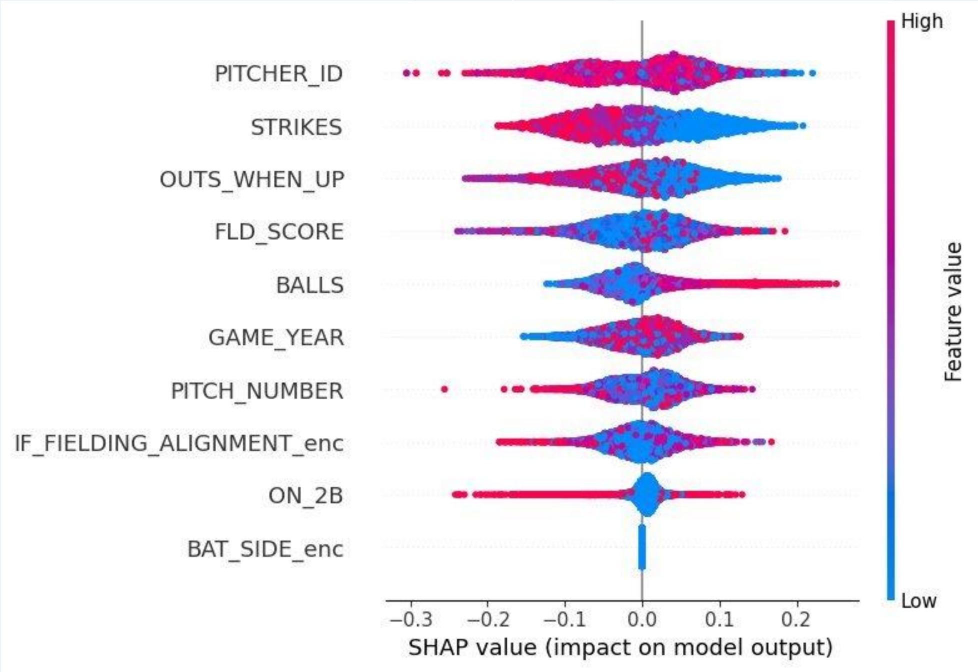
Batter Pitch Mix Predictions for 2024

MOOKIE BETTS CASE STUDY

- From the model, Mookie Betts is projected to receive the following pitch mixes in 2024

Fastball proportion	Breaking Ball Proportion	Off-Speed Proportion
0.726	0.228	0.046

Summary Plot for Mookie Betts



Using a tool called SHAP, we can visualize the influence of each variable on Mookie Bett's probability that he receives a fastball

- Y-axis indicates the feature names in order of importance from top to bottom.
- X-axis represents the SHAP value, which indicates the degree of change in log odds or influence that the model will predict a fastball
- The color of each point on the graph represents the value of the corresponding feature, with red indicating high values and blue indicating low values.
- Each point represents a row of data from the original dataset.

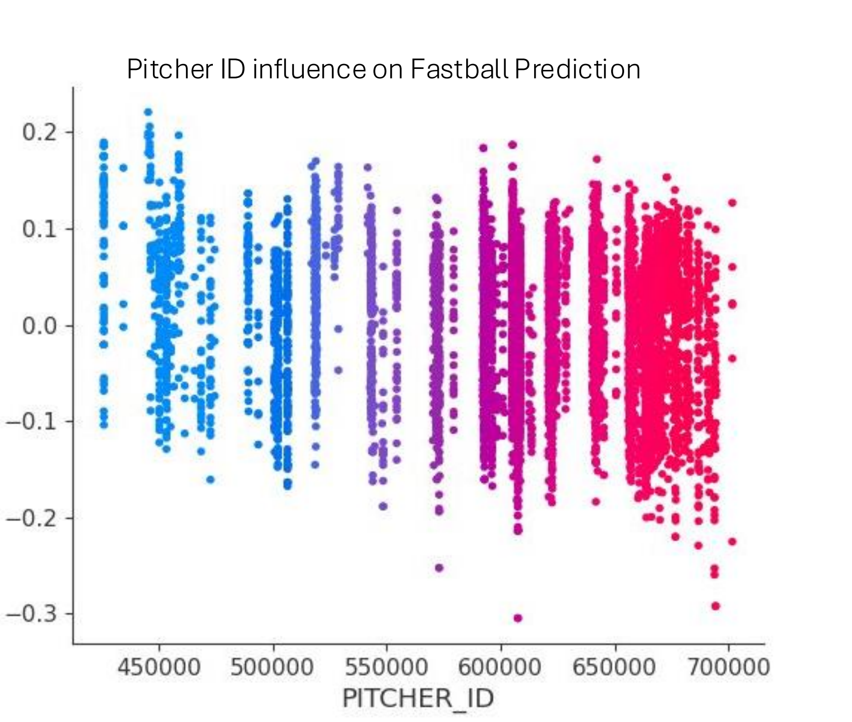
If you look at the feature “Balls", you will see that it is high ball counts that are associated with positive SHAP values. This means higher ball counts tend to positively affect the fastball probability. This makes sense since counts such as 3-0 are commonly fastball counts.

