Rotman

Master of Management Analytics

Datathon 2022

Assemble a new hockey team for the NWHL

June 17, 2022 Prepared by Team Lavender
(Ojaswita Kalra *, Manas Joshi *, Elaheh Dehghan*, Jad Krisht*)
*** shows equal contribution**

Rotman School of Management
UNIVERSITY OF TORONTO

Rotman

Problem Statement

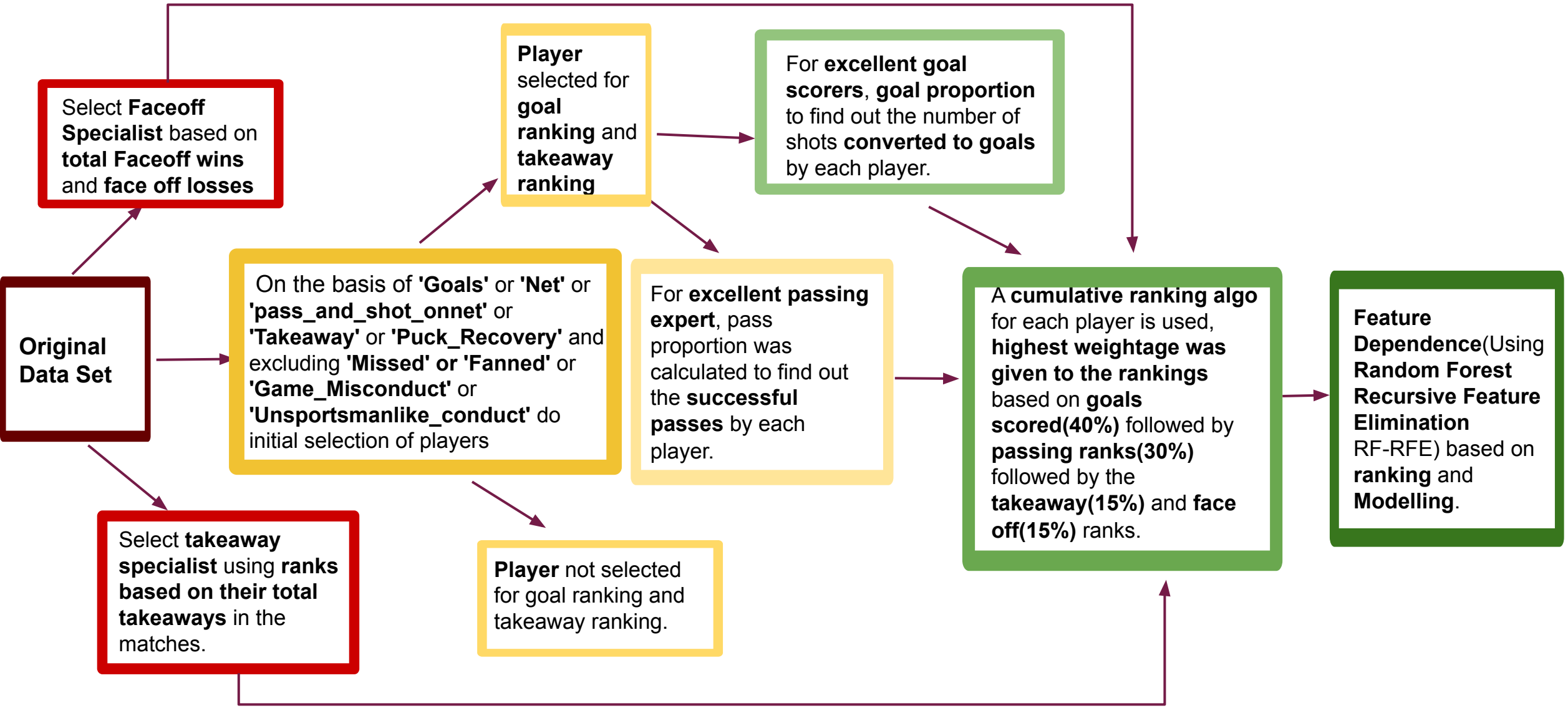
Building a proficient hockey team of five members from analysing all players' performance in all year round games. Players are to be chosen from the NWHL teams. Three excellent goal scorers and at least two should be excellent passers. In addition, at least two faceoff specialists and one takeaway specialist need to be selected.

