NHL Player Evaluation

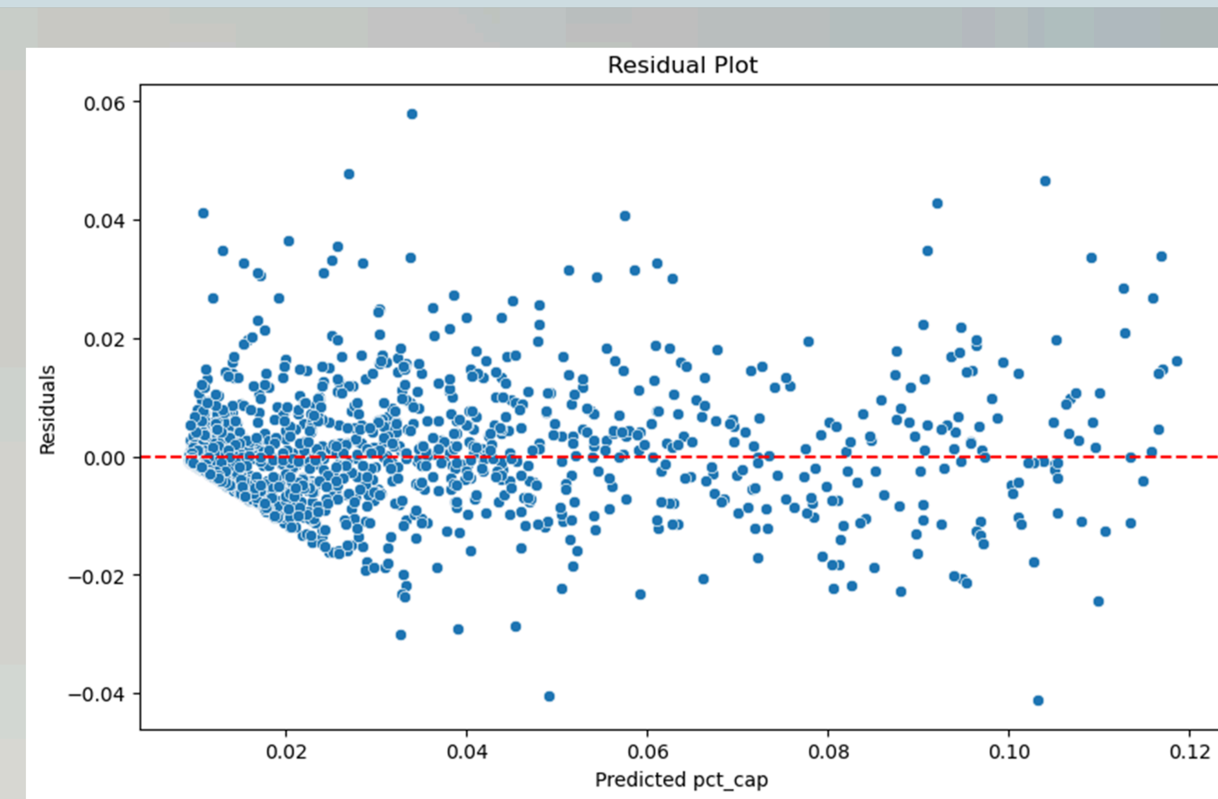
Trevor Hogland | Information Analysis Undergraduate
Dylan Pham | Computer Science Engineering Undergraduate