MOOKIE BETTS CASE STUDY PT.2

- The graph labeled "Pitcher ID influence on Fastball Prediction" shows all of the pitcher ID's plotted against their influence on the model's prediction for Mookie receiving a fastball
- Negative terms mean that Mookie Betts is less likely to receive a fastball against that pitcher, while positive terms mean that Mookie is more likely to receive a fastball
- The table below shows the top 3 most influential pitcher (taken from the greatest absolute value y-values in the graph)
  - Thus, we can see that against Kyle Freeland, Brandon Pfaadt, and Bryce Elder, Mookie is expected to see less fastballs since their coefficients are all negative

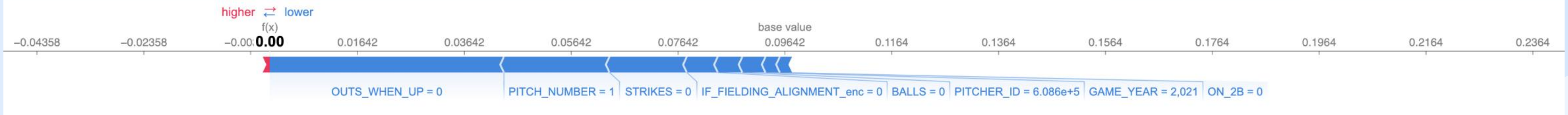
Top Three Influential Pitchers on Fastball Prediction		
Pitcher Name	Pitcher ID	Influence on Fastball Prediction
Kyle Freeland	607536	-0.304755
Brandon Pfaadt	694297	-0.292248
Bryce Elder	693821	-0.259720

Dependence Plot for Mookie Betts against all Pitchers

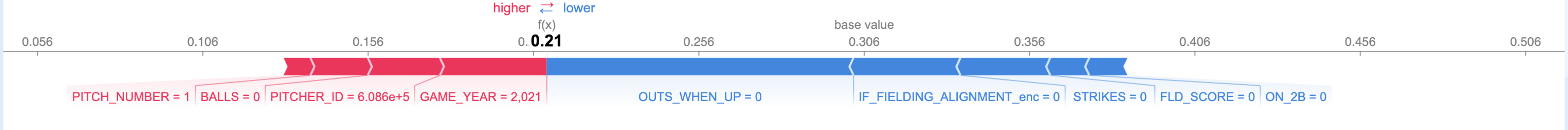


MOOKIE BETTS CASE STUDY PT.3

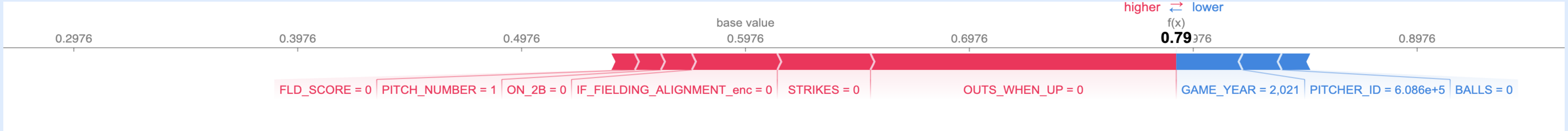
Force Plot for Mookie Bett's Probability for facing a Breaking Ball



Force Plot for Mookie Bett's Probability for facing a Off-Speed Pitch



Force Plot for Mookie Bett's Probability for facing a Fastball



The force plots above gives us a visualization that shows how each feature in a data set contributes to a machine learning model's prediction for a specific observation. For instance, in this case the model predicted that Mookie Betts most likely is going to receive a fastball since its probability (0.79) is higher than the probabilities for a breaking ball (0) and off-speed pitch (0.21). The features that were important to making the prediction for this observation are shown in red and blue, with red representing features that pushed the model score higher, and blue representing features that pushed the score lower. Features that had more of an impact on the score are located closer to the dividing boundary between red and blue, and the size of that impact is represented by the size of the bar. For instance, the fact that Mookie had 0 strikes and there were 0 outs during his at-bat made the biggest contribution to determining that he was going to receive a fastball.