Dataset Characteristics

Dataset used comprises of events by NWHL teams from Jan 23,2021 to Feb 1,2021 .There are 126 players and 6 teams in total from which 5 need to be selected. There were total 76 goals scored.

Here's where it changes.

Methodology



Rotman

Top Five Players

Rotman

Amy Curlew Faceoff and Pass specialist	Meaghan Pezon Takeaway and Pass specialist	Nina Rodgers Goal and Takeaway specialist	Mckenna Brand Play, Faceoff, Takeaway and Goal specialist	Taylor Woods Excellent Passer, Takeaway and Goal Specialist
				
Goal	Goal	**Goal**	**Goal**	**Goal**
Passes	**Passes**	Passes	**Passes**	**Passes**
Faceoff Win	Faceoff Win	Faceoff Win	**Faceoff Win**	Faceoff Win
Takeaway	**Takeaway**	**Takeaway**	**Takeaway**	**Takeaway**
Toronto Six	**Minnesota Whitecaps**	**Minnesota Whitecaps**	**Boston Pride**	**Toronto Six**

**Highlighted cells represent good track record of completing passes to goal scorers.

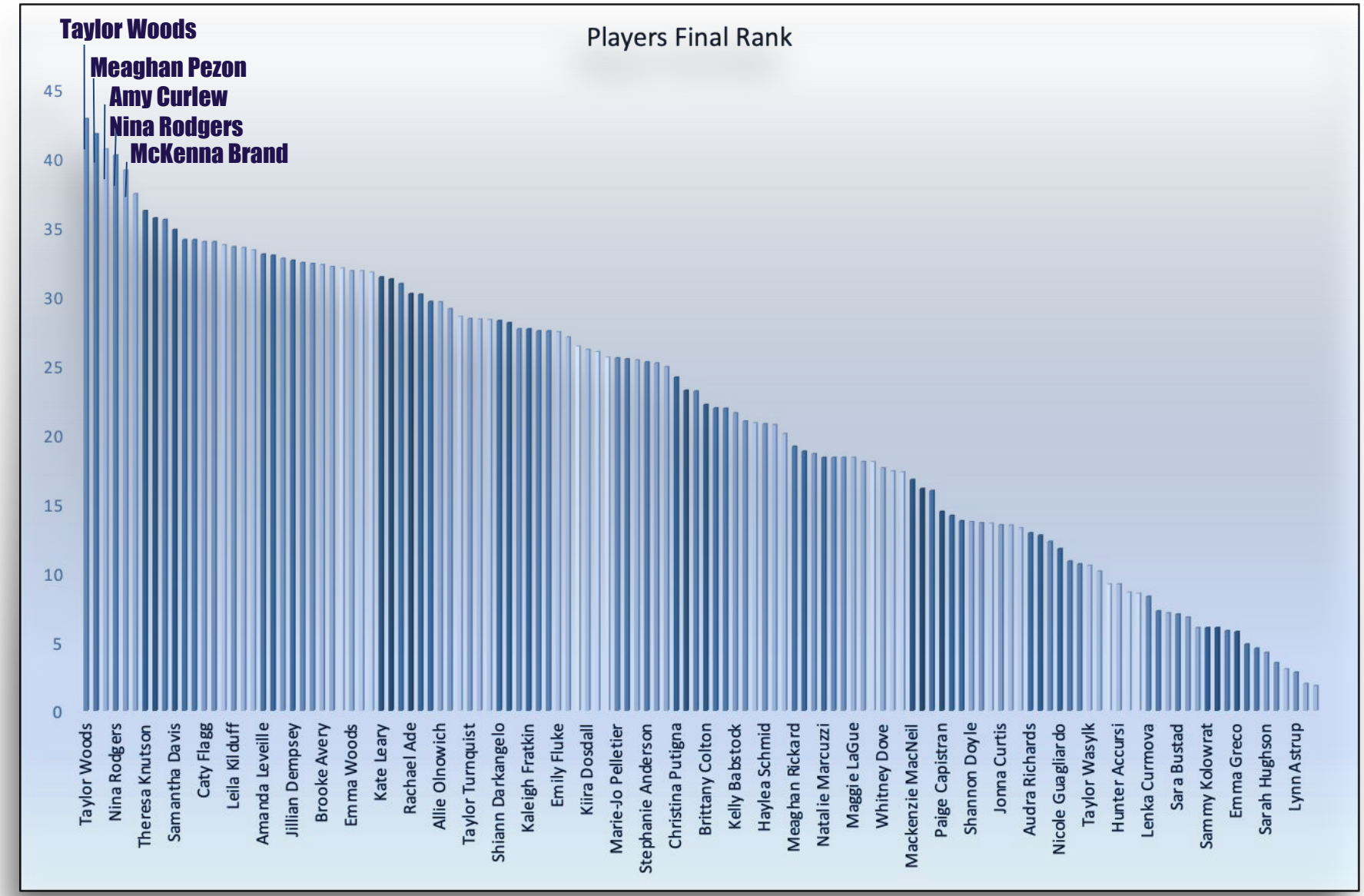
Final Ranking

Top 5 Players

Taylor Woods	42.8
Meaghan Pezon	41.7
Amy Curlew	40.6
Nina Rodgers	40.2
McKenna Brand	39.1

A **cumulative ranking algorithm** for each player is used in which **highest weightage was given to the rankings** based on:

- **Goals scored : 40%**
- **Passing ranks : 30%**
- **Takeaway : 15%**
- **Face off : 15%**

