Predicting Pitch Types for 2024 MLB Season

Alex Rintamaa

2024-10-12

Abstract

This report analyzes hitting data from 2021 to 2023 to predict the types of pitches hitters will encounter in 2024 MLB season. The pitch types are categorized into three main groups: Breaking Balls, Fastballs, and Off-speed pitches. Using a Random Forest model, known for its robustness in handling large data sets and multiclass predictions, I achieved a prediction accuracy of 59.9%. Results indicate that hitters are likely to see Fastballs approximately 60% of the time, while the proportions of Breaking Balls and Off-speed pitches will vary vary by batter. This highlights the importance of pitch recognition for hitters, as they seek every advantage at the plate.

1. Introduction

The unique characteristics of MLB allow for advanced analytical tools to effectively change how baseball is played. Each matchup between a pitcher and batter resembles a chess match, with pitches concealing their pitch intentions while hitters prepare for any possible delivery. While extensive research has focused on optimal pitch selection based on hitters' tendencies, less attention has been given to predicting what pitches hitters should expect to see. This report aims to address that gap with a robust model capable of multiclass predictions. Unlike many professional sports, MLB does not have salary cap restrictions, making it crucial for smaller-market teams to stay ahead of trends. Anticipating pitch types can fundamentally alter how both pitching and hitting are approached. Recent changes aimed at speeding up the game, such as the implementation of the pitch clock and PitchCom, underscore the need for teams to adapt quickly using data-driven predictions.

2. Methods

2.1 Data Preprocessing & Exploratory Analysis

The dataset comprises over 1,200,000 pitched balls to batters who faced atleast 1,000 pitches between 2021 and 2023. It includes 56 variables representing various aspects of th game at the time of each at-bat. Key variables include:

- **PITCH_TYPE** Type of pitch thrown
- **BATTER_ID** Unique Identifier for the batter
- **PITCHER_ID** Unique Identifier for the pitcher
- **BALLS** Number of balls in the count
- **STRIKES** Number of strikes in the count
- **INNING** Game inning
- **INNING_TOPBOT** Top or bottom of inning
- **BAT_SIDE** Right-handed or left-handed batter
- **THROW_SIDE** Right-handed or left-handed pitcher
- **ZONE** Specific area of strike zone
- **PLATE_X** Horizontal movement of the pitch
- **PLATE_Z** Vertical movement of the pitch

Pitches lacking a specified type were removed from the dataset, as they could not be classified as Fastball, Breaking Ball, or Off-speed. To classify pitch types consistently, I created a scatter plot (Figure 1) displaying the average horizontal and vertical movements of each pitch type.

Figure 1: Average Pitch Movement by Pitch Type (2021-2023)

