



Predicting Hits With AI

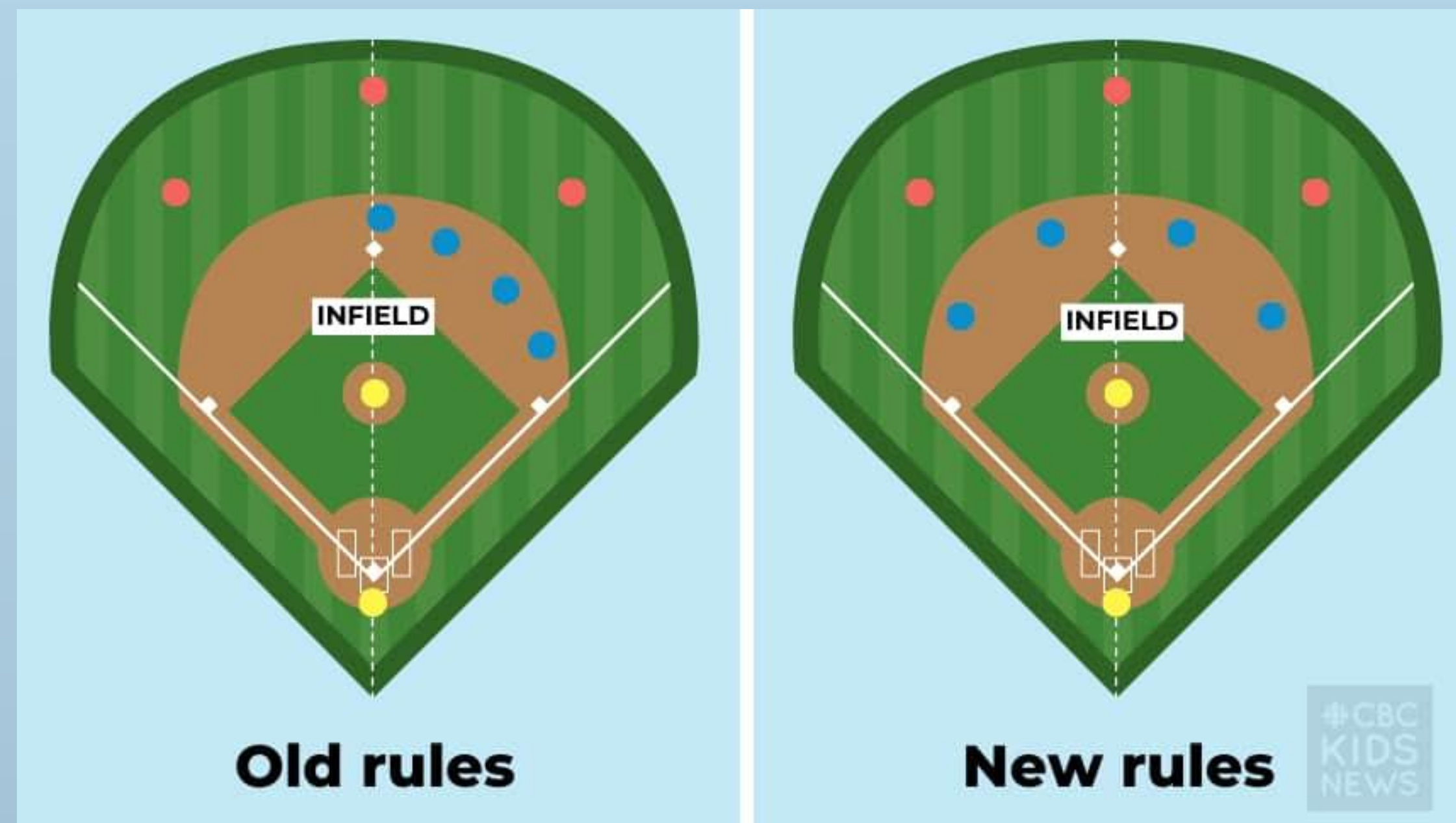


Student: Samuel Trumbley, SCSU

Thesis Advisor: Dr. Winnie Yu

Abstract

I present an analysis of the MLB's new rules by training a machine learning model on 2023 batted balls, then predict 2022 batted balls and compare to the actual data from that season.. The data from the two seasons were collected from Baseball Savant.



Introduction

The MLB decided to implement several new rules for the 2023 season that would hopefully lead to more offense and action, bringing in viewers and combating the trend of slower games. These rules include a pitch clock, larger bases, and the banning of infield shifts. In the first season, statistics like stolen bases, hits, and batting average all showed an increase from the 2022 season to the 2023 season.

