Baseball Pitchers Decision Tree

Drew Gotbaum, Eli Rosenberg, Mitch Novak, and Dennis Kachintsev

November 16, 2014

ABSTRACT

Project Type: Programming

Statement: Given a dataset of training data, we would like to predict the success of any given pitcher over the course of their career to come.

General Approach: Having collected and manipulated a dataset of training data, we will construct a decision tree to determine the pitcher's success on a year-by-year basis. We will be able to estimate future performance.

Data: There is an abundance of baseball databases available for our program, but we have not determined which is the optimal data set given our project's demands.

Evaluation: We will be using past player performance to determine their furure performances that are already known to produce a margin of error.

Background Reading:

- ★ FiveThirtyEight is statistician Nate Silver's blog consisting of a very resourceful sports section with different approaches to predicting different data points of the game.
- ★ The Bill James Historical Baseball Abstract offers different revolutionary formulas for predict the prduction of a player and the success of a team.

THE TEAM

Research (November 21) Look into the history of baseball prediction, similar artificial intelligence projects within the field, and the many sports blogs for potential formulaic ideas.

Architect (November 25) Given the datasets at our disposal, the architect must construct object oriented data structures to easily hold, manipulate and interact with the data.

Statistician (November 27) Analysis of the research, which will be used to help structure the formulas used within our algorithm.

Algorithm Design (December 12) Construct the optimal decision tree to work alongside the 'ideal' algorithm that we will choose. From the decision tree, expand the algorithm to take our tree as input.

Testing Engineer (December 15) Provide multiple test cases, consisting of edge cases, recent players, and 'outlying' players.

We have adhered to the honor code on this assignment.