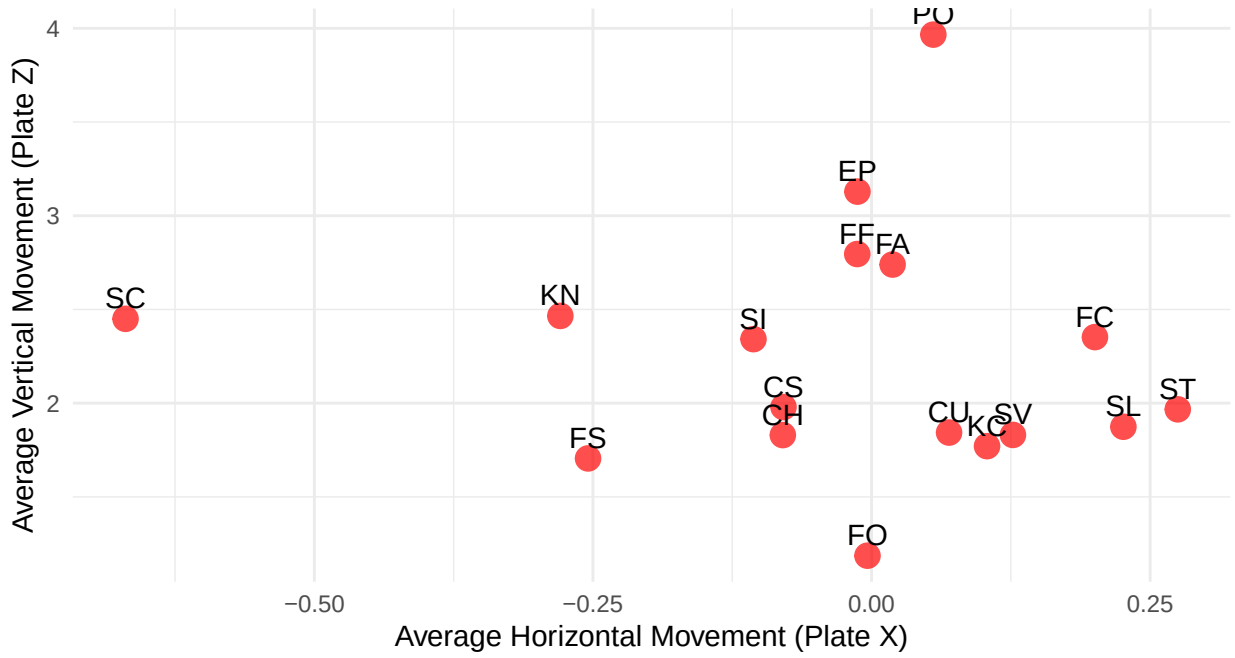

Fig. 1 displays 17 different pitch types that were seen from 2021 - 2023. For putting each pitch type into their pitch group, vertical and horizontal movement and the speed at which the pitch has historically been thrown are used as references. Breaking Balls are characterized by high lateral movement (Table 1), while Off-speed pitches typically exhibit lower speeds or significant vertical movement (Table 2). Fastballs include standard varieties such as 4-seam fastballs and sinkers (Table 3).

Table 1 Breaking Ball Group

Pitch_Type	Description
FC	Cutter
ST	Sweeper
SL	Slider
SC	Screwball
SV	Slurve

Table 2 Off-speed Group

Pitch_Type	Description
CH	Changeup
PO	Pitch Out
KN	Knuckleball
FO	Forkball
EP	Euphus
FA	Other (FA)
FS	Split-Finger
CS	Slow Curve
KC	Knuckle Curve
CU	Curveball

Table 3 Fastball Group

Pitch_Type	Description
FF	4-Seam Fastball
SI	Sinker

2.2 Modeling & Predicting

I employed a Random Forest model, well-suited for multiclass predictions with large datasets. This model reduces overfitting, which is particularly important given the frequency of Fastballs compared to other pitch types. The model was trained using an 80%-20% split of the dataset, utilizing 80% of the pitches from 2021-2023 for training.

Figure 2: Correct vs Incorrect Predictions

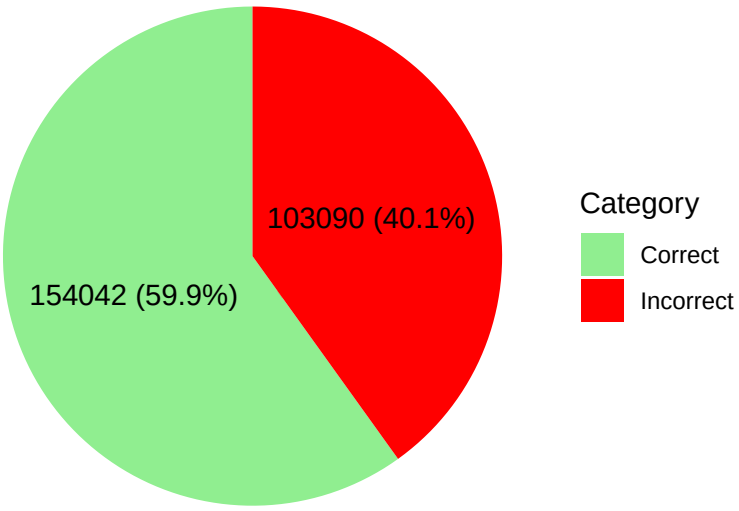

Fig. 2 presents the model's accuracy, illustrating the total number of correct predictions in a pie chart format, with

green indicating correct predictions and red indicating incorrect ones. Despite challenges in achieving high accuracy due to the datset's complexity, the Random Forest model outperformed other tested models

