

# Capstone Two Project Proposal: Predicting NHL Player Performance

Matt Feller

## 1 Problem statement

The goal of this project is to train a model to predict each NHL skater's goals, assists, shots-on-goal, and (possibly) a variety of other statistics, based on their statistics from previous seasons and other features such as age, height, and weight.

## 2 Scope

An eventual finished product would be a model predicting all relevant player statistics, for use as a tool in fantasy hockey drafts. However, since there are at least a dozen such statistical features commonly used in fantasy, I plan initially to limit the scope just to predicting goals, assists, and shots on goal. (For the purposes of completing this program in the coming months, I may stick to this more limited scope, but I do intend to develop it into a finished product that I can put to use for my own fantasy hockey draft in October 2024!)

## 3 Dataset

I plan to use this robust dataset which contains highly detailed information about NHL games, teams, and players from 2000 to 2020 as a relational database of CSV files. The amount of data available in this dataset is certainly too much to feed directly into a machine learning algorithm (at least on my machine!) but I hope to explore it to engineer some helpful features that can unlock better predictions than the basic stats do alone.

## 4 Baseline

We get a few naive models by simply taking a weighted average of a player's stats across recent seasons, or even just by predicting that their stats will be identical to the previous season's. The main challenge will be to train a model with significantly higher accuracy than any of these naive models. An additional, more ambitious challenge is to compare the model's accuracy against the projections provided by Yahoo Fantasy Sports and other such websites.