Connor Phillips

Data Scientist

May 15th, 2023

General Assembly

Executive Summary - Baseball Prediction App

This project was created with one goal in mind, to create the most user interactive modeling application for Major League Baseball (MLB) data. In order to carry out this task, I acquired data from a package in python named Pybaseball. Pybaseball has an enormous amount of data for MLB players, with over three hundred different metrics for both pitchers and batters. With so many metrics, only those that were most recognizable to a user with limited background information about Baseball were chosen. In addition, different metrics are provided based on the user's selection of a pitcher or batter.

While the current version of the project that is available today still has room to improve in future iterations, the project was still a success by large. The application allows for the user to select between pitching or batting data, and then head right into testing their own models. The user is allowed to choose up to ten variables to predict with and one variable to try and predict. After selection, the application provides a few metrics to inform the user of the effectiveness of their model.

The application also features forecasting, this feature is more specific to individual players. The user will be able to select a player from over 150 choices, and with the use of time series forecasting, be able to predict their statistics for the following season. This feature also allows the user to pick three metrics to predict from a list, following the trend of interactive modeling.

Overall, the application is a success in the goals that it set out to achieve. Improvements can be made to the application however, starting with an increase in the number of metrics available to the user to play around with. It should also be noted that the forecasting is limited in the amount of data available, as there are only so many seasons that an individual player can play. A future goal will be to acquire monthly data for these players, dramatically increasing the amount of data available and allowing forecasting to move to a monthly dynamic.