Put the Biscuit in the Basket:

A data mining approach to predicting NHL free agent salaries

SYS 6018 Final Project Proposal

Jack Prominski, Court Haworth, Tyler Lewris

Identifying the Problem

General managers in the National Hockey League are tasked with a difficult task: they must allocate limited resources in a competitive and uncertain environment, all while having their performance and decisions the subject of relentless public scrutiny. The increasingly widespread adoption of advanced analytical techniques in the hockey community has both increased this scrutiny and decreased the opportunities that were available in the less efficient market of the past.

NHL GMs, broadly, are responsible for putting a team on the ice. They are intimately concerned with how they can optimize their spending on player contracts to maximize their team’s success, subject to the constraint of the salary cap. Particularly, they would like to avoid both low-balling contract offers, at the risk of the player going to another team, as well as overpaying players beyond their market value. These decisions are further complicated by salary cap rules that affect negotiating rights as well as random fluctuations in player performance from year to year. If GMs can more accurately predict a player’s value on the open market, they will be in a better position to assemble the best team possible, thus leading to a higher chance of winning more games, and ultimately, the Stanley Cup.

Objectives & Approach

We propose to explore several modeling techniques to accurately predict the market value of an NHL free agent. We will be predicting the AAV (Average Annual Value) of a contract. The dataset we are working with are NHL performance metrics and salaries over the past 4 seasons. Our hypothesis is that we will be able to effectively predict a free agent’s value. And we will measure the accuracy of our models using the Test MSE obtained during K-fold cross validation. The methods for prediction we will use include decision trees (Random Forest/Boosting), multiple linear regression, GAMs, KNN, and some ensemble models combining different techniques.

Understanding the State of the Art

There is an active hockey blogging community, and we have found cases of bloggers using linear regression[[1]](#footnote-1) and KNN[[2]](#footnote-2) to tackle this problem. These models have been somewhat effective, with the regression model delivering an R2 of .76 and the KNN model delivering a standard error of roughly $650,000. Many teams likely have their own models to value players, but these techniques are proprietary and not publically known. This problem is especially difficult for several reasons. One is that the hockey environment changes relatively quickly. Styles of play go in and out of fashion, and thus the market value of players with particular skill sets fluctuate. Supply of players is also not constant. A player in a relatively weak free agent class may command more in salary than he would in another year. Another factor that makes this problem difficult is that GMs sometimes make irrational decisions, which introduces more noise in the data. One consideration is that this model attempts to model what a player gets paid in the possibly irrational free agent market, not an objective value of how much a player is worth.

1. "Predicting Free Agent Salaries - Puck Plus Plus." 28 Jun. 2015, <https://puckplusplus.com/2015/06/28/predicting-free-agent-salaries/>. Accessed 13 Nov. 2017. [↑](#footnote-ref-1)
2. "Hockey And Euclid: Predicting AAV With K-Nearest Neighbours | Corsica." 31 Oct. 2016, <http://www.corsica.hockey/blog/2016/10/31/hockey-and-euclid-predicting-aav-with-k-nearest-neighbours/>. Accessed 13 Nov. 2017. [↑](#footnote-ref-2)