Umpiring in the Age of Technology

A Study of Pitch Calling by MLB Umpires

Alexander J. Sutherland

Overview: In 2024, the pitch calling abilities of home plate umpires are under more scrutiny than ever: entities like *Ump Scorecards* publish summaries of home plate umpire performance for every *Major League Baseball* (MLB) game on social media (Ump Scorecards, 2024) and all AAA-level games in *Minor League Baseball* (MiLB) used the automated ball-strike (ABS) challenge system as of June 25th (Harrigan, 2024). Given the importance of correct calls to both game outcomes and fans – as well as the looming possibility of an ABS system in MLB games – we investigate both:

1) the efficacy of machine learning models to predict umpires' ball/strike calls, and
2) what features most impact calling balls and strikes (outside of pitch location).

Project Stakeholders: MLB, MiLB, Professional baseball organizations, Professional baseball umpires, Professional baseball players, Baseball fans.

Data: We use pitch data from the entire 2023 MLB season, which is originally provided and hosted by *Baseball Savant* (MLB Advanced Media, LP, 2024) and accessed through *pybaseball* (LeDoux & Schorr, 2024). We join this pitch information with home plate umpire information accessed from box scores on Baseball Reference (Sports Reference LLC, 2023).

Methods: Regarding task 1), we used both logistic regression and linear support vector classifier models trained on pitch location data and umpire information with 10-fold cross validation and hyperparameter tuning via grid search. These models were compared against coin-flip baseline models, where the odds of a pitch being called a strike matched the percentage of strikes present in the data frame.

For task 2), we provide visualizations of umpires' correct call percentages across a much wider class of features that are not location-based. We then perform feature selection using both forward selection and L^1 regularization to determine which features are most important in determining when an umpire will make a correct call.

Key Performance Indicators (KPIs): For all our models, we prioritize model accuracy. We additionally track F1-score and PR-AUC as secondary metrics.

Conclusions and Future Work: Each of our models – trained on real umpire calls – consistently outperformed the actual umpires, with gains as large as 2.95% (with umpire correctly calling 91.76% of pitches and our model predicting 94.72% of pitches correctly). Additionally, we found that the most important (non-location-based) factors for predicting correct umpire calls were the pitch count, height of the regulation strike zone, effective pitch speed, and vertical pitch movement. Future work should consider a longitudinal study of non-location-based factors on predicting correct umpire calls, as well as a thorough study of the new ABS system in MiLB.

Find out more on this project's Github repository:

References

Harrigan, T. (2024, Jun 18). *Triple-A to employ challenge system over full ABS for rest of season*. Retrieved from MLB.com: https://www.mlb.com/news/triple-a-abs-challenge-system

LeDoux, J., & Schorr, M. (2024, Oct 03). *Pybaseball Github Repository Readme*. Retrieved from Pybaseball Github Repository: https://github.com/jldbc/pybaseball

MLB Advanced Media, LP. (2024, Oct 03). *Statcast Search*. Retrieved from Baseball Savant: https://baseballsavant.mlb.com/statcast_search

Sports Reference LLC. (2023, Oct 03). *Baseball Reference*. Retrieved from Baseball Reference: https://www.baseball-reference.com

Ump Scorecards. (2024, Oct 03). Retrieved from umpscorecards.com: https://umpscorecards.com/home/

Find out more on this project's Github repository: