Batter Pitch Mix Predictor – Coaching Report
Timothy Clay
October 28th, 2024

Introduction

The objective behind this project is to predict what percentage of fastballs, offspeed, and breaking balls a batter might expect to see in their 2024 season. The results of this report also have practical implications, as they can inform the hitters on what aspects of their batting profile may lead pitchers to attack them with certain pitch types, which can help identify and address any weaknesses.

To make my projections, I created three different machine learning models – one for each type of pitch – that predict what percent of pitches the player will see of that type. To create these models, I used data from 2021 through 2023 to see what batter statistics in one year contribute to changes in their pitch distribution for the next. For each model, I used a variety of player-level statistics (swing rate, whiff rate, wOBA, etc.), as well as pitch-specific statistics (swing rate/whiff rate/wOBA/etc. on fastballs/offspeed/breaking balls). Each model outputs their projection independently of the other two, and I scale each of their predictions to add up to 100%.

Dashboard Overview

To better understand the intricacies of my models, I created an interactive dashboard that allows you to experiment with different players and different values of these statistics. This dashboard is split into two parts: the inputs sidebar on the left and the outputs content in the remainder of the screen. The dashboard can be accessed on this webpage: timothyclay.shinyapps.io/pitch-mix/.

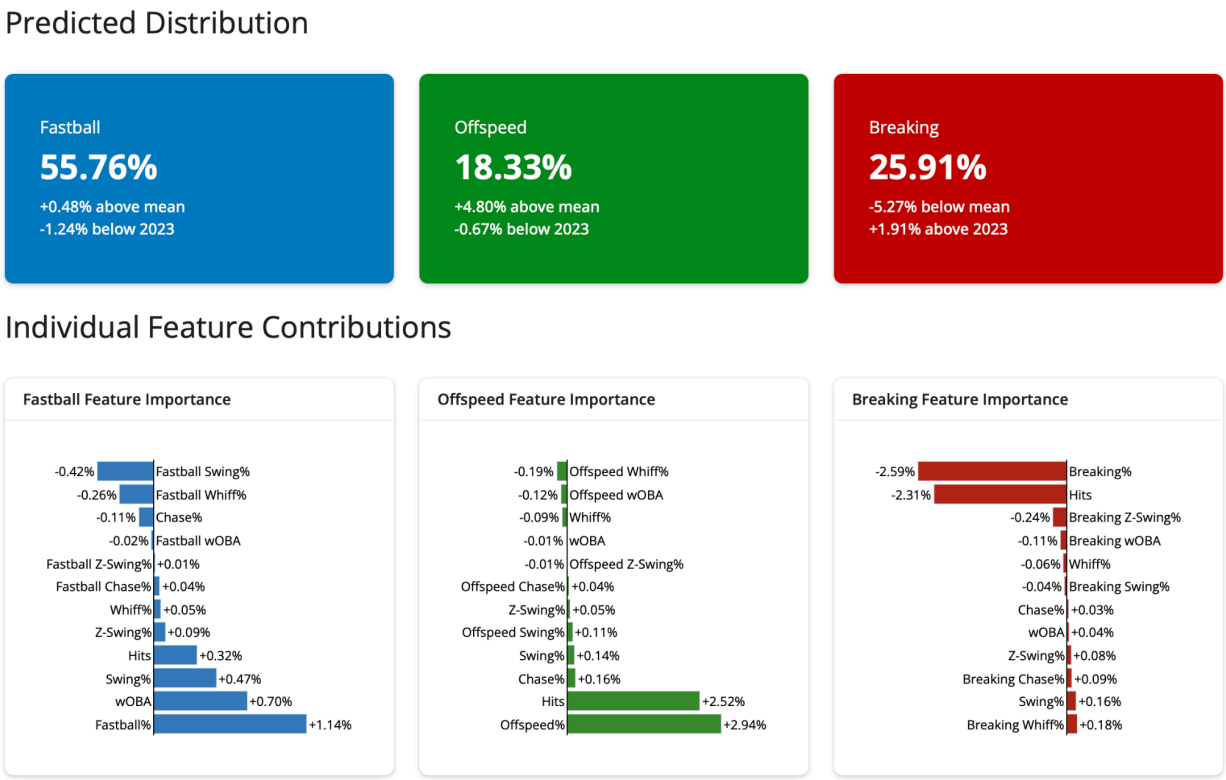
The input sidebar contains every statistic the model considers, grouped into 7 different sections. While you *can* manually input values for each of these statistics, the best way to interface with this sidebar is to choose a player from the dropdown in the “Biographical” section, whose 2023 statistics will be auto-filled into all the fields. I find this feature the most useful, as you can start with an existing player and experiment with what would happen if they, for instance, cut their whiff rate by 50%. The button at the bottom of the input values section will run the models with the provided inputs and the content section of the page will update accordingly.

