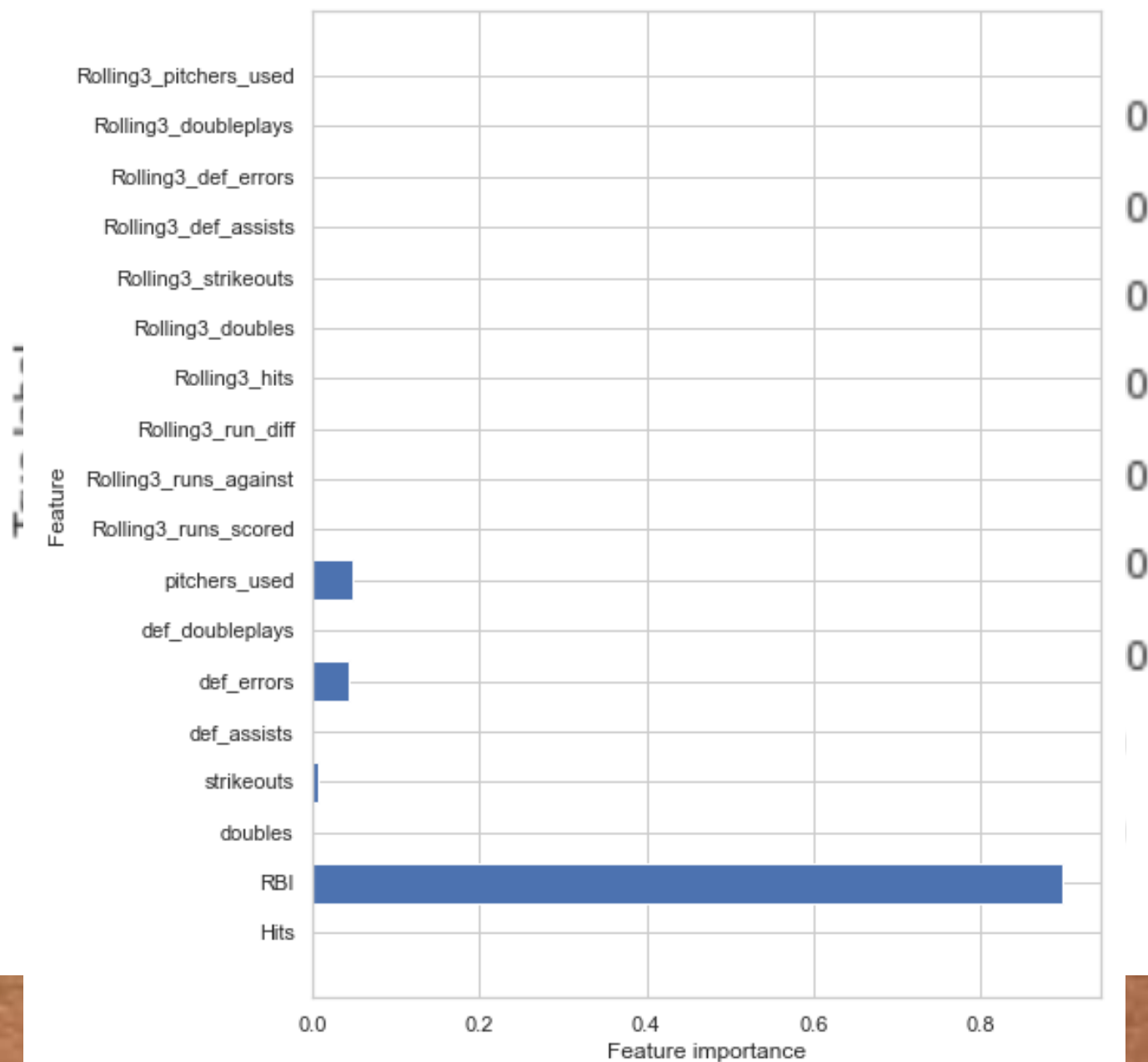
Built predictive machine learning algorithms

Evaluated models to make recommendations

Results

- Recall score: 80%
- Accuracy: 76%
- AUC: 86.5%

Data Scientist

vs.

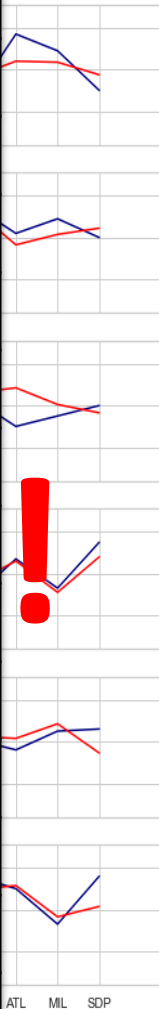Data Scientist

Team	Wins	Vegas Line	Absolute Diff	% Diff
HOU	551	508	43	9%
MIL	494	464	30	7%
NYY	549	524	25	5%
CHC	544	524	20	4%
OAK	494	476.5	17.5	4%
CLE	486	517.5	18	4%
LAD	579	560.5	18.5	3%
ATL	473	459	14	3%
MIN	471	482	14	3%
TBD	491	478	13	3%
STL	538	524	14	3%
WSN	546	546	0	0%
PIT	490	492.5	-2.5	-1%
ARI	472	475	-3	-1%
KCR	462	469	-7	-2%
COL	458	465	-7	-2%
NYM	489	497	-8	-2%

FROM 2014-2019,
ONLY 5 TEAMS > 5% DIFF!!!

Season Total Wins

ATL   MIL   SDP

Recommendations

1. Bet one of the following teams when wagering on a team's season total wins
   - Detroit Tigers – 10% - Under
   - Houston Astros – 9% - Over
   - Cincinnati Reds – 8% - Under
   - Milwaukee Brewers – 7% - Over
   - San Diego Padres – 6% - Under

2. Bet on teams that have high run differential and produce a lot of RBIs

Future Work

- Start analysis from the most micro level (1 pitch, 1 swing) and build up to the most macro (team data)

- Incorporate ELO ratings for players and teams

- Track daily odds from the bookmakers to then build a database that I can then compare to actual results in pursuit of back-engineering the models that bookmakers use

Thank You!

Name: Michael Tiernan

GitHub: @MJ-AllDay-67

LinkedIn: https://www.linkedin.com/in/michaeljohntiernan/