

High School to NCAA: Predicting Freshman Impact in College Basketball with Machine Learning

Author: Timothy Held (theld1@vols.utk.edu) Faculty Advisor: Adam Spannbaauer (aspannba@utk.edu)

Company: Tennessee Men's Basketball (garrett.medenwald@utk.edu)

Background

- As of April 2025, judicial rulings have removed limitations and guidance related to name, image, and likeness (NIL) usage and compensation, transfer student eligibility, and school-athlete revenue sharing.
- Now, more than ever, athletic departments, teams, and coaches need to focus on financial roster management, transparent recruiting, and relationship building.
- NBA teams use free agent and draft models to make informed roster and budget management decisions.
- Research Objective:** Build a predictive tool to inform coaches on potential impact of high school players at high major schools (SEC-level) in their freshman year.

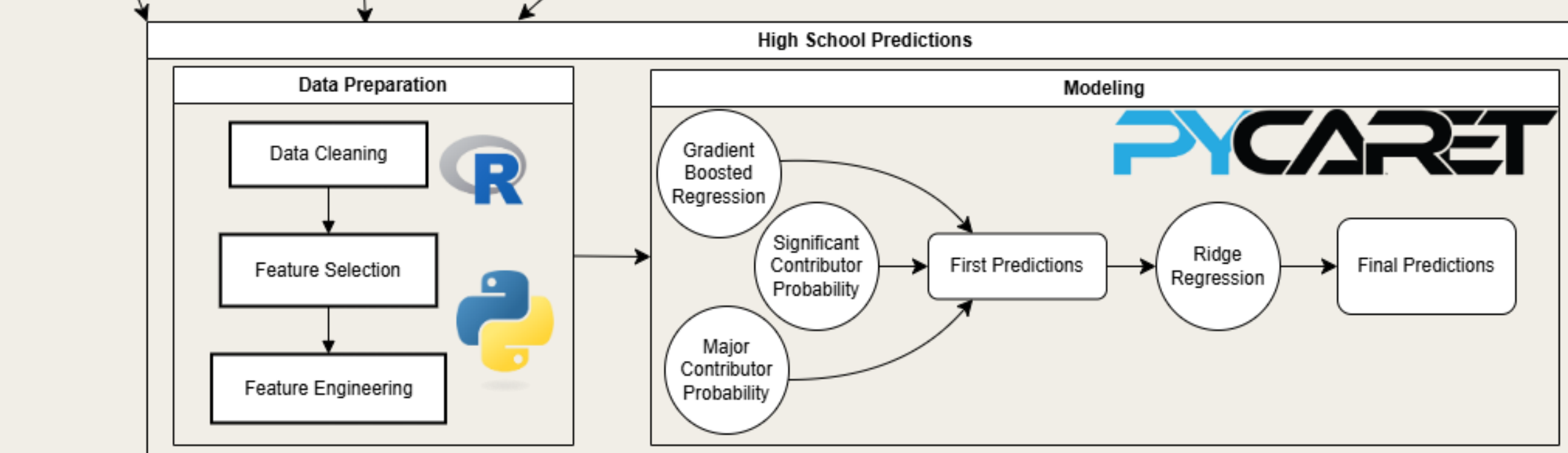
Data:

- We partnered with **Synergy Sports**, a division of Sportsradar, to access their API to collect game and player data for the model
- We compiled 4 years of game-level statistics from high school summer leagues (AAU) and college basketball (NCAA)
- The data is composed of statistics in 5 main categories: Player aggregated box scores, team aggregated box scores, opponent aggregated box scores, league aggregated box scores, and player aggregated advanced statistics.
- The data totaled over 150 statistics for over 350 individual college freshman seasons. For modeling, this was reduced to near 70 variables.