	2021	2022	2023
BA	.244	.243	.248
Hits	39484	39675	40839

References

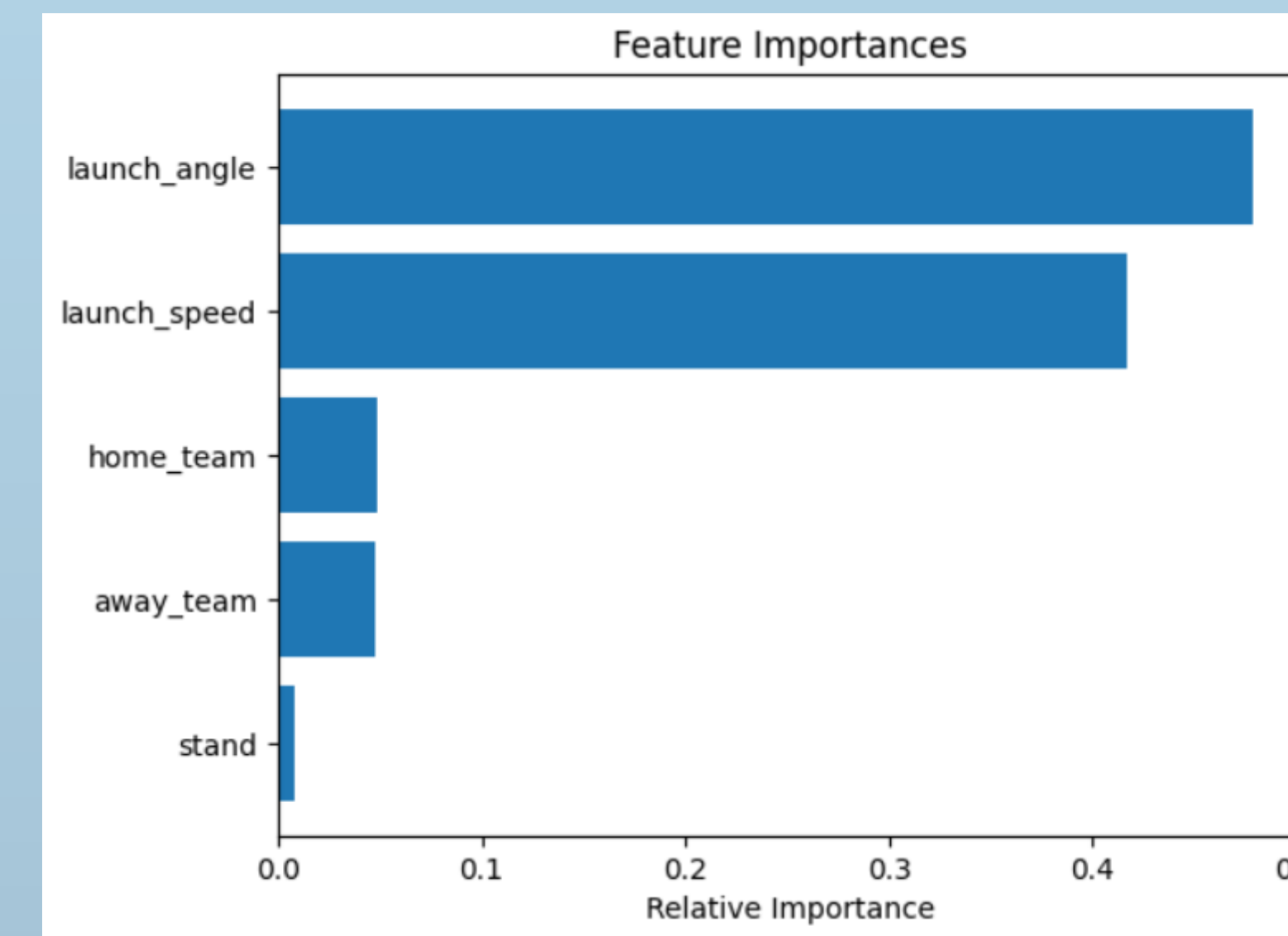
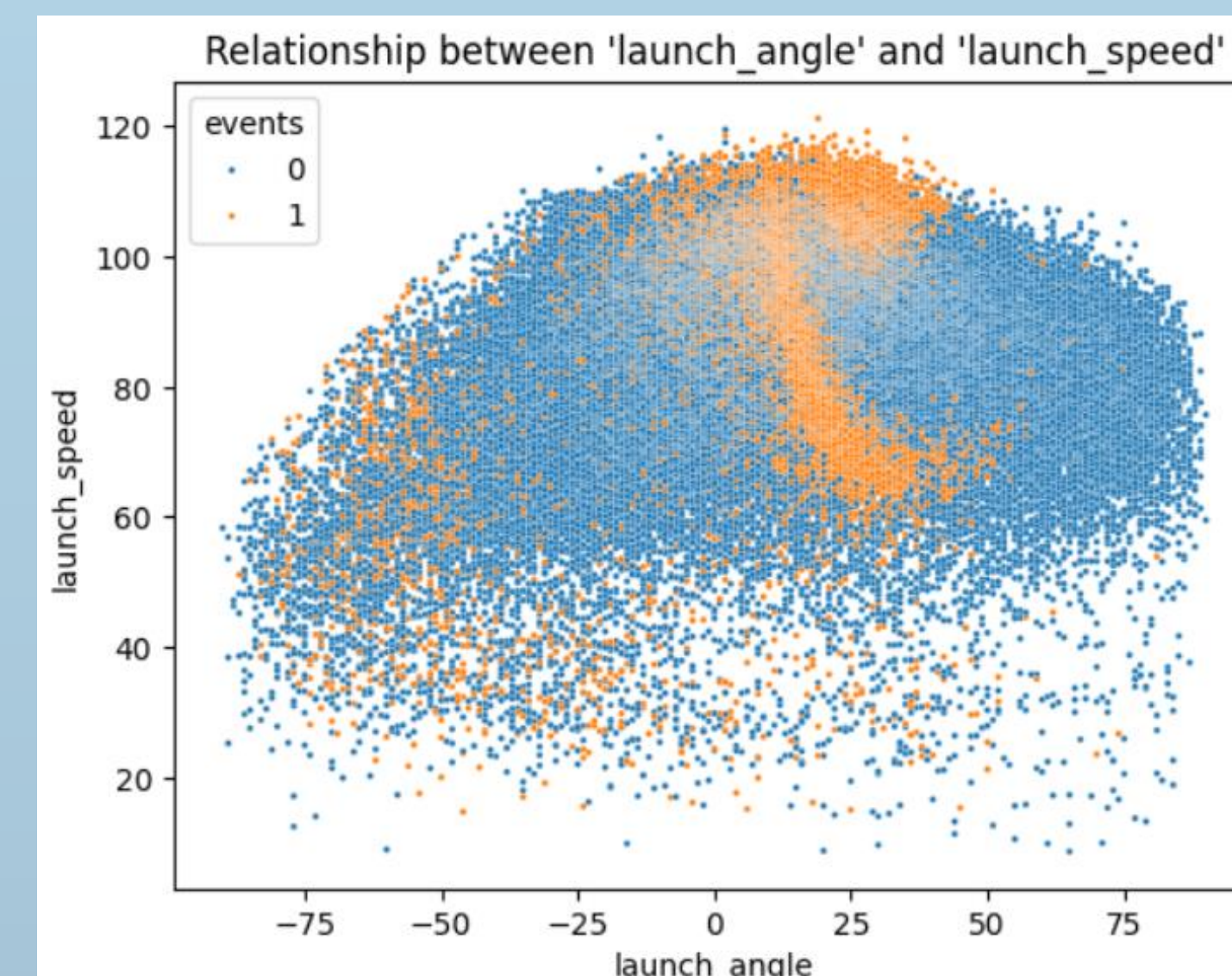
"Major League Baseball Logo." *Wikipedia*, Wikimedia Foundation, 27 Oct. 2023, en.wikipedia.org/wiki/Major_League_Baseball_logo.
 Mandala, Love. "Download Baseball Ball Transparent for Free." *Vecteezy*, Vecteezy, 20 Feb. 2023, www.vecteezy.com/png/13362729-baseball-ball-transparent.
 "Will New MLB Rules Make Baseball More Fun to Watch?" *CBCnews*, CBC/Radio Canada, www.cbc.ca/kidsnews/post/will-new-mlb-rules-make-baseball-more-fun-to-watch. Accessed 9 Apr. 2024.

