

The aim of this project is to attempt to determine whether the pitch type (i.e. 4-seam fastball, slider, change-up, etc.) can be predicted in advance. Specifically, we attempt to use game state information, namely the score, inning, the positions of the runners, the count, and the number of outs, to predict what pitch is coming. This project was inspired by the Houston Astros, who were recently caught using a hidden camera to intercept the catcher's sign and relay the information to the hitter.

We decided to change pitch type to a binary classification: 1 for "hard" pitches, those generally thrown at high velocity, and 0 for "soft" pitches, pitches that are not thrown with as much velocity, but usually have high spin rates. This simplification allows for the model to compare pitchers with different arsenals. It is also not entirely inaccurate for relief pitchers, many of whom only throw two pitches: one fast and one slow. We also created a model to confirm how accurately pitches can be classified, after the fact, based on spin rate and velocity.

Ultimately, we found that while our model has more predictive power than a baseline model (i.e. one that predicts fastball all the time), it does not perform quite as well as a camera trained on the catcher. We were able to conclude that those with moral scruples can predict pitch type with some accuracy. Future models can improve upon this one by including player data, such as a pitcher's FIP using a certain pitch, which was not part of the data set.