

Applying the Moneyball approach to assemble a winning roster

Faiz Rasool & Viet Dang

STAT 306 - Spring 2024

Introduction

In the 2002 MLB season, facing budget constraints and the departure of key players, the Oakland Athletics's general manager Billy Beane decided to turn to a data-driven strategy, focusing on sabermetrics to uncover hidden-gem players—players whose potential value did not match their market value. The decision revolutionized the baseball, making the A's immediately competitive again. Often referred to as *Moneyball*, Billy Beane's approach changed the way MLB front offices work even until now. Placing ourselves in a hypothetical situation of drafting for an MLB expansion team, we used the Moneyball philosophy to uncover hidden gem players using machine learning techniques.

Methods

- **Data**

The data used for analysis were collected from *fangraphs*, which included all players' performance statistics from 2017-2023. The data was split into three main sections: batters, starting pitchers, and relief pitchers. The selections were decided to be centered around WAR (Wins Above Replacement), often considered as a metric that best summarize the impact of a player on the team.

- **Techniques**

Linear regression analyses determined the important factors that significantly contributed to predicting WAR for each group of players. Neural networks were used to predict players' WAR in the 2024 season, given their 2023 performance.

Results

wOBA, SLUG, and BsR were important factors in predicting batters' WARs. IP, K/9 and HR/9 were important factors in predicting pitchers' WARs. Analyses resulted in a 23-man roster, spread evenly over the main different positions.

Conclusion

Our analyses provided an interesting experiment for a hypothetical draft scenario. It is worth noting that this approach was purely number based and performance based, and did not account for intangible factors such as squad harmony and players' morale. This highlights the necessity of a human decision with deep knowledge of the sport and the league in order to come up with the best draft decisions. Statistical analyses remain a tool to assist decisions, and cannot decide entirely.