Classifiers

KNN
 Decision Tree
 Naïve Bayes

Features

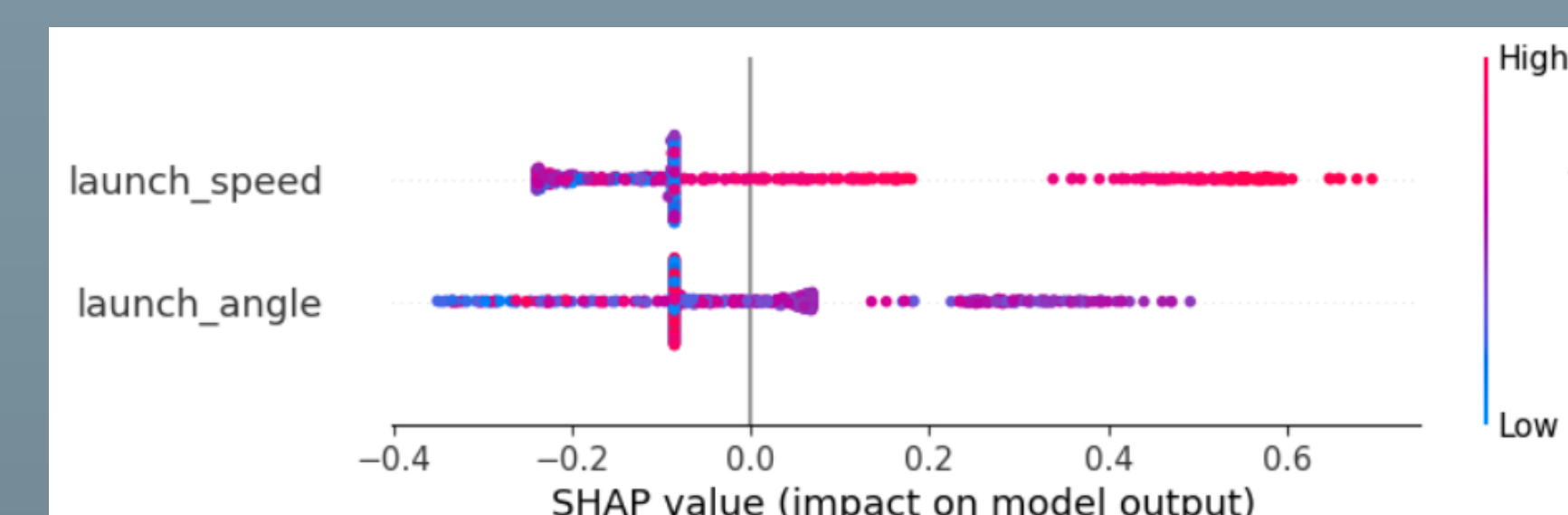
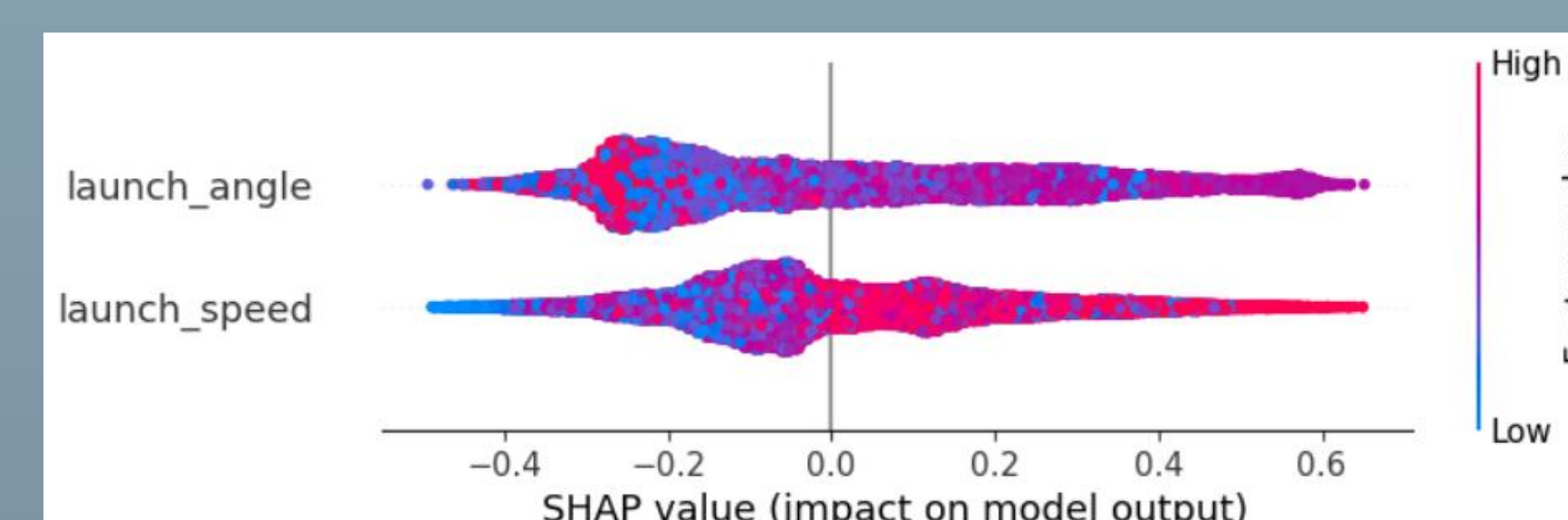
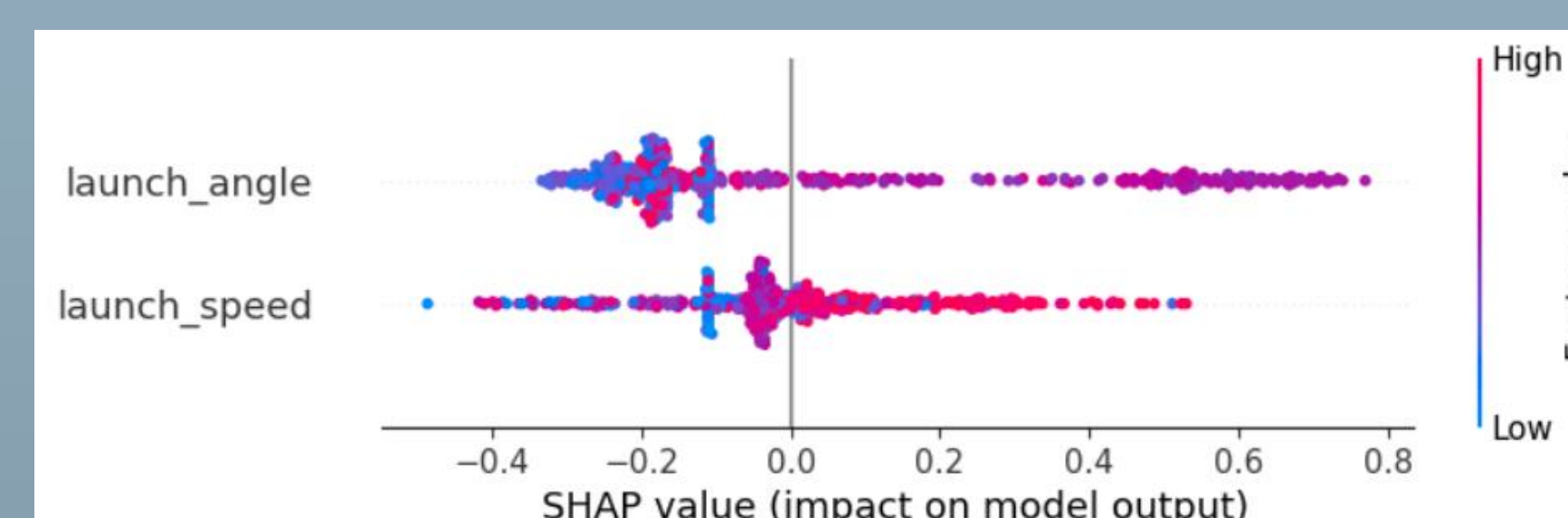
Launch Angle Home Team
 Launch Speed Away Team
 Stand (Lefty/Righty)