The content section of the page is broken down into two rows, which are further separated into the different pitch types. The top row shows the predicted percentage of each pitch. In addition to just the raw percentage, it also shows where that percentage ranks against all other players,

as well as against their own percentage from 2023. This provides a sense of whether that percentage is high or low for the pitch type, as well as what the player's overall trend is. In the bottom row are feature importance bar charts, which approximately show how each feature contributes to the model's overall prediction. Each bar is labeled with what the feature is, as well as the percentage that the feature added to or subtracted from the final predictions. Additionally, if you hover over the bars, you can see the value of the feature for the player, as well as where that ranks across the league by percentile. While these numbers are only estimates of feature importance – understanding the *actual* contributions for each feature is nearly impossible for these types of machine learning models – they can be valuable in getting a sense of which features are most driving the outcomes.

For this report, I will be focusing on three different Reds batters – Jake Fraley, Jonathan India, and Elly De La Cruz – and using the dashboard as a tool to analyze their predicted pitch distributions.

Jake Fraley



The first batter I looked at was Jake Fraley. In 2024, according to the models, Fraley was expected to see fastballs 55.8% of the time, offspeed pitches 18.3% of the time, and breaking pitches 25.9% of the time. Comparing those rates against the rest of the league, Fraley's

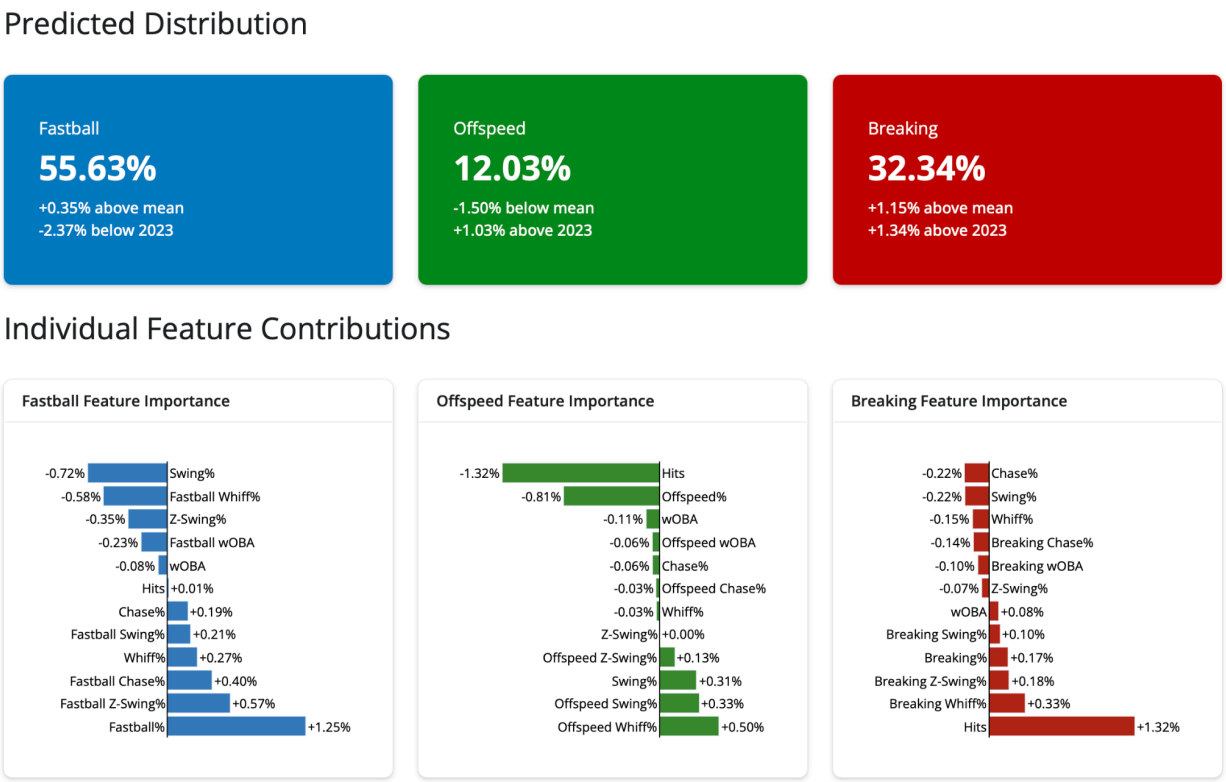
projected fastball rate is fairly normal, but his offspeed and breaking rates differ by about 5% each.

