Predicting NHL Goal Scoring

Capstone Sprint 2

Taylor Gallivan

10 Nov. 2023

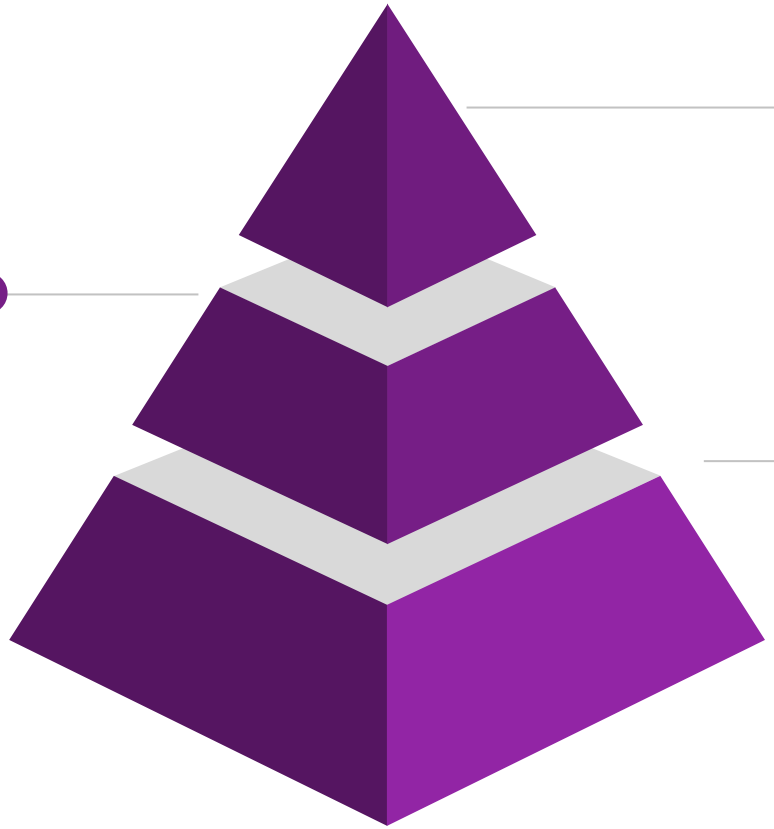
Project Overview

- The sports analytics industry has been growing consistently since the early 2000's → first widely adopted in the MLB, but now a prominent feature in every major North American league (MLB, NBA, NFL, NHL, MLS)
- Have been a hockey fan as long as I can remember: fond memories attending games, cheering on Team Canada, watching Hockey Night in Canada
- With the bevy of NHL data available, merging my hockey fandom with data science was a perfect fit for a capstone project

Problem Question:

Can an ML model (or models) be trained to accurately predict a player's goal output for a season, based on their individual characteristics (a mix of statistical and categorical features)

Progress to Date



Ran initial model

Split and scaled data
Ran linear regression model
Model overfit on training data

R2 Score (test): 88%
R2 Score (train): 87%
MSE: 0.2%

Processed Dataset

Flattened duplicate seasons
Solved for missing statistics
Made franchise names consistent
Encoded categorical data