JUSTIN TURNER CASE STUDY

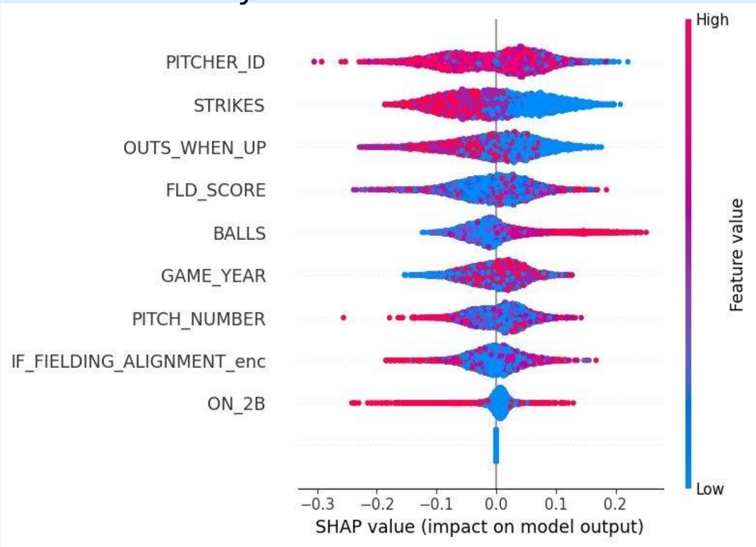
- From the model, Justin Turner is projected to receive the following pitch mixes in 2024

Fastball proportion	Breaking Ball Proportion	Off-Speed Proportion
0.796	0.178	0.027

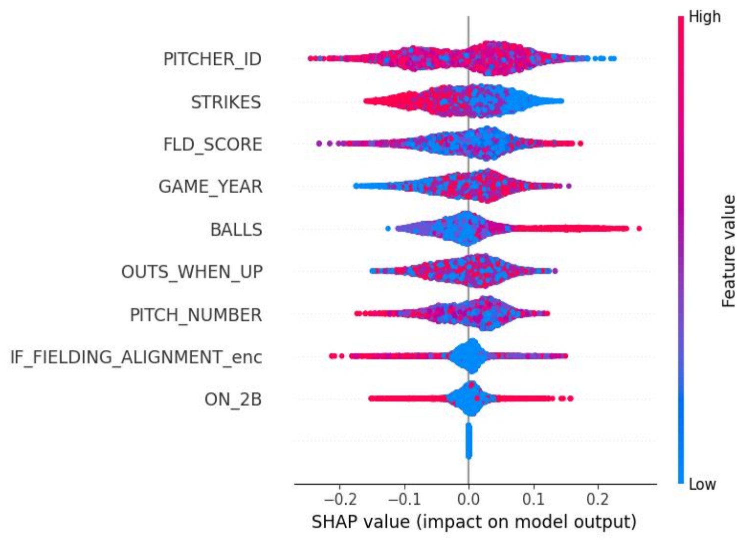
We can now compare the summary plots for Mookie Betts and Justin Turner. For instance, we see that more higher values of *ON_2B* have negative SHAP values for Betts (the points extending towards the left are increasingly red). Unlike Turner's distribution which was relatively the same amount of high ON_2B values for both positive and negative SHAP values. This indicates that Betts is more likely to not see a fastball if there is a runner on second while for Turner, a runner on second is less of a clear tell.

The distribution of points can also be informative. Looking at the OUTS_WHEN_UP for Betts, we see a dense cluster of low out instances (blue points) with small but positive SHAP values. Instances of higher outs (red points) extend further towards the left, suggesting more outs during the at-bat has a stronger negative impact on fastball probability than the positive impact of less outs on price. However, for Turner we don't see this trend as clearly, since his graph doesn't have a left tail that is mostly red.