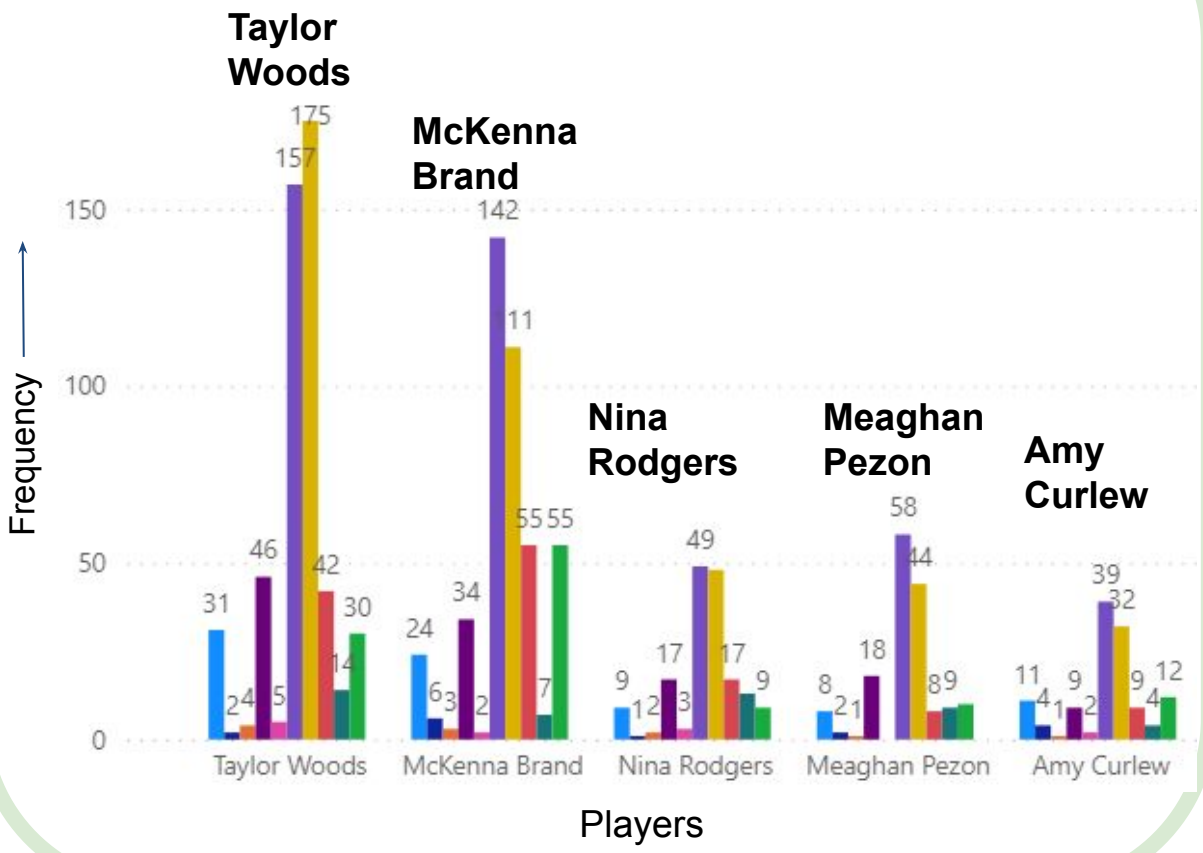


Rotman

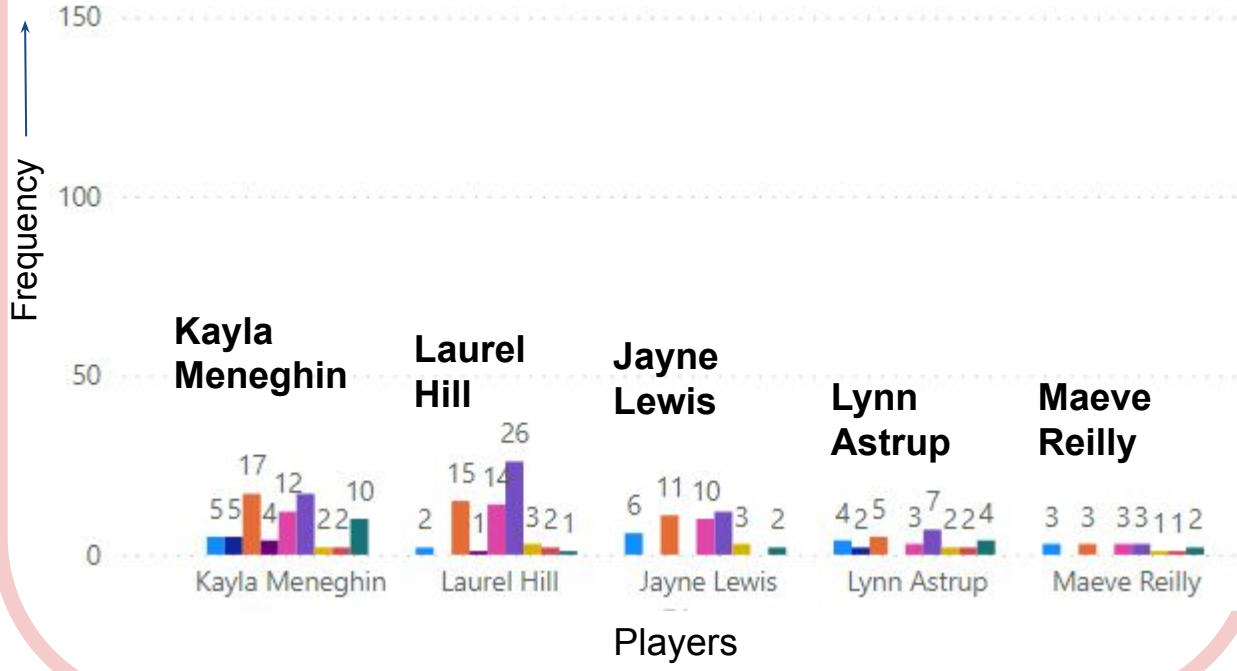
Best vs. Worst Players (Ranking Higher the better)