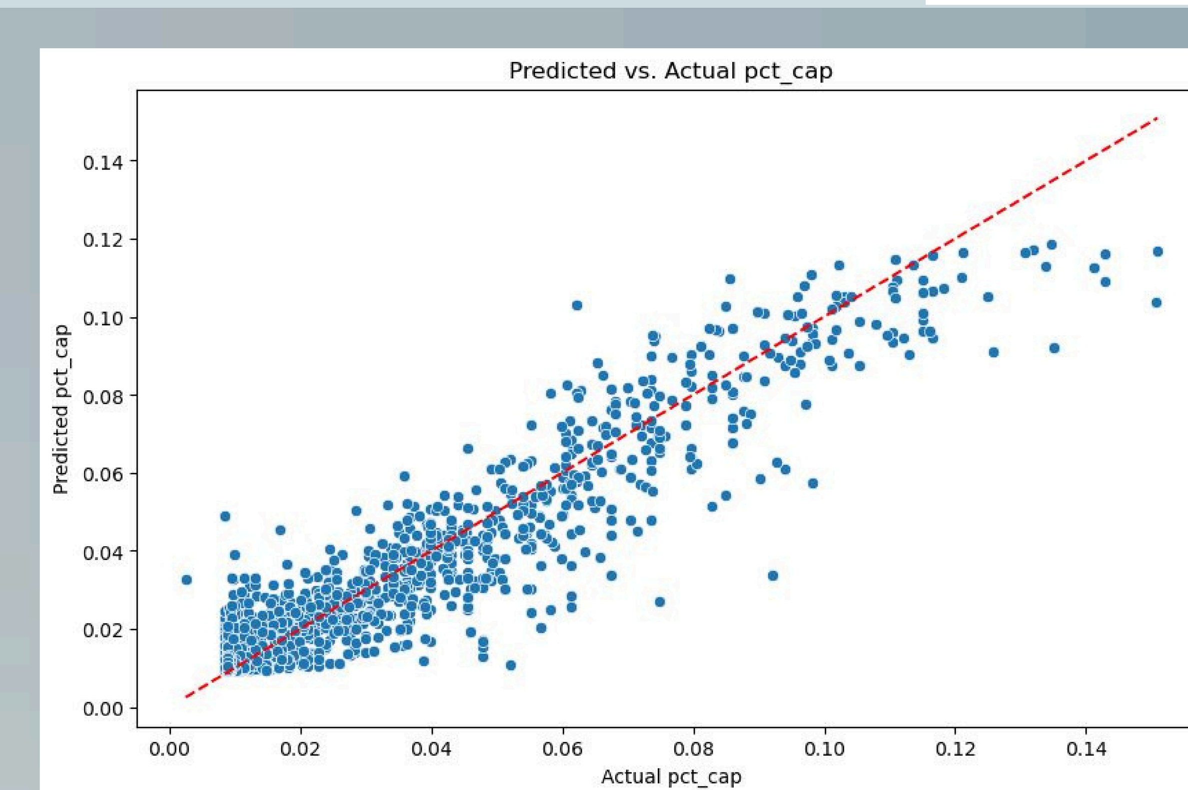
Introduction

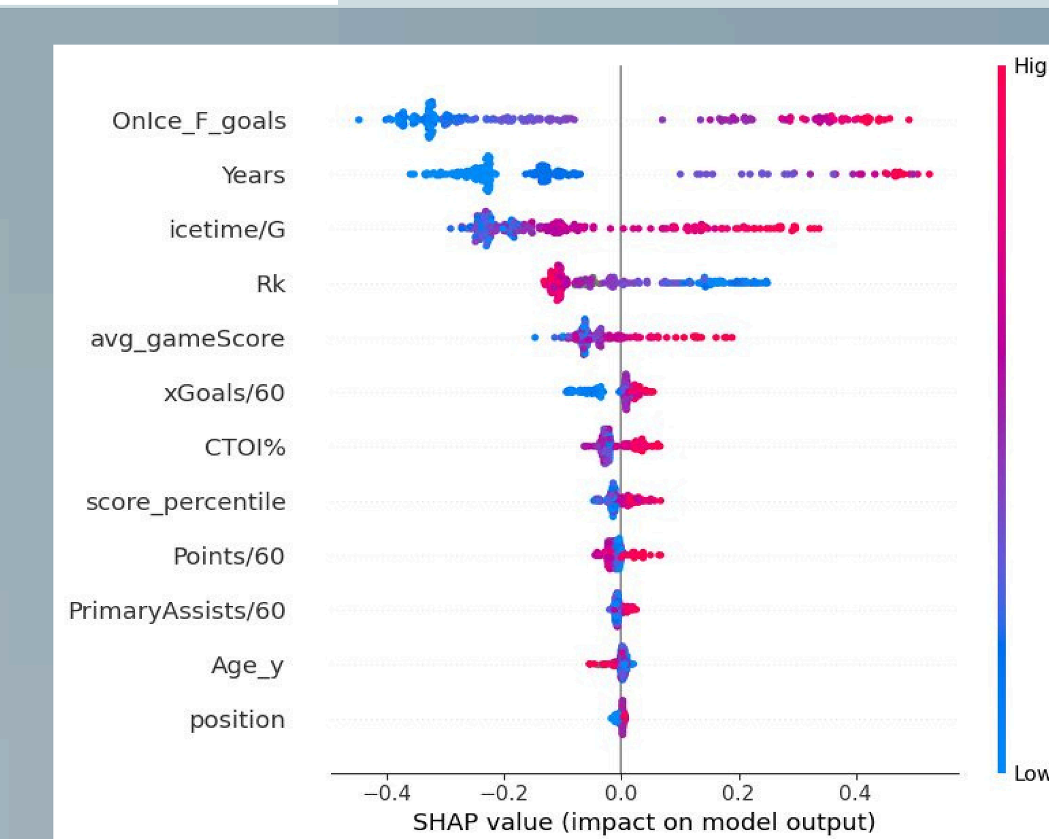
Being a general manager (GM) for any professional sports team is an incredibly challenging job. Predicting how good a player will be and deciding how much they should be paid are difficult decisions historically made through a combination of intuition and experience. The outcomes of these choices can etch a GM's mark in sports history or leave them jobless.

Fortunately, with the use of linear regression models, data scraping, and machine learning, GMs now have access to tools to make more stat-backed decisions.

This semester, we developed a machine learning model that estimates the monetary value of a player based on a variety of performance metrics using SciKit Learn, PYMCDM Topsis, BeautifulSoup, XGBoost, and other Python libraries.



