Baseball Informatics – from MiLB to MLB Debut

Chung-Hao Lee & Woei-jyh (Adam) Lee
University of Maryland, College Park, Maryland 20742, USA

Abstract

Drafted baseball players typically begin their professional baseball career with Minor League teams, and are not guaranteed opportunities in the Major League. Accurate estimation of players' likelihood to advance to the Major League debut can reduce the cost and increase value for both players and franchises. We mined both baseball performance stats and non-baseball data of players drafted from 2001 to 2010. We applied machine learning techniques to analyze and rank stats and data variables. We compared four sets of variable selections to train and validate our models, which predict the likelihood of a drafted player reaching the Majors. We fitted extreme gradient boosting, random forest, decision tree, and support vector machine to determine the high impact variables in the prediction. We successfully translated our model results into guidance for drafted players in the Minor League on what they should improve to increase their chance to play in the Major League.