Vivek Divakarla

In predicting pitch mix for the 2024 season, using a Random Forest classifier allowed for modeling of complex interactions within pitch-related data. Random Forest is a machine learning technique that builds an ensemble of decision trees, each trained on a different subset of the data. The ensemble then averages these predictions, improving both model accuracy and stability. This approach works well for pitch prediction because it can effectively capture non-linear relationships and interactions between variables, such as pitch location, batter and pitcher hand dominance, and game context like balls and strikes. These elements often influence pitch selection in ways that might not be evident with simpler models.

The model built on the “all balls and strikes” data provided insights into pitch decisions throughout at-bats, while the model based on batted ball data added a layer of analysis focused on pitches that resulted in contact, providing further nuance on pitch patterns and batter outcomes. Selecting columns like Weighted On-Base Average and Launch Angle were particularly important for the batted ball model because these statistics correlate closely with hit quality and batter success, helping the model detect trends in pitch types that batters tend to put into play with varying degrees of success. Features like Plate_X and Plate_Z, which capture pitch location, allowed the model to analyze the strategic positioning of pitches. The final model result—54.7% fastballs—was close to the actual fastball rate observed in 2024, reinforcing that the model captured meaningful patterns in pitch choice.

The Random Forest model's use of previous data provides predictive insights by learning from established patterns in pitch usage. For instance, it captures how a pitcher might use fastballs more in certain counts or in specific batter-pitcher matchups. This historical data-driven approach enables predictions aligned with real-life trends and tendencies, potentially guiding hitters and coaches in preparing for anticipated pitch mixes in various game situations.