Methodology:

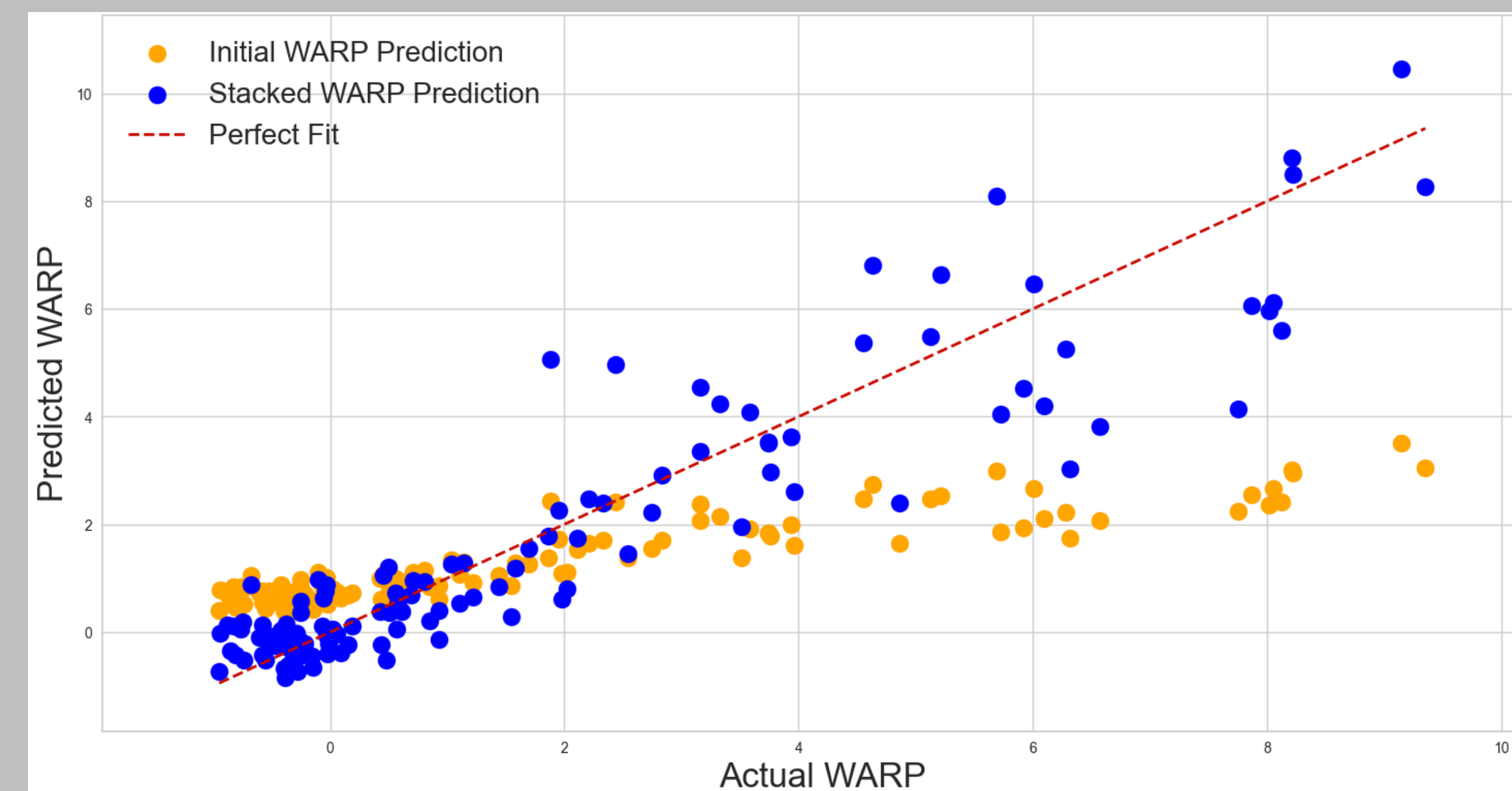
- Calculated Wins Above Replacement Player (WARP)** for each season, high school and college. WARP is a comprehensive measure of a player's impact related to a "replacement-level player".
- Using the AutoML package **PyCaret**, we constructed the below ensemble model, training on past data and testing model accuracy on current freshman.

Predictive Model Name:	Predictive Model Type
<u>Preliminary High-Major WARP Model</u>	Gradient Boosted Regression (GBR): • 71 statistical inputs from high school • Outputs logged value of freshman year WARP
<u>Significant Contributor Model</u>	K-Nearest Neighbors (KNN): • 71 statistical inputs from high school • Outputs probability of an incoming freshman playing 450 minutes (~40% of max possible minutes played)
<u>Major Contributor Model</u>	K-Nearest Neighbors (KNN): • 71 statistical inputs from high school • Outputs probability of an incoming freshman playing 600 minutes (50% of max possible minutes played)
<u>Stacked High-Major WARP Model</u>	Ridge Regression: • Stacked ensemble model using the 3 outputs from the above models as inputs • Outputs a final prediction value of freshman year WARP



Why Use a Stacked Model?:

First model underpredicts consistently on past data while the final stacked model better reflects reality.

