WILDCARD WEIGHTING

PLACE NAMES OF GROUP MEMBERS HERE

ABSTRACT. The MLB postseason is known to be unpredictable. Over the last decade, Wild Card teams have outperformed division winners. This paper aims to investigate and quantify the underlying factors driving this phenomenon. Using a dataset of MLB games from 2015–2025, we train both Random Forest and XGBoost models to identify the variables that most strongly influence postseason success and to evaluate whether Wild Card teams possess structural advantages that are not captured by regular season standings.

1. RESEARCH QUESTION AND OVERVIEW OF THE DATA

Recent MLB postseasons have shown an interesting trend: Wild Card teams often perform better than division winners, even though they typically finish the regular season with weaker records. This raises an important question for both baseball fans and analysts:

> *What makes a Wild Card team outperform a division winner in the MLB postseason?*

This question is interesting because it challenges the assumption that regular season strength is a reliable predictor of postseason performance. It also relates to the design of playoff formats and whether certain types of teams are structurally advantaged in short playoff series.

2. BACKGROUND AND PRIOR WORK

Several previous studies have examined MLB postseason performance. Baseball Prospectus and FanGraphs have written extensively about the role of randomness, bullpen strength, and matchups in postseason series. Academic work (e.g., Berri & Schmidt, 2010) has shown that regular season winning percentage is a relatively weak predictor of postseason outcomes.

3. DATA SET

We use MLB game data from 2015-2025. The data includes variables such as team OOA, team WOBA, bullpen FIP, bullpen usage, whether the team is a wildcard, and postseason results. The data set is strong in that the data is specifically geared towards matchups and understanding the impact of the wildcard on the postseason. However, with only so many postseason

---

Date: 4 October 2024.

games (317), it is difficult to fully understand the impact of the wildcard teams.

Based on our prior experience with baseball, we form the following hypotheses:

(1) Wild Card teams may enter the later rounds of the postseason carrying more momentum than their division winner counterpart, causing them to win.
(2) Bullpen strength will be more important in the postseason than in the regular season.
(3) Matchup-based factors (pitcher handedness, lineup construction) may be more important than overall regular season strength.

4. Data Cleaning / Feature Engineering

Tell us what you did when you were cleaning your data and engineering features. Why did you make the choices that you did? What are the consequences of those choices?

(1) **FIP** stands for Fielding Independent Pitching, and is a metric designed to capture how well a pitcher prevents runs independently of the rest of the defense. It is a better metric of a pitcher's skill than Earned Run Average because it disregards events outside of the pitcher's control.
(2) **Bullpen** Because pitching strains a pitcher's arm quickly, one pitcher will start and throw about 5-7 innings before being switched out for a *relief pitcher*. The bank of pitchers ready to relieve are called the Bullpen.
(3) **OOA** stands for Outs Above Average and signifies how many more outs a player makes than an average fielder at the same position.
(4) **WOBA** stands for Weighted On-Base Average and signifies how often players make it on base, weighted by how likely they are to score based on the manner in which they made it on base.

We gathered the data we used in three main pulls. Our final dataset incorporates data from 1) The pitcher logs for each game for each pitcher during the regular season from 2014-2015, 2) the pitcher logs for each game for each pitcher during the post season from 2014-2015, and 3) team level statistics (such as outs above average or weighted on-base average) for each team for each year from 2014 - 2025.

Early on, we decided that we wanted our match-up model to include the following features for both teams: starting pitcher's FIP, starting pitcher's freshness, an aggregated value for the FIP and freshness of the pitchers in the bullpen, average OOA (as an indication of the efficacy of the team's defense), average WOBA (as an indication of the efficacy of the team's offense), and whether the team is a wildcard in the playoffs. We also decided to include a feature indicating which team played at home.

The pitcher's logs (collectively referred to as df) included the following pertinent features: Date, Team, a unique Pitcher ID, FIP, Innings played, and the result of the game. In order to assemble it into a useable format for our problem, we made a few tweaks and added some new columns. We first converted the Date column to pandas.DateTime. After sorting the df by Pitcher Id and Date, we grouped the df by Pitcher ID and shifted the Date column to create a Previous Game Date column, and subtracted the Previous Game Date column from the Date column to create a Freshness column for each pitcher in each game.

Additional changes to make the pitcher logs more useful included adding a Starter feature that one-hot-encodes whether the pitcher started that game, and a Game ID to uniquely identify each game, based on the date and the two teams playing.

The dataset of team level statistics by year included the following pertinent features: Team, Year, OOA (averaged over the season), WOBA (averaged over the season), and Is-Wildcard. This dataset was nearly immediately useable after cleaning the Team names to match those in the pitcher logs dataset. We merged it with the pitcher logs on Year and Team.

Equipped with a newly-merged datset (with features Date, Team, Pitcher ID, FIP, Innings played, Game Result, Freshness, Starter, Game ID, Team OOA, Team WOBA, Is-Wildcard for each pitcher in each game) we were ready to assemble a final games dataset. To do so, we grouped the newly-merged dataset by Game ID and then by team. For each team in each game, we assembled a dictionary containing the starter's FIP and freshness, the meaned FIP and freshness of the relief pitchers, whether the team was at home, team OOA, team FIP, whether it was a post-season game, and if so, whether the team was a wildcard during the game. After shuffling the order of the teams, we concatenated the two team's stats into a single row, with a results feature signifying whether the first team in the row won or lost. There were no ties in our dataset.

With our data in a tabular dataset with a binary label, we were ready to train!

5. DATA VISUALIZATION AND BASIC ANALYSIS

Analyze the data, draw conclusions, and effectively communicate your main observations and results.

- Calculate appropriate summary statistics.
- Use appropriate plotting techniques, visualizations, and other tools and techniques you have learned, to thoughtfully identify and evaluate what the data are telling you,how well suited the data are to answering your problem,
- Reference figures and plots, like Figure **??**.

6. Learning Algorithms and In-depth Analysis

Analyze the data using the machine learning techniques discussed in class. Explain what research questions you can answer using the machine learning techniques presented this semester, and if applicable, what you think you may be able to answer next semester.

Be able to explain the results of your analysis, whether the results are meaningful, and why you chose the tools that you used.

7. Ethical Implications and Conclusions

Thoughtfully analyze the ethical implications of your research questions, the data you gathered, and the analysis that was performed. Are there privacy or other implications from the collection or use of the data? Could your results and methods be misused or misunderstood? What can and should be done to prevent misuse and misunderstanding? Could your algorithms and methods result in a destructive self-fulfilling feedback loop? How could that be prevented or controlled? What other ethical implications does your work have?

This part should all be done before you get to *page 5*. The bibliography can spill on to page 6, but we won't read text that goes past page 5.