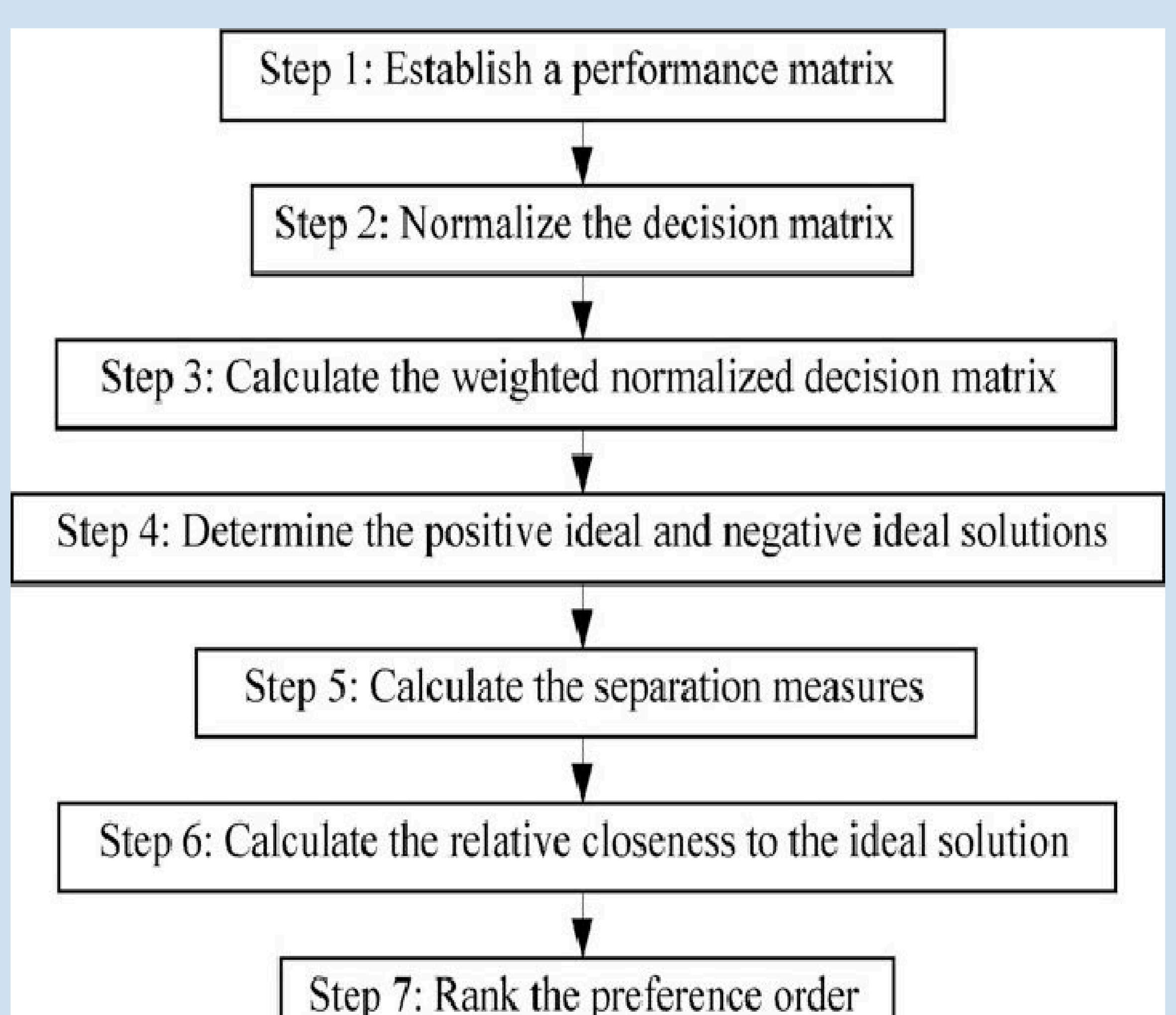




Defensive Metric

In hockey, offense can be easily quantified by points, goals, and assists. Defense is much more difficult to quantify. Determining defensive player productivity is challenging because there are no statistics that capture the number of poke checks, stick lifts, goal crease clear-outs, and other defensive contributions.

We endeavored to capture defensive productivity utilizing statistics such as high danger, medium danger, and low danger expected goals against. We created our defensive metric using the TOPSIS method (see image below) which grades players based on their distance from the ideal. The results were visualized on a basic player card (see Adam Fantilli card).



Salary Methodology

Once our defensive metric was in place, we created a model to predict the monetary value (salary) for each player. We first gathered data on every player who became a free agent or received a contract extension from the 2018/2019 season through the 2023/2024 season. We trained the model using a variety of advanced statistics, our defensive metric, and the player's extension salary.

