

Forecasting the Major League Baseball Draft Success: A Predictive Model for the Cape Cod League Talent

Fr. Humbert Kilanowski, O.P., John Incantalupo, Emily Gelchie, Gianpaolo Colannino, Colin Fitzgibbons, Cameron Gouveia, Ryan Biancavilla

Math and Computer Science Department, Providence College

Introduction

WAR, or wins above replacement, is an all-encompassing statistic that seeks to measure a baseball player's total contributions to his team in the form of a single number. This number represents how many wins a team won because they had that particular player instead of a 'replacement-level' player. The Cape Cod Baseball League is a summer collegiate league in Cape Cod, Massachusetts that has operated for over 100 years. This league is known for attracting top talent across college baseball, as many Cape League players go on to play professional baseball.

Data Overview and Preparation

Utilizing TrackMan's detailed ball-tracking technology, we've deepened our player analysis for the Cape Cod Baseball League, capturing intricate play details that traditional stats miss. Our initial data faced player name discrepancies, which we resolved using Tableau Prep's grouping functions, thus sharpening our dataset for more precise predictive modeling.

Predictive Results

X	Top 10 Batters	Top 10 Pitchers
1	TRAVIS BAZZANA	CAMRON HILL
2	JO OYAMA	DEREK CLARK
3	CAMERON SMITH	SMITH PINSON
4	DEREK BENDER	TRISTAN SMITH
5	COLE MATHIS	CAM SCHUELE
6	MATT HALBACH	SEAN MATSON
7	HUNTER HINES	JOEY DECHIARO
8	JEFFERY HEARD	JOSH TIMMERMAN
9	JAMES TIBBS	FINNEGAN WALL
10	DERIC FABIAN	CHRISTOPHER KAHLER

Figure: Top 10 Batters and Pitchers Predicted (Overfitted Model)

Barrel Plots

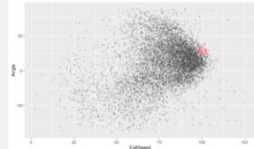


Figure: The relationship between a batted ball's exit speed (in mph) and launch angle (in degrees)

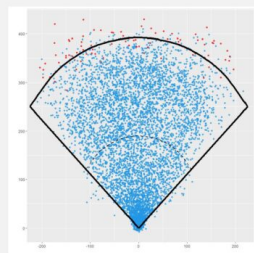


Figure: Hit location of every batted ball from the 2023 Season (Barrels in red)

Methods

We aimed to use WAR, along with other key baseball statistics, to generate a model that can predict the draft position of a young baseball player. The model would accurately predict the value of a player to their Cape League team based on their individual contributions. The objective is to show the great value that baseball statistics hold, and specifically the success of our model in indicating tangible results from summer league stats.

Strike Zone Plot

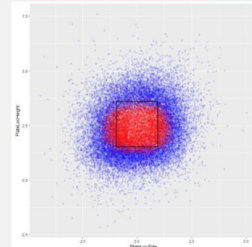


Figure: The location of every pitch tracked during the 2023 Cape League season, color-coded based on whether the pitch was called a strike (red) or ball (blue)

Importance of Variables

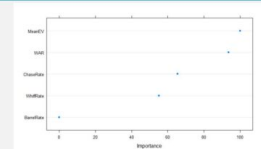


Figure: Significant Variables in the Batters Prediction Model

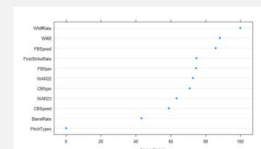


Figure: Significant Variables in the Pitchers Prediction Model

Conclusion

After working with multiple models including ridge regression, lasso regression, and elastic net, we found the linear model to have the best fit. We found that both pitching and hitting models were overfitted, due to the whiff rate. The predicted values for draft pick for each of the three models tested are concentrated between the 100th and 500th picks, which will not necessarily reflect the actual draft position.