Figure 3: Confusion Matrix Heatmap

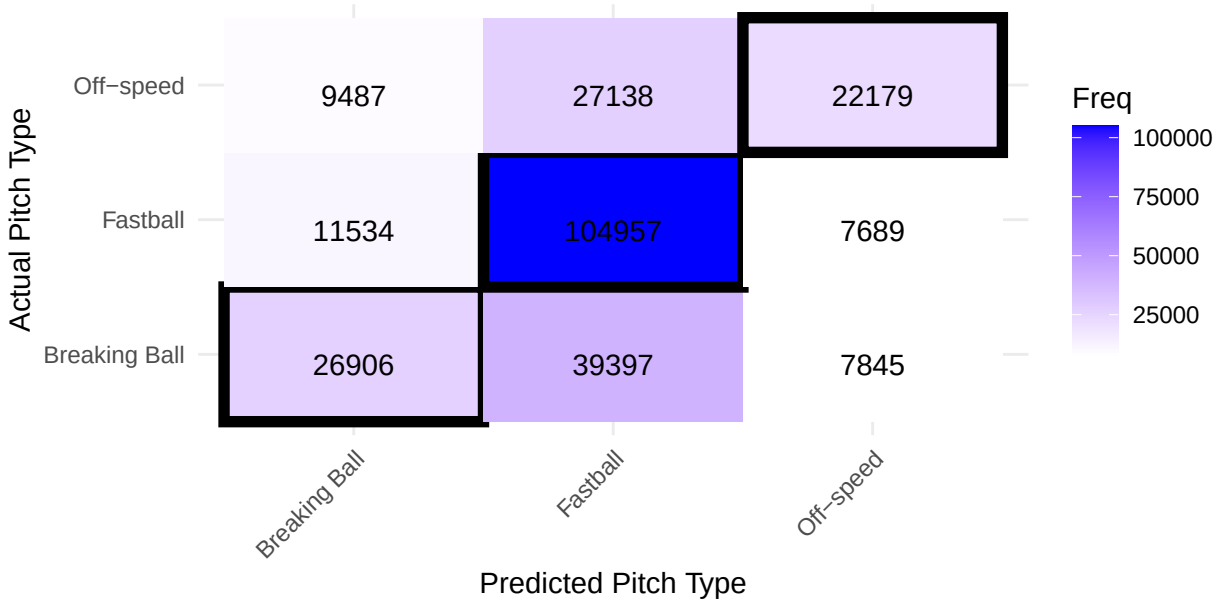

Fig. 3 displays a confusion matrix, which helps identify correct and incorrect predictions. The diagonal boxes indicate accurate predictions, while off-diagonal boxes represent errors. Notably, Fastballs were predicted were predicted incorrectly more frequently than Breaking Balls or Off-speed pitches, suggesting an area for improvement in model accuracy.

Results

To determine the proportions of pitch mixes, we grouped predictions by specific pitch categories for each batter's unique identifier. Each pitch grouping was divided by the total number of pitches expected for that batter in 2024, as shown in Table 4.

Table 4: Ten Player Pitch Mix Proportions

BATTER_ID	PLAYER_NAME	GAME_YEAR	PITCH_TYPE_FB	PITCH_TYPE_BB	PITCH_TYPE_OS
444482	Peralta, David	2024	0.638	0.108	0.254
453568	Blackmon, Charlie	2024	0.736	0.106	0.158
456781	Solano, Donovan	2024	0.684	0.230	0.086
457705	McCutchen, Andrew	2024	0.675	0.243	0.082
457759	Turner, Justin	2024	0.674	0.245	0.081
467793	Santana, Carlos	2024	0.682	0.091	0.227
500743	Rojas, Miguel	2024	0.676	0.256	0.068
502054	Pham, Tommy	2024	0.679	0.243	0.078
502110	Martinez, J.D.	2024	0.642	0.256	0.101
502671	Goldschmidt, Paul	2024	0.661	0.269	0.070

Table 4 provides the expected pitch percentages for the 2024 season: - **PLAYER_TYPE_FB** Percentage of fastballs - **PLAYER_TYPE_BB** Percentage of breaking balls - **PLAYER_TYPE_OS** Percentage of off-speed pitches Something

important to notice is that the percentage of fastballs for each player is relatively similar; however, the percentage of breaking balls and off-speed changes with a high degree of variability between each batter. For instance, David Peralta is predicted to see a Fastball 63.8% of the time, a Breaking Ball 10.8%, and an Off-speed pitch 25.4%.

Table 5 Reds Players Pitch Mix Proportions

BATTER_ID	PLAYER_NAME	GAME_YEAR	PITCH_TYPE_FB	PITCH_TYPE_BB	PITCH_TYPE_OS
641584	Fraley, Jake	2024	0.685	0.092	0.223
663697	India, Jonathan	2024	0.684	0.232	0.084
666181	Benson, Will	2024	0.651	0.121	0.228
670770	Friedl, TJ	2024	0.683	0.085	0.232
682829	De La Cruz, Elly	2024	0.631	0.106	0.263

Table 5 shows the pitch mix proportions for five players from the Cincinnati Reds. Notably four out of the five players are expected to face Off-speed pitches more frequently than Breaking Balls. This suggests that pitchers may tailor their strategies based on individual batter profiles, a trend further explored in Figure 3.