Looking at the feature importance plots, we can start to understand how the models arrived at these projections. For Fraley's fastball rate, the prediction is close to league average due to opposing contributions from a variety of features. The most negative of these features is Fraley's swing rate on fastballs, which, according to the dashboard, is in the 83rd percentile. Such a high swing rate on fastballs might discourage pitchers from throwing the pitch, since Fraley is unlikely to simply take it. This, in turn, may decrease the pitch's effectiveness. On the other hand, the most positive of these features is Fraley's fastball rate from last season. For many players, one of the most indicative features of how they will be pitched to in the future is how they've been pitched to in the past, and Fraley is no exception. The model says that, since Fraley had a 65th percentile fastball rate in 2023, his rate in 2024 will likely also be high.

For the offspeed and breaking feature importance plots, the results are a lot less balanced. Both plots are driven in the opposite direction by two features: that pitch's rate from 2023 and the fact that Fraley hits left-handed. For offspeed pitches, these features suggest that Fraley will see more offspeed pitches than average, while the same features suggest the opposite for breaking pitches. Considering the context that Fraley is left-handed, these values make sense when considering the movement of these individual pitches. For right-handed pitchers, their breaking pitches are typically more effective against right-handed batters, and they might instead choose to go to an offspeed pitch against a lefty. Since most pitchers are right-handed, it would therefore make sense that Fraley's distribution would look like this.

Using the dashboard in this manner can be helpful to understand why a player's distribution may look the way that it does. Analyzing the model's outputs in this way can further help your understanding of how the model works, as well as why a pitcher may choose to attack a hitter a certain way, which are both valuable insights.

