Machine Learning Applications in Projecting Future Performance of NHL Draft Prospects

### Student Name: Alex Church

### Supervisor Name: Dr. Tony White

## Overview

The goal of this project is to use machine learning applications in an attempt to predict the probability of success a draft-eligible NHL prospect will enjoy over the course of their prospective NHL careers, based solely on information available to us prior to, and including, their draft-eligible season.

## Background

Each year during the offseason months the NHL holds an entry draft where teams select the rights to junior-aged prospects. These are players who generally fall between the ages of 18-20, and whose rights are not held by an NHL franchise at the time of the draft. The most recent iteration of the NHL draft was held on June 21-22, 2019, where 217 eligible prospects were collectively selected by the NHL’s 31 franchises. The annual order of selection is predetermined largely based on a team’s performance in the NHL season prior to the draft. Teams are also permitted to exchange future draft selections via trade at any point. Since the introduction of the salary cap in 2005, the NHL Entry Draft has become more and more significant in terms of helping teams gain a competitive advantage. This is largely due to the fact that a player’s first professional contract is capped using a standard scale, meaning that on average players on their entry level contracts (i.e. recent draft selections) often offer franchises better value per dollar than their more tenured counterparts. As players age and their salaries increase, teams can often only hope to remain relevant if they continue to replenish their stocks with younger and more cost effective options, the best way to accomplish this generally being via the Entry Draft.

As analytical thinking has crept its way into the NHL sphere, surprisingly little public work has been done to-date pertaining to the NHL’s Entry Draft. Gabriel Desjardins’ NHLe equivalency model is likely the most widely cited non-proprietary draft related work. Desjardins devised this model in an attempt to normalize the production of draft eligible prospects across the many different junior and pro leagues that feed into the NHL’s talent pool.[[1]](#endnote-1) Following on the heels of Desjardins’ NHLe model, Josh Weissbock developed a model dubbed the Prospect Cohort Success model (PCS), which he used in an attempt to predict the expected future value of any given prospect by generating cohorts of NHL players deemed to be most similar to the prospect in question based on their pre-draft inputs.[[2]](#endnote-2)

## Objective

We would like to build a model that follows in the footsteps of the aforementioned NHLe and PCS models, allowing us to predict the expected future value of a draft-eligible prospect with reasonable accuracy. We hope to accomplish this by calculating the value a prospect can be expected to provide over the course of their NHL career, measured using metrics including, but not limited to, games played, goals, assists, and points. If time permits, we may even wish to take the model one step further by considering additional adjustments to the model’s inputs, such as era adjustments (to adjust for the variation in league-wide scoring levels year over year), along with team-strength based adjustments (to normalize the production of players who play for objectively stronger or weaker teams within the same league). Those types of adjustments may prove to be overly ambitious, however, due to the limited availability of data relating to smaller amateur and junior leagues. Furthermore, and again if time permits, we can also look into extending the model so that we can adjust our expectations for each passing season that proceeds a player’s draft year. This would have additional real-world applications by allowing teams to continue to adjust the expected future value of any given prospect following their draft seasons, as more information becomes available.

## Methodology

We will first need to collect our data, which will be done by calling the API made available by <https://www.eliteprospects.com/>, widely regarded as the hockey world’s most comprehensive publicly available database. Using their API, we can contrsuct a database of each individual player’s career production broken down by age/season, and league. Next, we will need to develop a method for normalizing production across the various junior-aged leagues from which the NHL’s prospect pool is drawn. While Desjardins’ NHLe model did so by comparing the production of players who jumped from feeder leagues directly to the NHL in subsequent seasons, we hope to take that a step further by comparing the year over year production of players as they progress through intermediary junior and amateur leagues as well. Once we have normalized our pre-NHL figures using these newly computed equivalency weights, we can then move toward developing a clustering model that can generate similarity scores for newly draft eligible prospects, based on their historical comparables. These comparables will be computed using pre-draft inputs such as each player’s amateur statistics, along with additional variables such as age and height. In order to generate the clusters, we will use an existing clustering algorithm such as k-means. Finally once we have developed our cohort model, we can then use the NHL career stat lines of the most comparable players in each cluster in order to project the future NHL career trajectory (i.e. the expected future value) for all draft eligible prospects in a given year.

## Timeline

Week 2: background research

Week 4: data collection and cleaning

Week 6: develop feeder league equivalency model

Week 8: test feeder league equivalency model

Week 10: develop clustering model

Week 12: test clustering model

Week 14: develop expected value model using our pre-computed clusters, final testing and refining of full model

Week 16: prepare final report

\*Python will be used as the chosen language of execution, thanks in large part to its extensive catalog of libraries including numpy, pandas, scikit, tensorflow, requests/beautifulsoup, sqlite, etc.\*

1. <http://www.behindthenet.ca/projecting_to_nhl.php> [↑](#endnote-ref-1)
2. <https://canucksarmy.com/2015/05/26/draft-analytics-unveiling-the-prospect-cohort-success-model/> [↑](#endnote-ref-2)