Figure 3 Reds Pitch Mix Proportions by Player

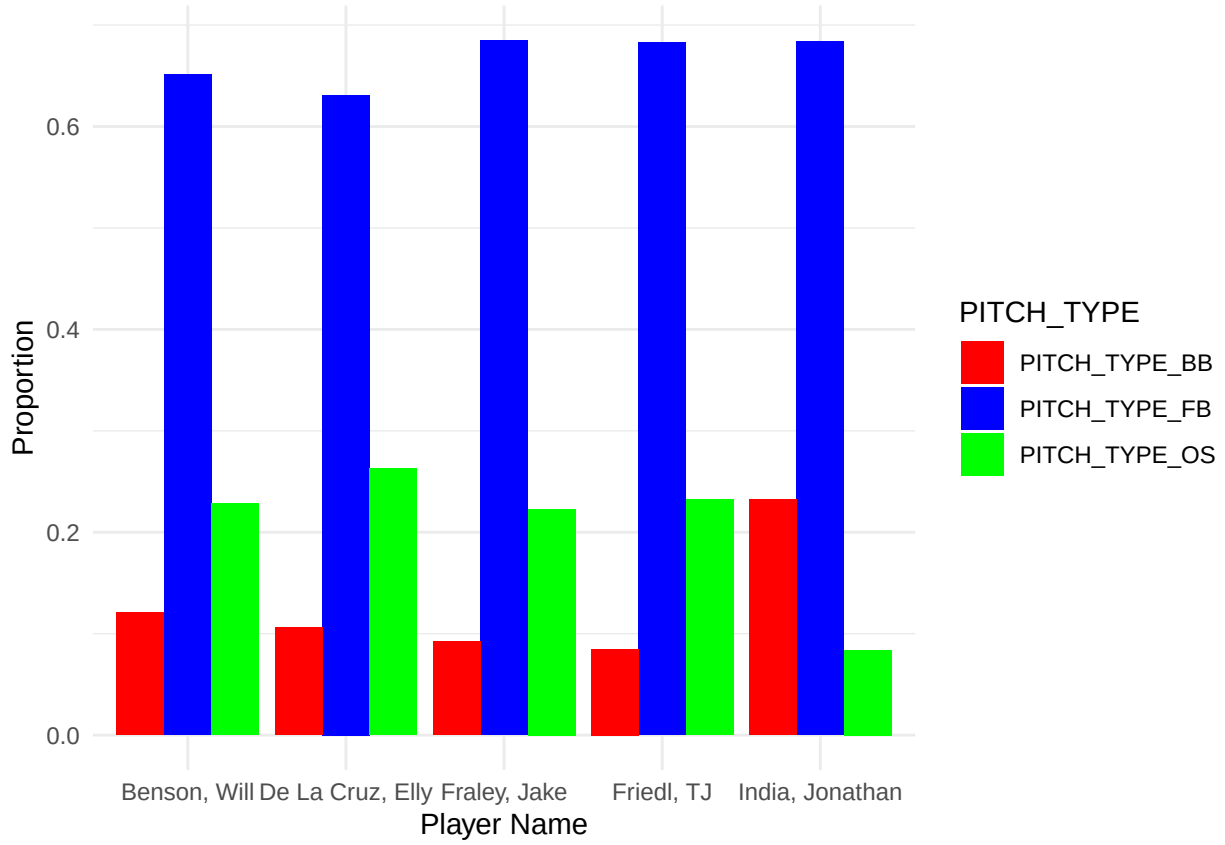

Fig. 3 is a bar graph comparing the expected pitch mixes for each Reds player. Interestingly, left-handed batters such as Will Benson, Jake Fraley, and TJ Friedl, show different expectations compared to right-handed batters like Jonathon India. This indicates the potential for strategic advantages, particularly for switch-hitters like De La Cruz, who can adapt their approach based on pitch type.

Discussion

The Primary objective of this report was to use hitting data from 2021 to 2023 to predict the pitch mix proportions that hitters can expect in 2024. Despite the model's overall accuracy being 60%, clear trends emerged, notably that Fastballs are the predominant pitch type. The Analysis reveals the influence of batter handedness of pitch expectations, providing valuable insights for managers and player development staff. While the model indicates potential trends, the accuracy limitations suggest that the proportion of Fastballs may be overstated. Improved processing systems could enhance the model's performance. Overall, this report provides significant insights into the pitch mix proportions hitters should anticipate, offering a strategic advantage for organizations looking to excel in an increasingly data-driven game.