Jonathan India



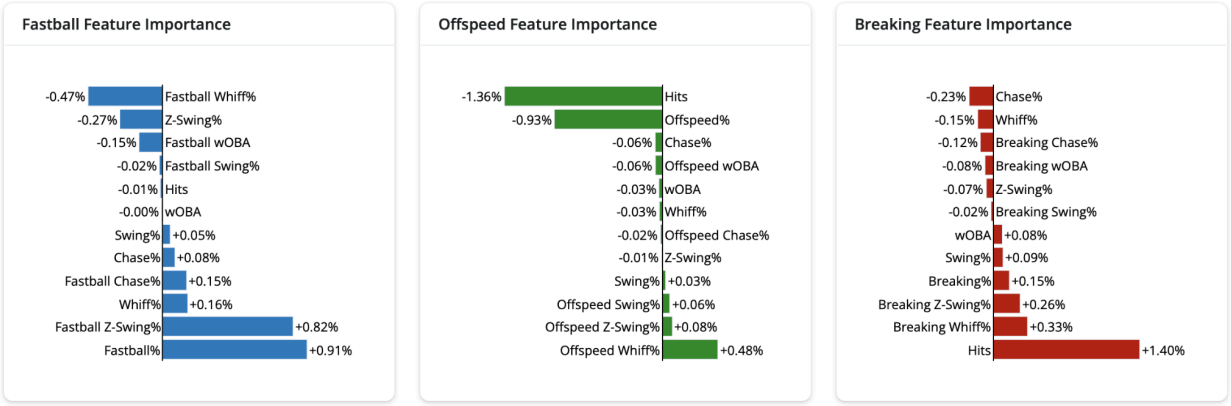
The second batter I focused on was Jonathan India. As seen on the dashboard, India is projected in 2024 to see 55.6% fastballs, 12.0% offspeed pitches, and 32.3% breaking pitches. All of these predictions are relatively close to their respective means across all players.

However, it is worth noting that India is projected to see almost 2.5% fewer fastballs than he did in 2023. As was the case with Fraley, India's largest positive contributor to his fastball percentage is his fastball percentage from 2023. As for the negative features, the model expects to see a decrease in India's fastball percentage due to his extremely low swing rate (5th percentile) and his similarly low fastball whiff rate (22nd percentile). These features play a role in his other pitch types as well, with swing rate and pitch-specific whiff rates playing a large role. Ultimately, it appears that India's patience and his ability to make contact when he does swing greatly impact the proportion of pitches he sees.

Given this, we can further explore India's profile by investigating how his pitch mix might change if he were to become more aggressive and start swinging more. To do this, I updated the swing rates in the input section of the dashboard to be 50% across all pitches and reran the models. Doing so resulted in the output below.