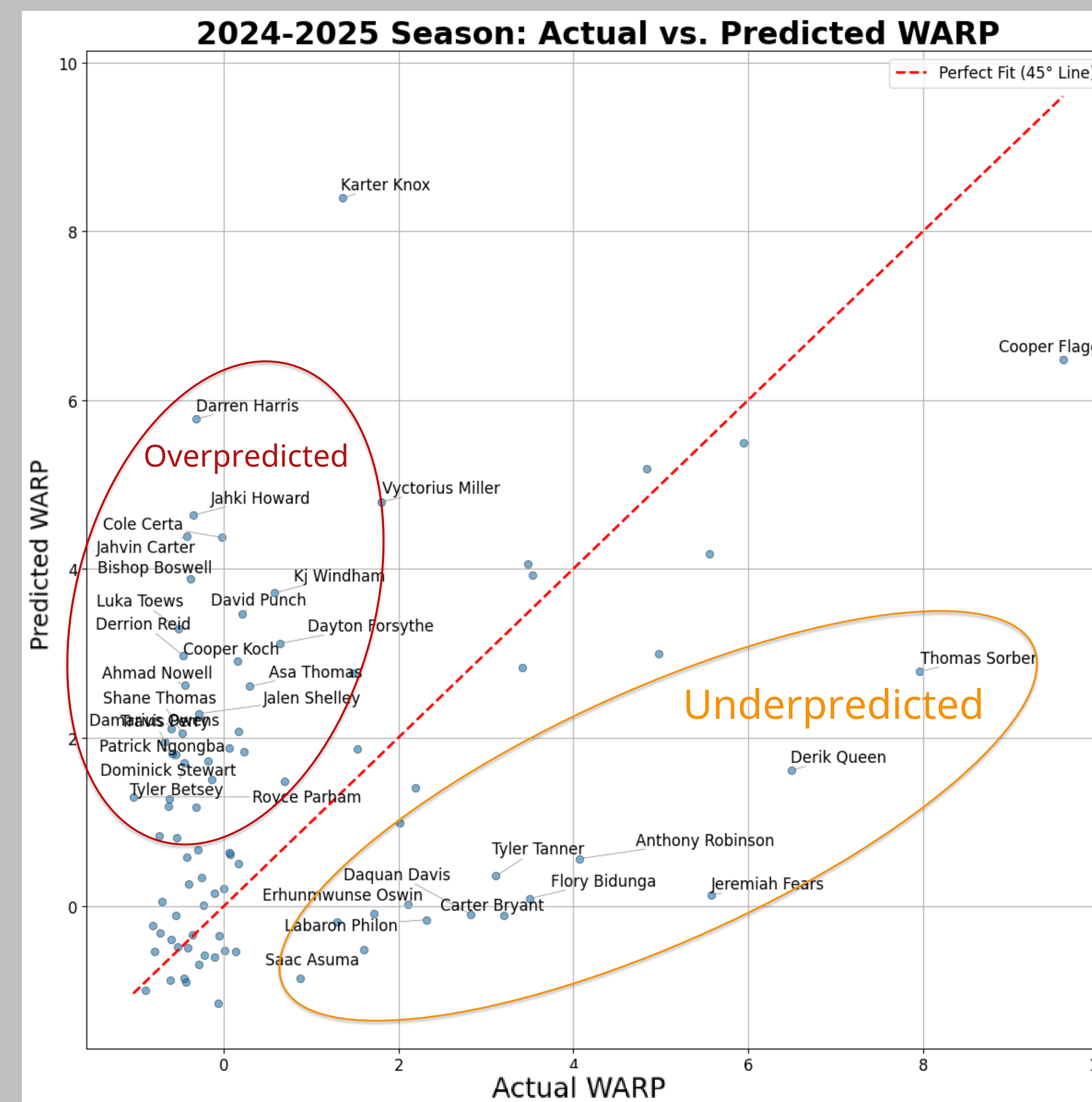


Results:

The stacked high-major WARP model projects a freshman's WARP at an average high-major program (ACC, Big 12, Big East, Big Ten, or SEC) using 17U AAU stats.