Comparison of the greatest and worst players in terms of event performance

Highest rank players

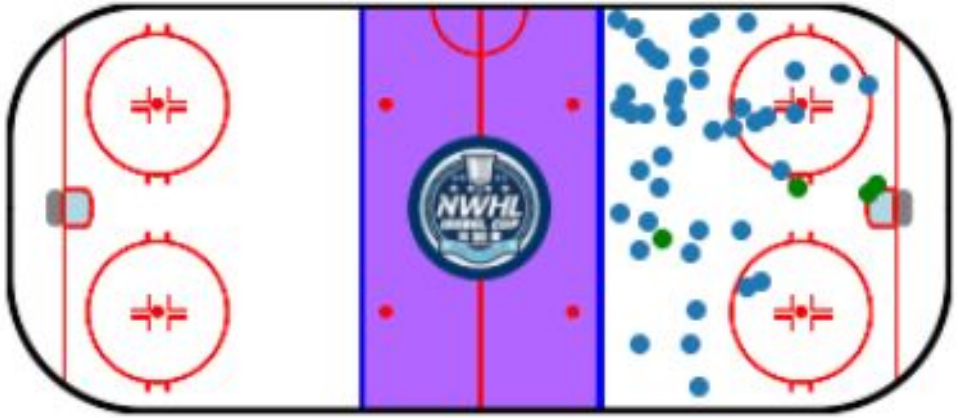


Lowest Rank players

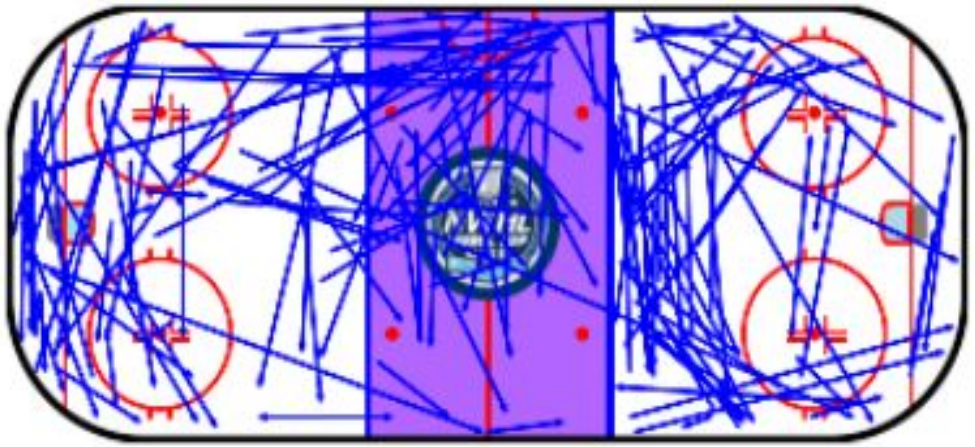


Highest Rank vs Lowest Rank player performance comparison

Taylor Woods (Selected)



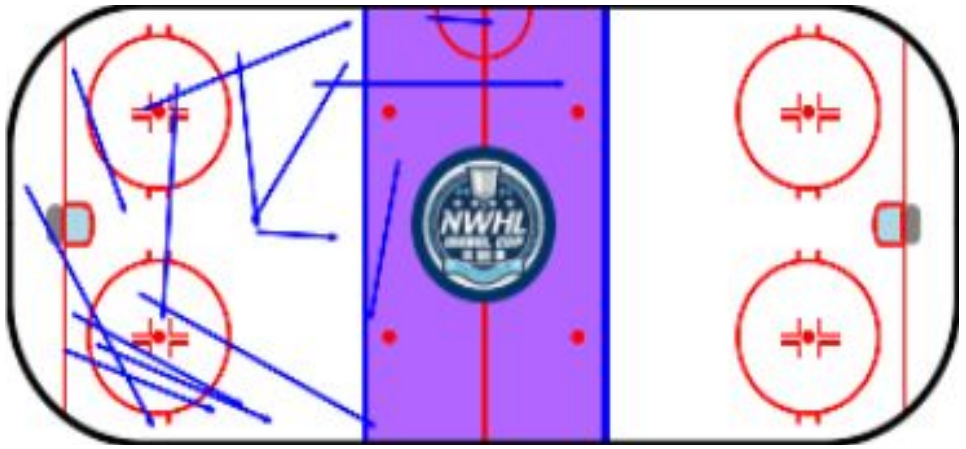
Plays, shots (blue dots) and goals (green dots) of Taylor Woods (Selected) on the rink



1. Here we do a comparison between the two player Taylor woods (Selected) and Laurel Hills(Not Selected) based on our ranking algorithm.
2. We can see that the *number of goals* as well as the number of shots made by Taylor woods is significantly higher than Laurel hills.
3. The reach of Taylor woods on field in the upper and lower half is significantly higher than Laurel hills.
4. The ranking algorithm was able to recommend a very good player for selection.

Laurel Hills (Not selected)

