Capstone Proposal

NHL Player Analysis

Business Understanding

NHL owners need to evaluate player statistics when signing or re-signing new players. Through machine learning we can develop a system to predict player contract prices allowing team owners more insight through contract negotiations. This projects target audience will be NHL management or any person in within a team's organization that have contract negotiation powers. This model will impact a team by giving them the capable resources to identify player contract prices.

Data Understanding

The data will be collected through the NHL's API and scraped from Spotrac's website. Spotrac has extensive statistical information for any professional sport. A few features in the dataset will include player name, position, jersey number, salary, height, & weight.

Data Preparation

The data will be stored and cleaned within a pandas dataframe. Preprocessing will include converting JSON files to dataframes and merging tables based on common columns. I expect the dataframe to be ~950 rows of active players. At this time, I do not have data on inactive players.

Modeling

Since this will be a regression problem the project will be using Linear Regression, Random Forest, SVM, and XGBoost. Our target variable will be the players salary. To start, we will use a Linear Regression model as our baseline and move forward using pipelines to find our best model.

Evaluation

Our evaluation will be based on RMSE. I believe this task can be completed within a week. My stretch goals would be to incorporate a UI allowing end users to filter stats based on player. Also, incorporating some form of time series forecast would a huge stretch goal.

Deployment

Deployment will consist of a user interface that will be filtered based on player.

Tools/Methodologies

We will be using all of the data science tools a few are:

- Sklearn
- Pandas
- XGBoost
- Matplotlib

The analysis will be completed on my local machine storing the data locally and on github.