Predicted Distribution

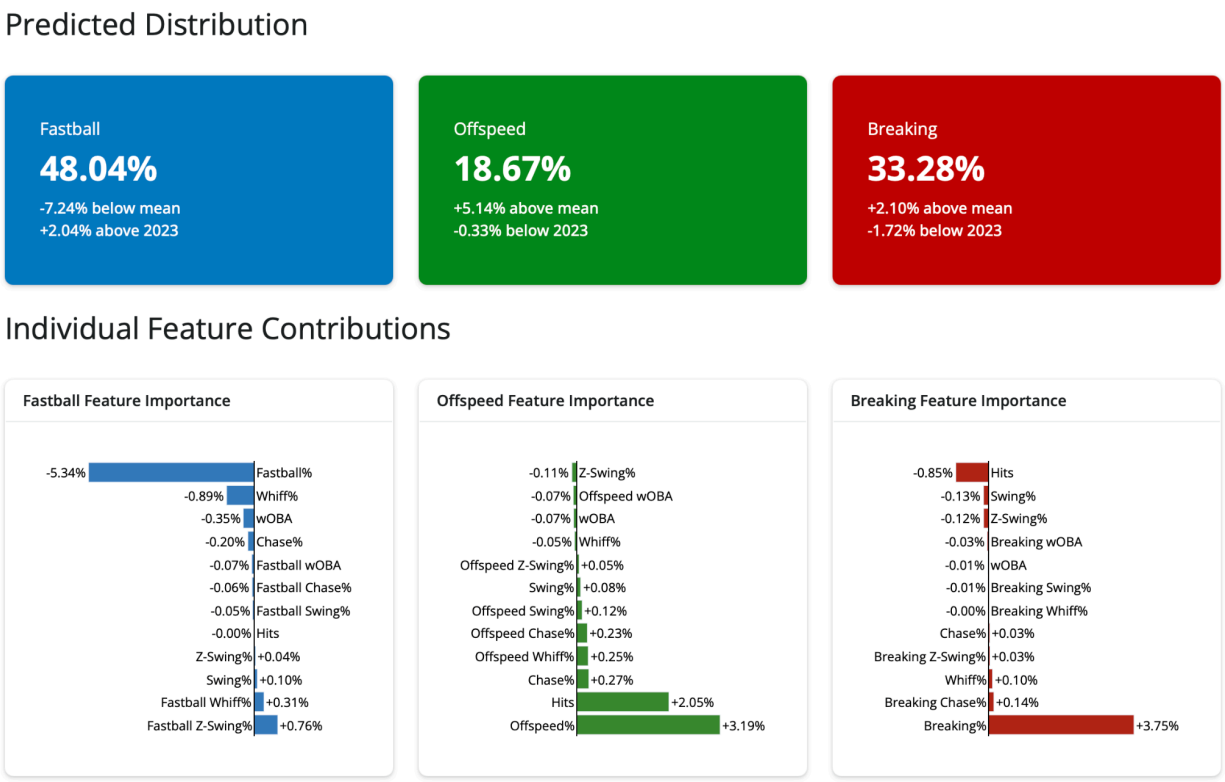
Individual Feature Contributions



As you can see, his projections change, but not drastically. However, the feature importance plots are noticeably different, with swing rate no longer being a negative contributor in the fastball plot, for instance. While not a huge effect, this might indicate that if India were to start swinging more often, he might start to see more fastballs.

This feature of the dashboard can be incredibly helpful to identify what impacts a change to a player's strategy might result in long-term pitch distributions. Additionally, a coach might expect a change in approach would have a great impact on pitch mix, when in reality it might not. This dashboard can be a great way to test that hypothesis before implementing any changes. Both of these use cases can be incredibly helpful in a developmental context.

Elly De La Cruz



The last batter I analyzed using the dashboard was Elly De La Cruz. According to the models, Elly is expected to see 46.9% fastballs, 18.3% offspeed, and 34.8% breaking pitches. While these projections are fairly in line with his previous pitch mixes, his fastball rate is extremely low at 8.5% lower than the mean of all players. Meanwhile, his offspeed and breaking distributions are both higher than league average, at +4.7% and +3.7% respectively.

Looking at the feature importance plots, the model is mostly driven towards these projections due to how pitchers have attacked Elly in the past – the contributions of Fastball%, Offspeed%, and Breaking% features are all the largest for their respective plots. As I mentioned before, these features are large drivers of future pitch mixes for many batters, and in the case of De La Cruz, that effect is extreme. When looking at these distributions on a percentile basis, it makes sense; Elly's 46% fastball rate in 2023 was in the 1st percentile, while his offspeed and breaking rates were 90th and 80th percentile respectively.

Given these feature importance plots, there's not many "what-ifs" that come to mind regarding changing approach. However, Elly is also unique since he is a switch hitter. Because of this, we can further look into Elly's season by evaluating him as both a right-handed and left-handed batter. The resulting pitch mixes are shown below.

Elly De La Cruz as RHH

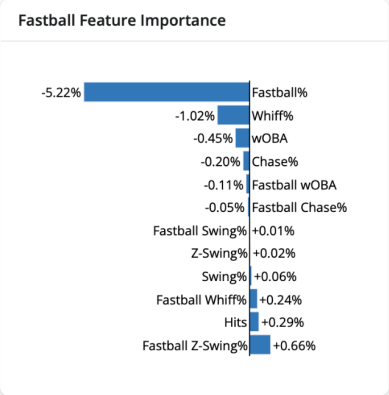
Predicted Distribution

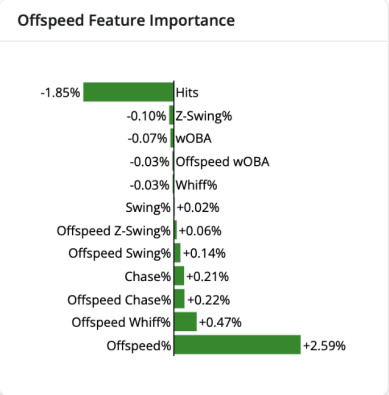
Fastball
49.89%
-5.39% below mean
+3.89% above 2023

