xReboundsPlus, Creating a Statistic to Predict Rebound Quality

Evan Lawing

elawing1018@gmail.com
https://medium.com/@elawing1018

Rebounds are one of the most important aspects to any team's success when it comes to hockey. Scoring off of rebounds is becoming increasingly important and starting to play a larger role in hockey every season. In an analysis by Neil Pierre-Louis, he found that the number of rebounds per game increased dramatically from the 2017-18 season to the 2023-24 season, rising from 3.53 rebounds per game to 6.17. The increase in rebounds highlights the fact that teams are using rebounds more than ever in an effort to score goals where they catch the goalie and defense off guard. To show the value of each rebound, a new statistic, xReboundsPlus, will quantify the quality of the rebound of a given shot. This will be calculated in three steps: first, modeling the probability of a rebound (xRebounds); second, modeling the expected shooting angle of the rebound (xAngle); and finally, combining xRebounds and xAngle to determine the goal probability.

The shot data used for these models was collected from moneypuck.com; specifically, the shot data from 2023-24 were used to train these models and applied to the data from the 2024-25 season. The variables that were used for this analysis were if the shot resulted in a goal, if the shot was a rebound, if the shot generated a rebound, if the shot was on a rush, the handedness of the shooter, the shot type, the angle of the shot, and the arena adjusted shot distance. For this analysis, a rebound was defined using the Moneypuck standard: a subsequent shot occurring within three seconds of the initial shot.

The first step in creating the xReboundsPlus statistic was to create an expected rebound model that outputs the likelihood that a given shot will result in a rebound. Extreme gradient boosting (XGBoost) was selected as the model to be used for rebound classification due to its ability to classify well and create the model rather quickly. The dependent variable for the model was whether the shot generated a rebound and the independent variables were the angle of the shot, the shot distance adjusted to the arena, the type of shot, the handedness of the shooter, and if the shot came off the rush.

K-fold cross validation with five folds was used to determine the best XGBoost model and the parameters were tuned to prevent overfitting. These tuned parameters included adjusting for a class imbalance, changing the maximum depth, and adjusting the learning rate of the model. After tuning the parameters, the final xRebounds model had an area under the ROC curve of 0.9017 and an accuracy of 0.8097. Additionally, the model had a specificity of 0.8418 and a sensitivity of 0.8071, making this a usable piece of the xReboundsPlus model.

The next step is creating the xAngle model, unlike the previous model for xRebounds, this model needs to output a continuous variable: the rebound shot angle in degrees. To account for this, multiple different types of models were created and compared to determine the best possible fit. The models that were considered for this expected angle prediction were a neural network, another XGBoost model, a random forest model, multiple linear regression, and Bayesian regression. A switch to Bayesian regression was made after the other methods failed. Using Bayesian regression allowed the model to be able to make predictions based on prior events and is better for modeling uncertainty. The Bayesian regression model did converge with all Rhat values being 1 meaning there were no signs of divergence or sampling issues. The selected Bayesian regression model had a low Bayesian R-Squared value of 8%, but this was comparatively better than the other models.

Once the models were created for xRebounds and xAngle, the xReboundsPlus model was ready to be created. Before the model was created, additional data manipulation was needed to bring in the xAngle and xRebound predictions for each shot. Additionally, if the next shot was a goal and the distance of the next shot were brought over to each shot observation as well. Finally, the data was filtered to only shots that generated rebounds to have the proper sequencing for the training.

XGBoost was selected as the machine learning method with K-fold cross-validation. The independent variable were xRebounds, xAngle and the distance of the rebound shot. After tuning, the model had an area under the ROC curve of 0.9242, which means that it was able to classify well but was not at risk of overfitting as the original model. A confusion matrix was created as well achieving an accuracy of 80.99%.

After fully creating xReboundsPlus, the model was implemented on shot data for the 2024–25 NHL up until the morning of March 26, 2025. Since there was no actual distance that could be used for each individual shot, the average distance of a rebound was used for each calculation. Once the statistic was applied to each shot, rankings were created based on the sums of xReboundsPlus, xRebounds, and shots. Brady Tkachuk led the league in the statistic with 94.85 xReboundsPlus, followed by Matt Boldy and Nathan Mackinnon with 89.39 and 86.09 respectively.

This metric provides a framework for measurement of rebound opportunities, but there are still many ways to improve the model. With data that is able to track where the shot would hit the net and the positioning of the goalie, the prediction could become even more accurate as the likelihood of a rebound and the expected angle could both become more accurate as the model will more accurately depict hockey in three dimensions. Additional shot velocity or goalie specific data would improve the model as well due to differences in reaction time. Overall, the project was able to meet the original goal by creating a new rebound-based statistic, but still has room for improvement. The full code used to create this project can be found on my Github, https://github.com/elawing40.