## Problem Definition

In this project I hope to classify National Hockey League shot selection by some level of quality of the shot (i.e. poor, fair, good) by the level of success from position on the ice and player amongst other features. As I continue the analysis, this could be compared to things such as game results. The NHL lags behind most other sports leagues in this type of analysis but there are a few foundations out there that I'd like to expand on. For instance, others have tackled this topic from an interest in gambling or visualzation , but if machine learning is approached all that is mentioned is libraries and packages with little mathematical support.

## Dataset Selection

The NHL provides all necessary data, but does not share it easily. There is a workaround for this developed by the community which pulls the data from the API (which is open and fair-game) in JSON format easily. This is an example for the final game of the season last season that could easily be expanded over a series of game id's and compiled into a season(s) long dataset: https://statsapi.web.nhl.com/api/v1/game/2018030414/feed/live .

## Network Information

I would like to try both Perceptron and Decision Tree networks to compare (or perhaps combine) the results as they both are capable of classifying data given to it into different categories. To test the results, I plan on implementing a confusion matrix style test to see how the classification compares to true positives/negatives with MSE or another statistic that leads to a findable conclusion. I've already listed a few references that can help me learn to tackle this problem, but I will include another: https://hockey-graphs.com/ . This has a breadth of sport-specific statistic studies that could help drive my project in the right direction.

## Rough Project Schedule

Phase 1 (Completed by 8/3)

- Scrape all needed data from internet (API/Selenium/etc…)
- Complete data preprocessing
- Use clean data and pre-analysis to develop numerical goals for objective classifications "good", "poor" , "fair"

Phase 2 (Completed by 8/16)

- Develop and train network
- Make any changes to problem statement that could be necessary to finding new, interesting discoveries
- Re-develop, re-train network
- Test Results

Phase 3 (Completed by 8/20)

- Draw Conclusions and complete report