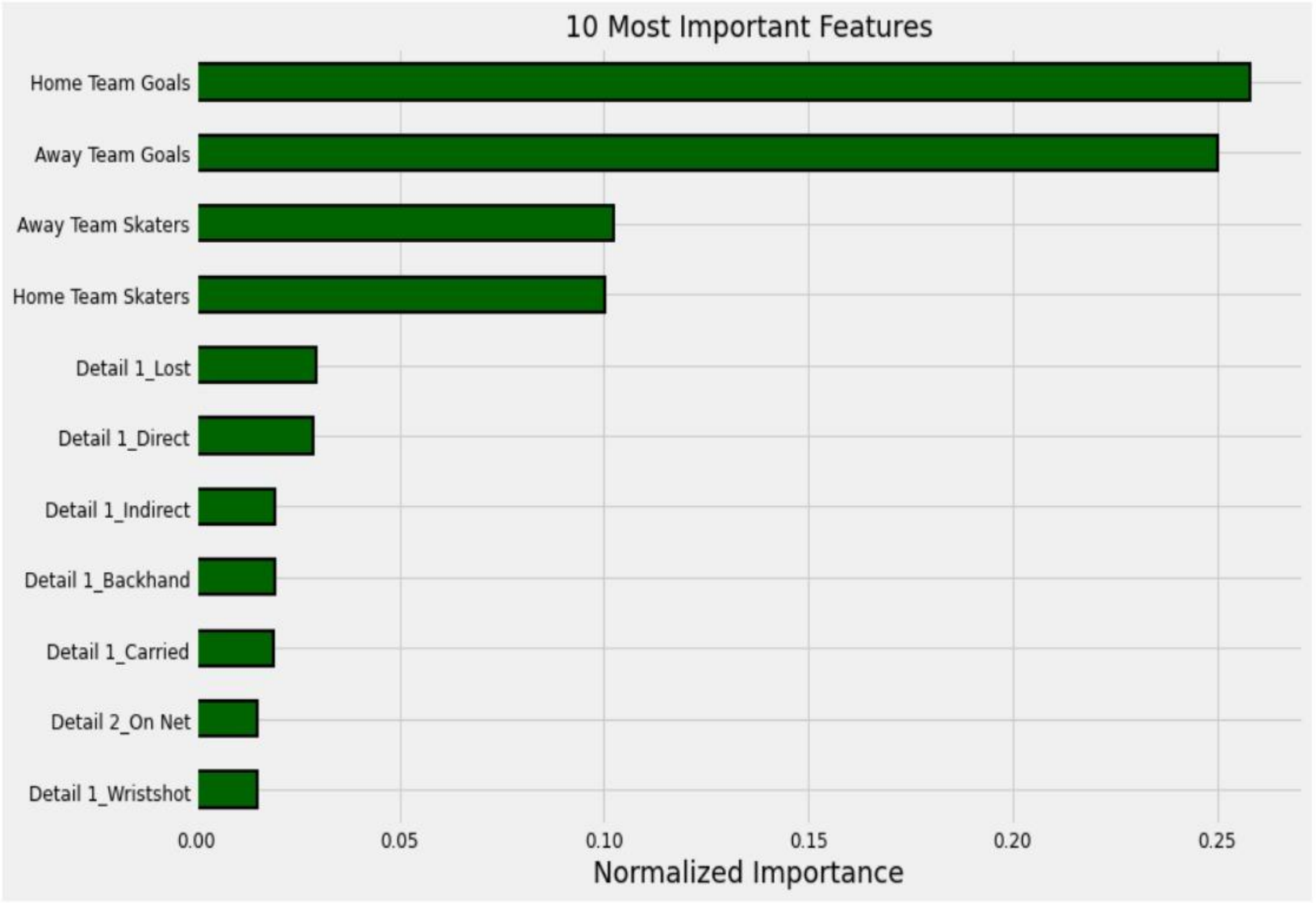
Plays, shots (blue dots) and goals (green dots) of Laurel Hill (Not selected) on the rink



Rotman

Feature Importances

Top ten most important features that were used for selecting a player.