Prediction Performance

The following is the performance of the three classifiers when predicting batted balls in 2023. The overall performance of the model was at the expected level. Overall, KNN was the best predictor, leading in virtually every evaluation.

	Accuracy	Precision	Recall	F1 Score
KNN	0.80	0.74	0.61	0.80
DT	0.73	0.59	0.61	0.73
NB	0.75	0.64	0.55	0.75



KNN	0.90	0.10
	0.39	0.61
DT	0.80	0.20
	0.39	0.61
NB	0.85	0.15
	0.45	0.55

Results and Conclusion

When predicting hits from 2022, both KNN and NB predicted **LESS** hits than there were in 2022.

	2022 Predicted
KNN	32592
DT	41085
NB	33299

2022 Hits
39136

Note: The discrepancy in 2022 hits is due to NaN values in the dataset

To better evaluate the performance of the classifiers, the same model was replicated, instead using training data from the 2022 season.

	Trained on 2023	Trained on 2022
KNN	32592	31188
DT	41085	39607
NB	33299	27395

This showed that the model was predicting more hits in 2023 than 2022, confirming the idea that the new rules would lead to more predicted hits. The difference between the KNN and NB predictions are likely due to outs being easier to predict than hits. KNN's high prediction seems to be a result of how the classifier makes its prediction, possibly leading to its lower performance.

This result further confirms that the new MLB rules were successful in increasing offense.

Further Research/Questions

- Improving the prediction of batted balls
Different features, different classifiers, more data, etc.
- More complete analysis of features and predictions
Why does KNN predict more hits?
 - Looking at data from 2021 and earlier
- Will the 2024 season continue the same trends?
- Will teams find any loopholes in the new rules?
- Could the MLB make more changes to further increase offense?