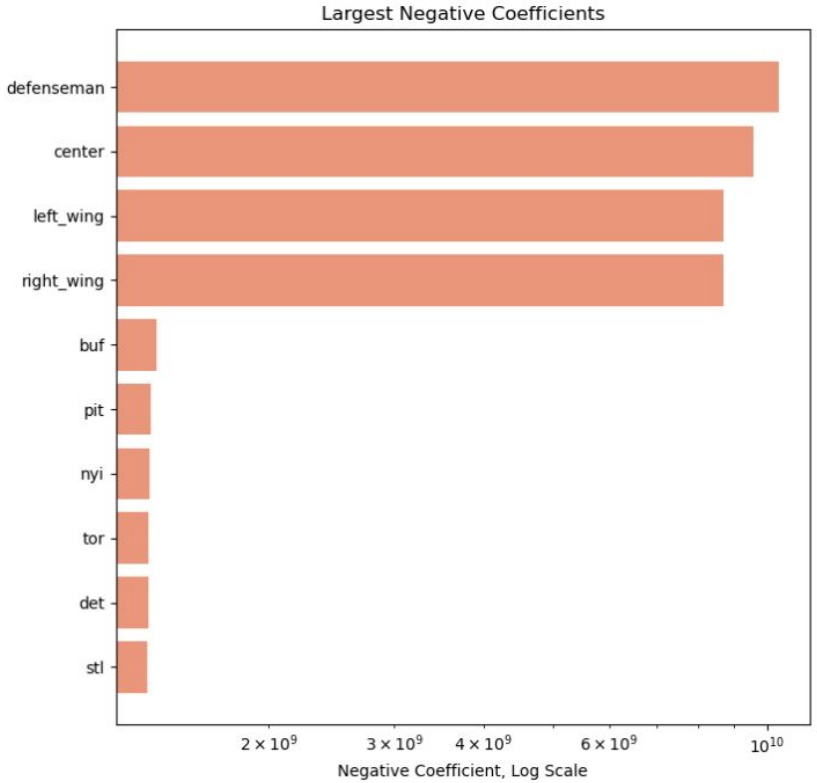
~25,000 rows, 62 columns

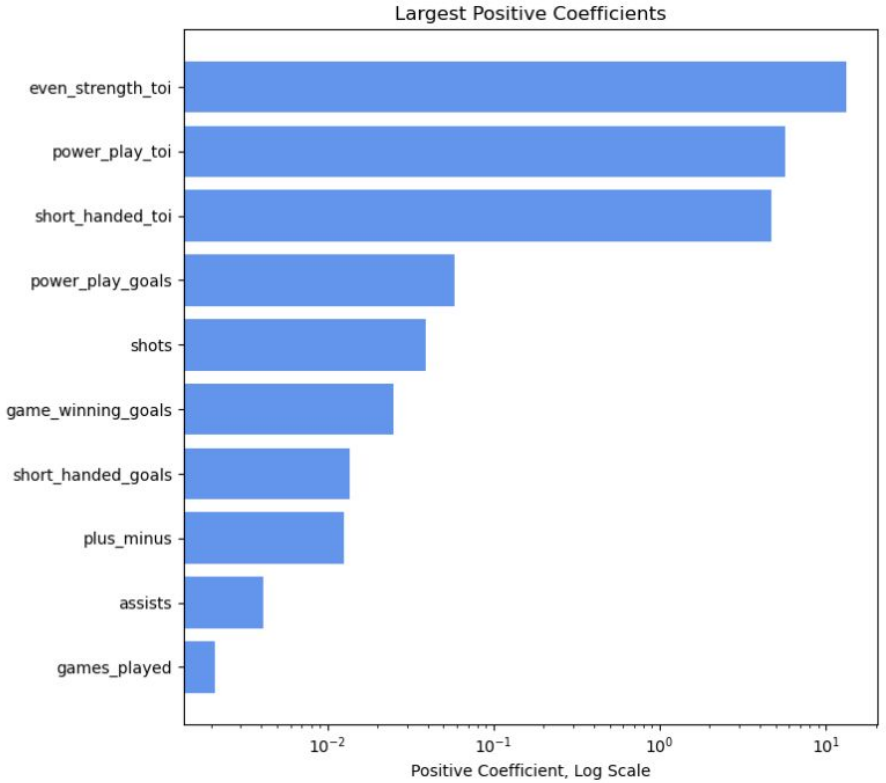
Collected Dataset

Data pulled from NHL API
Data filtered during collection

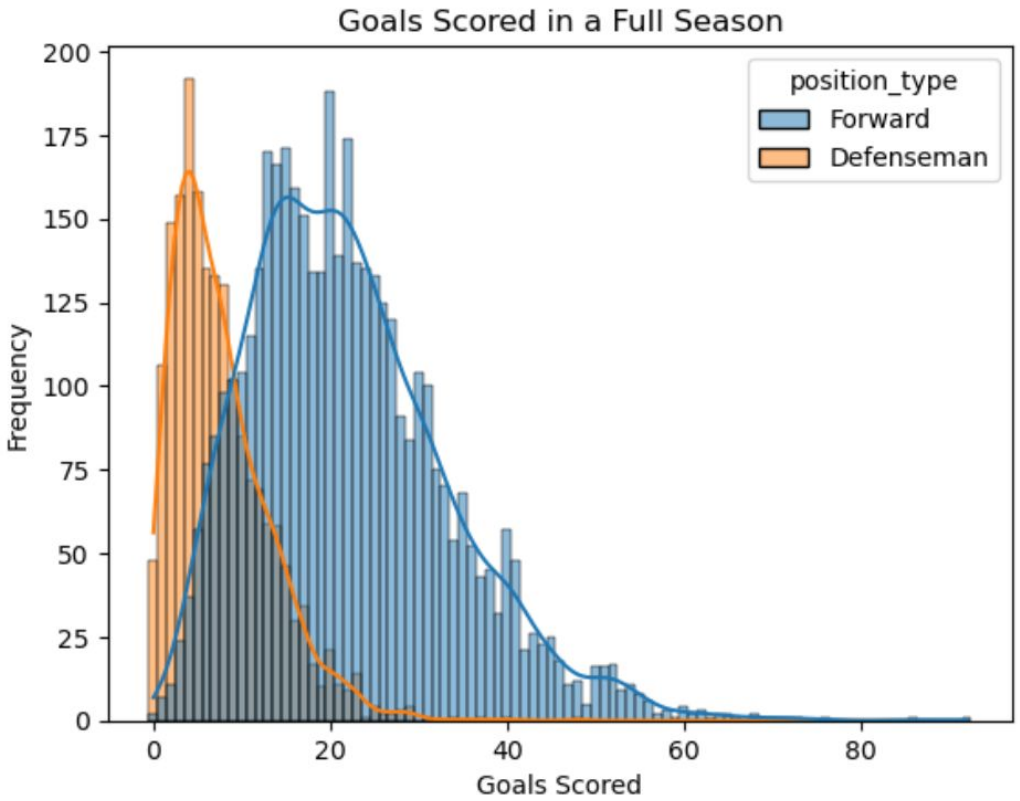
~38,000 rows x 46 columns

Modeling Insights





Outlook



Negatives:

➔ Dataset is smaller than anticipated

Positives:

➔ Features with strong predictive power

➔ Domain knowledge

Concerns:

➔ Time constraint

Next Steps

1. Additional feature engineering
    a. Adjust target feature
    b. Adjust scoring stats for era
    c. Use trailing seasons to calculate:
        i. Percentage change for time-one-ice features
        ii. Weighted averages for scoring stats
2. Evaluate p-values within features matrix
3. Remodel the data using:
    a. Linear Regression model
    b. Gaussian Process Regression (GPR) model
    c. Support Vector Machine (SVM) Regression model