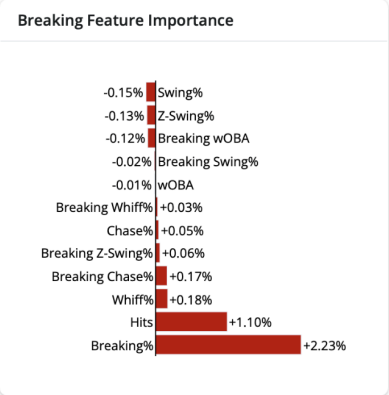
Offspeed
15.04%
+1.51% above mean
-3.96% below 2023

Breaking
35.07%
+3.88% above mean
+0.07% above 2023

Individual Feature Contributions







Elly De La Cruz as LHH

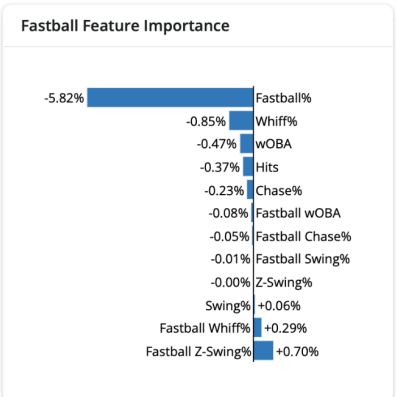
Predicted Distribution

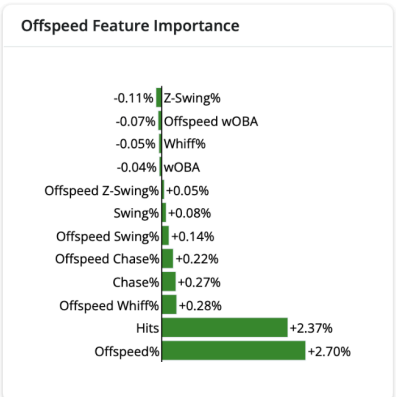
Fastball
48.05%
-7.23% below mean
+2.05% above 2023

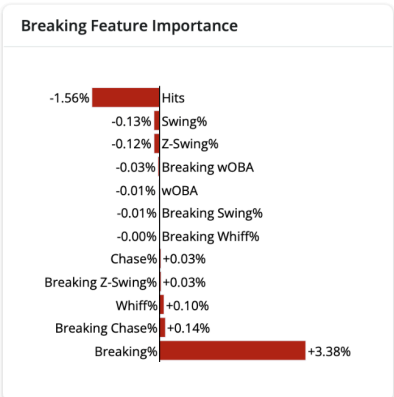
Offspeed
19.00%
+5.47% above mean
-0.00% below 2023

Breaking
32.95%
+1.76% above mean
-2.05% below 2023

Individual Feature Contributions







As a left-handed hitter, Elly's overall pitch mix distribution is not much different than overall as a switch hitter, which makes sense, since he'd be hitting left-handed more times than not. As a righty, however, Elly would start to see a greater percentage of breaking balls than before, at the expense of offspeed pitches. For players with unique quirks such as switch hitting, this side-by-side breakdown can be helpful to understand what they might expect for any given at bat, which is another valuable potential use case for the dashboard.

Conclusion

In summary, this project aimed to predict the percentage of fastballs, offspeed pitches, and breaking pitches that a batter could expect to see in 2024. As you can see in the player examples above, the model provides its audience with a great deal of valuable information that can be applied in a variety of contexts. As I mentioned with the players before, the dashboard can be used not only to get a sense of what's going on behind a player's pitch mix distribution, but also to test theories about different plate approaches and game plan for certain types of hitters.

Additionally, while I tried to demonstrate a variety of features of the dashboard, I highly recommend exploring yourself and playing around with the data. With different players and different feature values, there are an endless number of predictions you can generate, which will only familiarize yourself more with the inner workings of the model. I hope that this model and this dashboard are valuable pieces of insight into the pitch mix distributions of MLB batters.