For 2024-2025 freshmen, the Mean Absolute Error is 1.83 WARP

We're projecting performance on a hypothetical average team. Thus, we tend to overpredict WARP for those on stronger teams and underpredict those on weaker ones.



Overpredicted Players Averages:

- KenPom* team rank of 40
- 25 games played
- 11 minutes per game

Underpredicted Players Averages:

- KenPom* team rank of 56
- 32 games played
- 24 minutes per game

*KenPom is an advanced analytics system that evaluates NCAA men's basketball teams based on efficiency metrics

Player Prediction Examples:

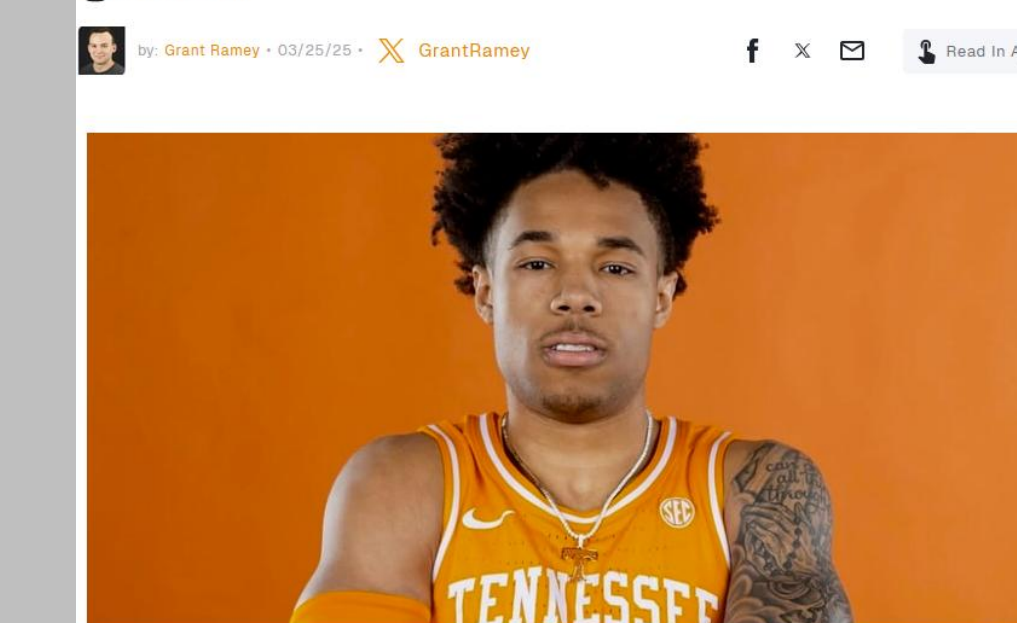
Player	Team	Category	'24-'25 WARP	Predicted WARP	'24-'25 Rank	Predicted Rank
Dylan Harper	Rutgers	Close to Prediction Below Average Team	5.95	5.49	#4	#4
Taheed Pettiford	Auburn	Close to Prediction Above Average Team	3.53	3.92	#10	#12
Cooper Flagg	Duke	Underpredicted Generational Talent	9.6	6.48	#1	#2
Bishop Boswell	Tennessee	Overpredicted Above Average Team	-0.38	3.89	#61	#13

This model shows solid insight into a freshman's potential impact at the high-major level but overlooks context—like Tennessee's four senior guards limiting Bishop Boswell's opportunities.

Implementation:

Player predictions and information, plus cost and transfer models by Noah Cooper (MSBA) and Scott Daughtry (MBA-MSBA), power roster, search, and scouting apps built in R Shiny, a web framework package.