Summary Plot for Mookie Betts



Summary Plot for Justin Turner

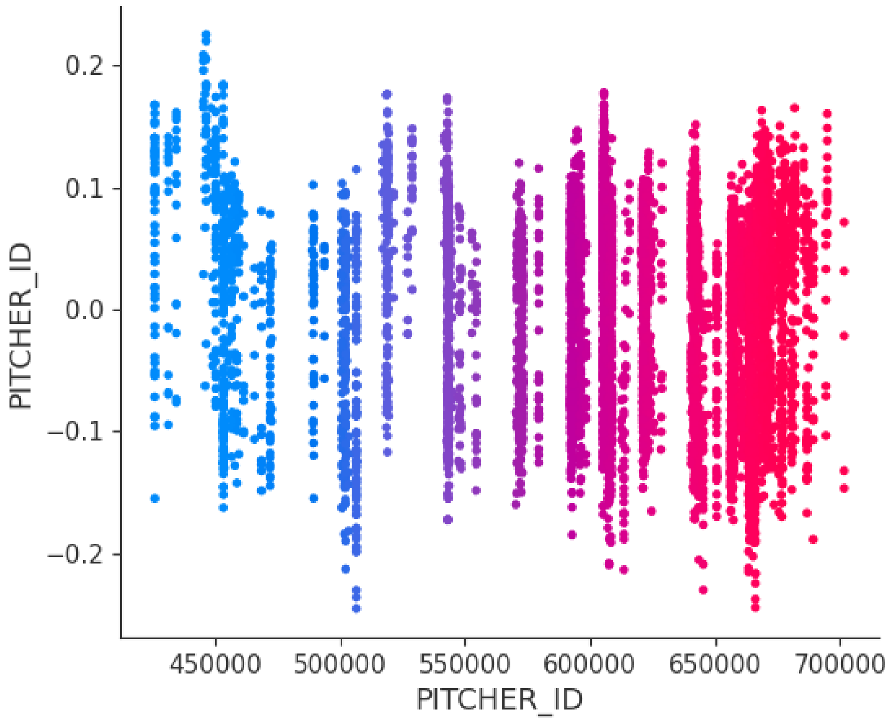


JUSTIN TURNER CASE STUDY PT.2

- The graph labeled "Pitcher ID influence on Fastball Prediction" shows all of the pitcher ID's plotted against their influence on the model's prediction for Turner receiving a fastball
- The table below shows the top 3 most influential pitcher (taken from the greatest absolute value y-values in the graph)
  - Thus, we can see that against Yu Darvish, Matt Manning, Cole Ragans Turner is expected to see less fastballs since their coefficients are all negative
  - These pitchers are different than Betts' suggesting that Turner and Betts have different pitchers that throw less fastballs against them
  - Therefore, if they see these pitchers more often during the season, they are likely to see a smaller proportion of fastballs during the season

Top Three Influential Pitchers on Fastball Prediction		
Pitcher Name	Pitcher ID	Influence on Fastball Prediction
Yu Darvish	506433	-0.24551
Matt Manning	666159	-0.24465
Cole Ragans	666142	-0.23773

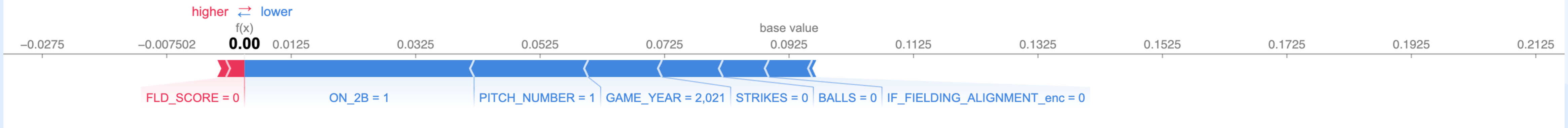
Dependence Plot for Justin Turner against all Pitchers



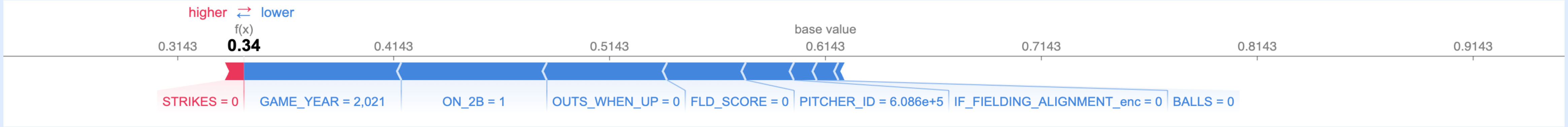
JUSTIN TURNER CASE STUDY PT.3

Force Plot for Justin Turner's Probability for facing a Breaking Ball

Force Plot for Justin Turner's Probability for facing a Off-Speed Pitch



Force Plot for Justin Turner's Probability for facing a Fastball



The force plots for Turner can also gives us a visualization that shows how each feature in a data set contributes to a machine learning model's prediction for a specific observation. For instance, in this case the model predicted that Turner is most likely going to receive a breaking ball since its probability (0.65) is higher than the probabilities for an off speed (0) and fastball pitch (0.34). Looking at what factors most influenced the prediction that Turner will receive a breaking ball, it appears that having a runner on second was the most influential. This makes sense since, when a runner is in scoring position, especially at second or third base, a single hit or a well-placed ground ball can easily result in a run. This puts additional pressure on the pitcher to make pitches that reduce the chances of solid contact or prevent the ball from being put in play in a way that could advance the runner. Breaking balls induce weaker contact or ground balls, reducing the likelihood of extra-base hits or deep fly balls that could score the runner.