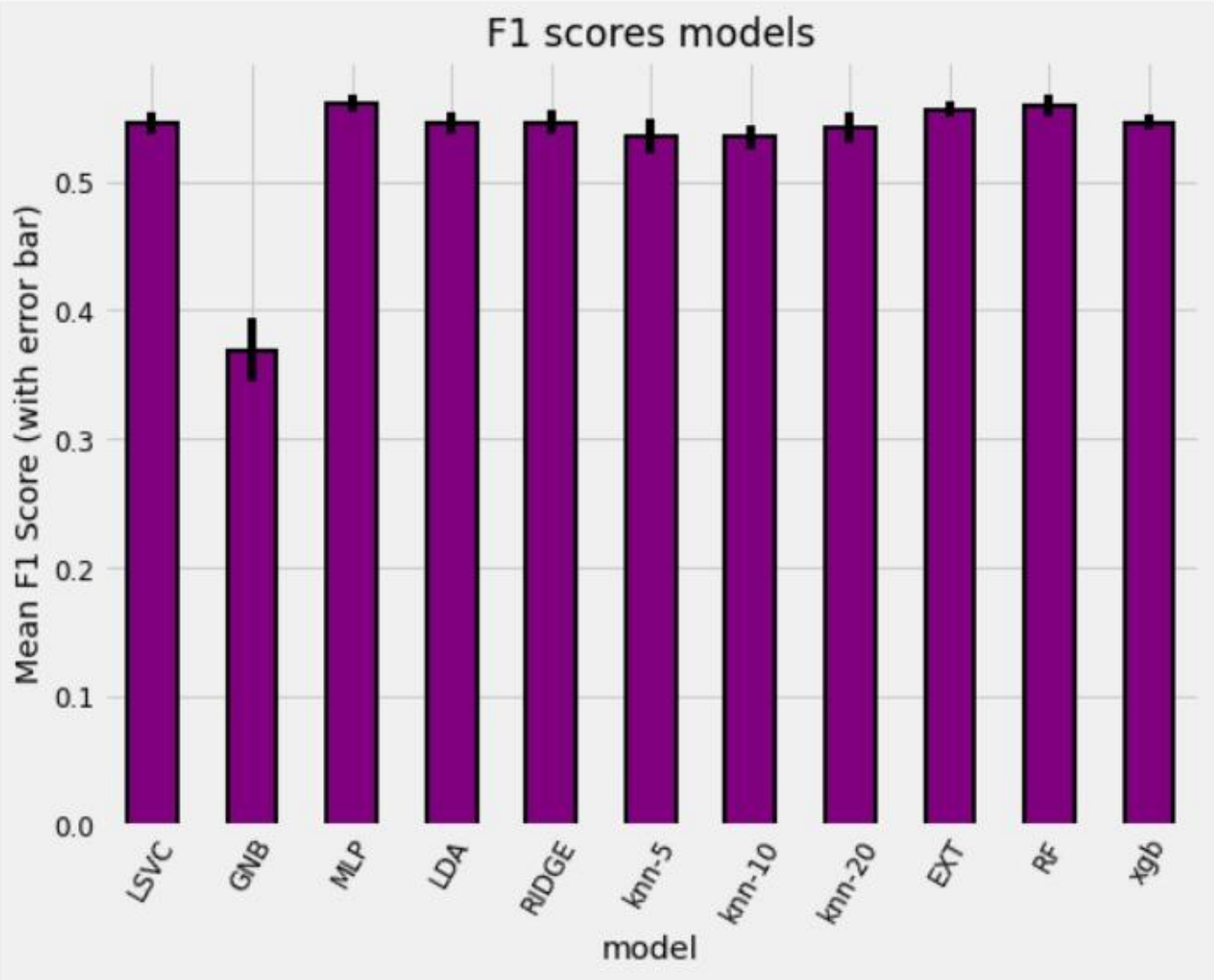
- We used **Random forest recursive feature elimination (RF-RFE)** to **recursively** find out the most **important features for selection** of a particular player.

- Using **cumulative feature importance** we were able to find out that these **ten features required for 95% of cumulative importance.**

Rotman

ML Modelling and its business use-case

The ML models were trained on binary labels target label 0/1 on basis of rank>threshold_rank.



The ML models were trained so that:

1. If we are dealing with an even larger dataset for example dataset from **multiple NWHL seasons** we can use these models as **initial sort** to select players on whom we can apply further analytics and **select top five.**

2. We got **F1 score** of about **56%** and **accuracy** of **67%** on test dataset (unseen dataset), this **can further be improved with DNN**.

3. We set **threshold_rank** for labelling to be **26.0** since this gave us **balanced data distribution**

Rotman

Thank You

1. **F1 score** : The F1-score combines the precision and recall of a classifier into a single metric by taking their harmonic mean. It is primarily used to compare the performance of two classifiers.

2. **Random Forest Recursive Feature Elimination (RF-RFE)** : This is a popular feature selection algorithm. RF-RFE is popular because it is easy to configure and use and because it is effective at selecting those features (columns) in a training dataset that are more or most relevant in predicting the target variable. (https://arxiv.org/abs/1310.5726) (https://machinelearningmastery.com/rfe-feature-selection-in-python/)

3. **Hockey Rink** : A Python library for plotting hockey rinks with Matplotlib.

4. **Github link**: https://github.com/coderXcode/MMADatathon2022

Appendix and Citations