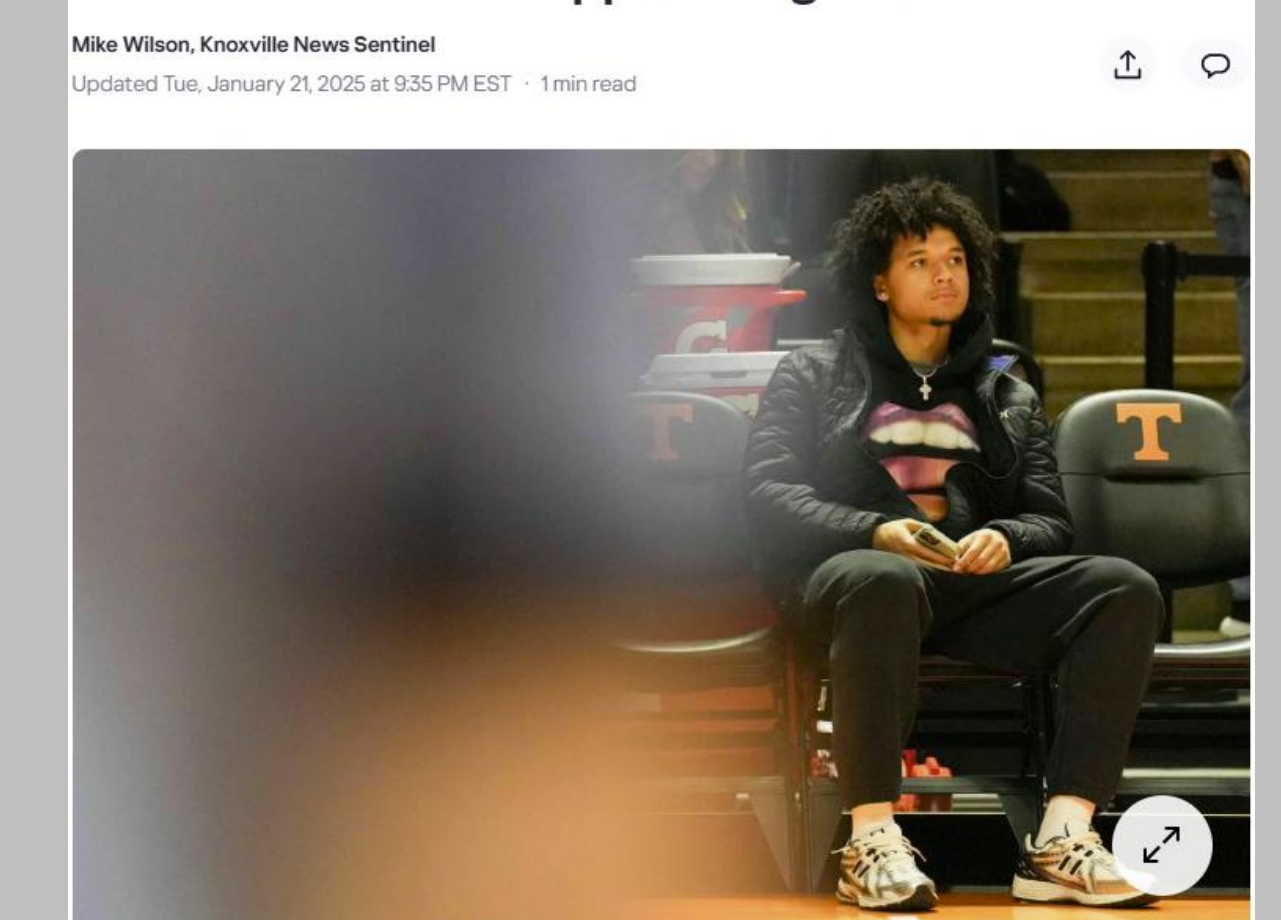
Tennessee Basketball adds commitment from Virginia point guard



Troy Henderson:

- 0-star recruit
- Predicted freshman WARP of ~3 at an average Power 5 team, ranking 44th amongst 2024 AAU players.

Five-star Brayden Burries visits Tennessee basketball for Mississippi State game



Brayden Burries:

- 5-Star recruit, ranked the 11th best recruit in the country by 247Sports
- Predicted freshman WARP of ~6 at an average Power 5 team, ranking 4th amongst 2024 AAU players.

Future Research:

- Build an **Expected Minutes Played (xMP)** model that considers team age, quality and potential positional needs to better predict how a player would perform on a certain team.
- Construct a new model that considers predicted WARP, xMP, and 247Sports Recruiting Rankings to predict Freshman Year Impact.
- Create a NIL model to predict the monetary cost of recruiting and signing players.
- Build an application that suggests high school and transfer players based on position, cost, and other parameters for coaches to utilize