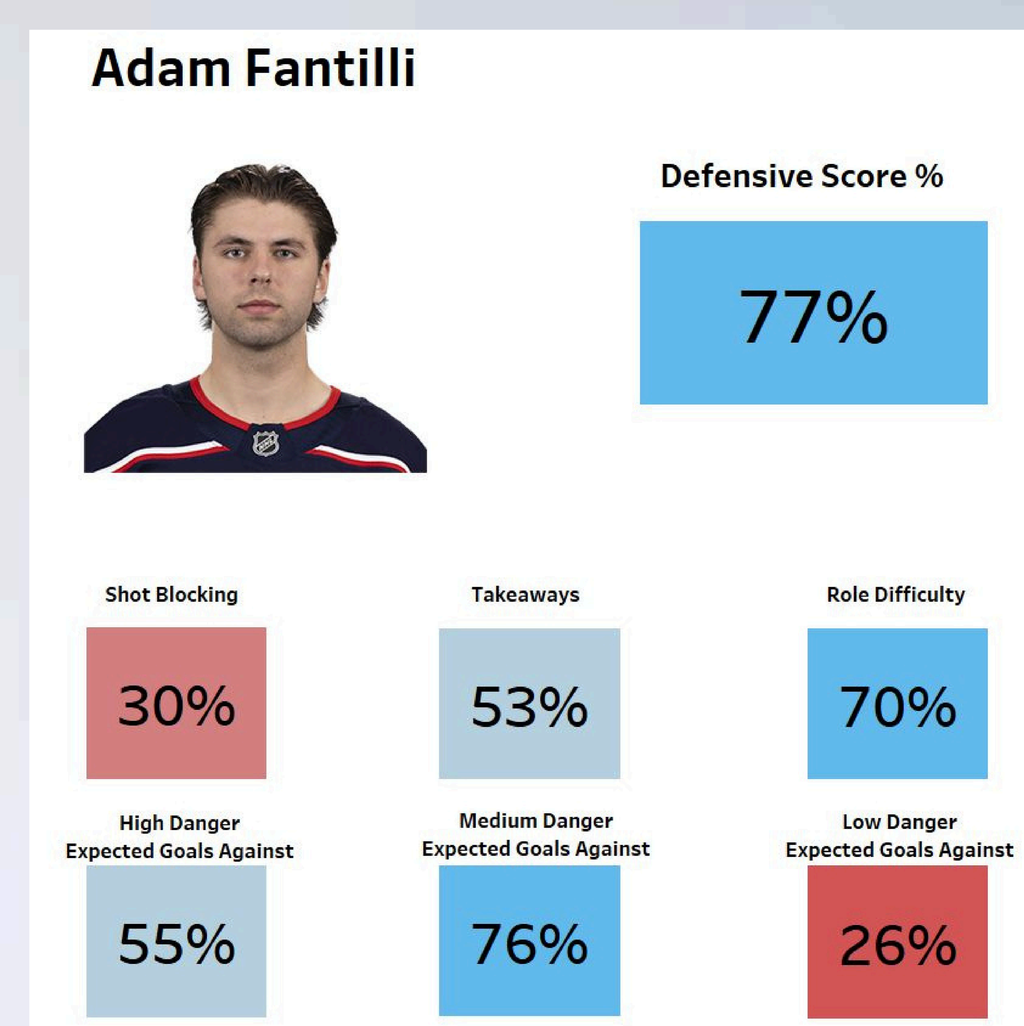
We used an XGBoost decision tree model that analyzed our inputs to project a salary for every player. By projecting each salary in terms of percentage of the salary cap, we were able to account for increasing caps and inflation. These projections were considered accurate if they fell within a 1% range from the actual salary. Based on this metric, our model achieved an accuracy of just over 83% (for more detail, see the graphs at the top).

By putting this data into Tableau, we were able to create an automated player card which displays the projected salary as well as offensive and defensive ratings for each player. See the final product below:

Why This Matters

With our salary prediction model, we can give an approximation of a player's monetary value within seconds, given their statistics from a certain season. Our efforts to create quality player cards will allow anyone to see the monetary value of a player and quickly determine if a player might be overvalued or undervalued. We hope that our model can help fans better understand the value of players.

Helping fans evaluate and understand roster changes is critical in all sports. Fans often react to roster changes and evaluate GM performance based on emotion or poor information. Creating player cards such as the defense card and overall evaluation card allows fans to see what makes a player valuable and how much that player is really worth. Our model has the potential to bring fans inside the mind of the front office and introduce a widespread understanding of both why and how players are valued.



Next Steps

The next steps for this project are fine-tuning our model and preparing for free agency this summer. We would love to have a report for each free agent this year. This would allow us to evaluate free agency as it occurs and offer our own opinions on signings from an analytical perspective.

In terms of model improvements, we hope to increase the amount of data we are using by including more seasons.

While our current model does achieve an acceptable level of accuracy, we only have ~2100 data points and believe it is possible (although time-consuming) to double that. This should improve the overall accuracy of our model. It would be particularly impactful for players who have higher salaries since there is currently less training data available for those cases. Currently, our model achieves an accuracy within our margin for error of ~50% on these players.

Acknowledgements

References

https://moneypuck.com/
https://www.capfriendly.com/browse/free-agents
https://www.spotrac.com/
https://www.quanthockey.com/