An Expected Goals Machine Learning Model in Hockey

Aashir Naqvi, Felix Jiang,
Ildi Hoxhallari, Lea Luchterhand

January 25, 2025

Hockey analytics has gained considerable attention in recent years as teams and analysts look for ways to gain a competitive edge. One key metric in hockey analytics is "expected goals" (xG), which is a statistic used to measure the quality of a team's scoring chances. It estimates the likelihood that a given shot will result in a goal, based on factors such as shot location, angle, and type of play. Accurate predictions of xG help in assessing player performance, evaluating goaltenders, and informing game strategies. This project will apply machine learning models to predict the likelihood of a shot turning into a goal, based on historical shot data.

Peter Tanner's shot data will serve as the primary data source of this project, containing over 1.8 million shots with over 100 possible features to choose from. This rich dataset provides the necessary features for training and evaluating machine learning models.

We will explore several machine learning models for predicting xG. Logistic regression will act as a baseline model for binary classification. Then, random forests will be used as an ensemble method to capture any nonlinear relationships. Finally, gradient boosting will be used as an advanced technique for further improving prediction accuracy if possible. Model performance will be evaluated using cross-validation, with metrics like accuracy and precision.