RISE RESEARCH INNOVATION SCHOLARSHIP ENTREPRENEURSHIP

Mentor Name. : Dr. Gerber
Subject/Category : Machine Learning
Class : DS4420 Section 02
Group # : #

Northeastern University

Comparison & Prediction Of NBA Player Success Based On Combine Performance

Ben Rice, Derrick Gan, Amaey Pandit

Motivation

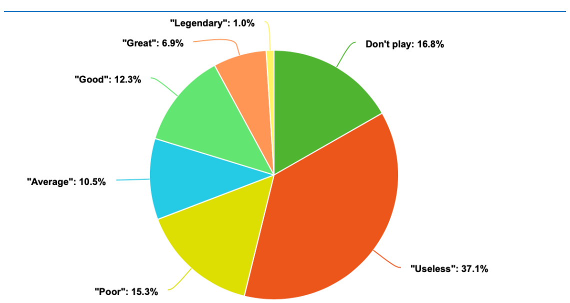
- As technology has evolved so has sports, working tangentially with more advanced analytics every day to measure player, team, and sport success.
- Significant research has gone into understanding the current NBA players trajectory whether it be in competition or for personal gain **but what about those that aren't in the NBA yet**?
- It's hard to predict the success of a prospective NBA player on more than their college stats - while those are good identifiers, **what's the point of the Combine then? We aim to use the NBA combine analytics to predict a rookies success in the NBA** to better suit teams for drafting with a glimpse into the future rather than a feeling.

NFL DRAFT

NFL Draft Pick Bust Rate Remains Very High

More analytics haven't increased the success rate

By Warren Ludford | @wludford | Apr 26, 2022, 8:53pm CDT | 48 Comments / 48 New



*Example of NFL bust rates, like the NBA showing how analytics aren't considering enough variables

Process And Methods

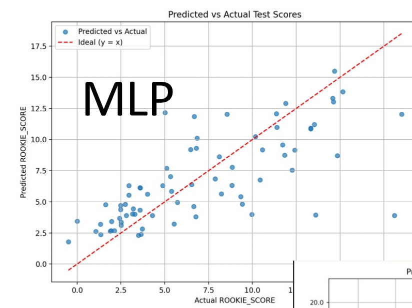
- We used machine learning methods of Collaborative Filtering and MLP to find comparisons and predict the player's success.
- Merged rookie performance metrics from the NBA Stats API with web-scraped strength and agility data and anthropometric measurements from another NBA endpoint to predict rookie performance solely from physical and athletic metrics.
- To achieve this, we developed a custom target variable called ROOKIE_SCORE that consolidates box score production into one number.

ROOKIE_SCORE =
POINTS +
0.4 * FIELD_GOAL_PCT *
POINTS +
0.7 * REBOUNDS + 0.7 *
ASSISTS + STEALS + BLOCKS – TURNOVERS

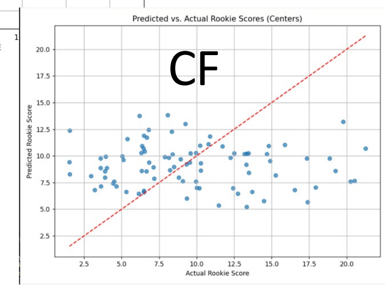
- Use MLP to build a two-layer algorithm using tanh for smooth initial transformations and leaky ReLU to keep gradients flowing, with a linear output predicting a rookie score.
- In CF there is a combine score which weighted the player attributed then filtered for only "Centers" who played at least 20 games then takes the target players to find similar players using either cosine or L2 based on k=5 nearest neighbors

Findings

MLP: The best-performing MLP model achieved a test MSE of 10.7493, significantly outperforming the baseline MSE of 21.7912, where predictions were simply the mean ROOKIE_SCORE across the training set. This model also achieved an R^2 score of 0.5066, meaning it explained just over half of the variance in rookie outcomes.



CF: Our CF Model performed poorer than expected with a MSE of 28.468. This suggests that there is a weak correlation between combine stats and rookie output. This may indicate our model might be overfitting or poorly specified which we tried different variations of constraints to yield better results but all were the same.



Conclusion And Next Steps

Conclusion:

- A simple neural network uncovered modest relationships between physical combine metrics and rookie-year performance, achieving a meaningful improvement over a baseline.
- Despite promising results for scouting and roster planning, limitations like selection bias and reliance on GAMES_PLAYED highlight challenges in pure pre-draft prediction.

Future Work:

- Explore modeling playing time separately and integrating additional data sources (e.g., college or international stats) to mitigate selection bias.
- Enhance the model's expressiveness with a richer set of combined metrics and unsupervised clustering techniques to identify player archetypes.