Lord Stanley Seekers

Crooks, Kevin
crooksk@colorado.edu

Lim, Brian
brian.lim@colorado.edu

May, Arthur
arthur.may@colorado.edu

May 3, 2020

Figure 1: A full penalty box during an NHL Game.[1]

1 Problem Space

In the world of sports, predicting a games outcome has myriad benefits. It can help coaches and players learn what areas to focus on. It can help sportswriters write articles outlining the predictions for season-final standings. It can help oddsmakers set profitable betting lines. Unfortunately for these professionals, current best NHL game prediction accuracy is less than 70 percent. This is generally attributed to simple randomness in sports, the general perception of higher parity in the NHL as opposed to other sports (we all know the Patriots are going to be in the postseason, right?), and of course the randomness of ice condition. Bad bounces, breaks, millimetres in difference in skate angle, etc. Nonetheless, 70 percent isnt bad, but naturally everyone wants these to be better.

Rather than retreaded an old problem space, which would simply amount to wondering if the science has improved in the past few years, or training algorithms have, we decided to look at a new problem: period prediction. An NHL game is divided into 3 periods of twenty minutes each. If the score is still tied after these three regulation periods, a 4th overtime period is played. Overtime rules have changed a few times in the past couple of decades, but the current standard is a single 5-minute period. We included these in our analyses as well. Despite the fact that statistics will be different between a full 20-minute period and a shorter 5-minute period, we are looking at predictors for each period individually, so our overall models will not be affected by this bundling.

The advantage to such a prediction might be immediately obvious to a coach, although probably not as useful for Vegas. If your team enters the second period trailing by 2, how do you overcome that lead? If hits have a high correlation with success, maybe play more aggressively. If blocked shots have a low correlation, dont worry about taking too many bad shots at the risk of turning the puck over. Our goal is to see if we can find which statistics, taken from the standard metrics used in NHL statkeeping, can be beneficial in predicting the outcome of a period.

Similar research has been done in this space on the depth of such research, but not using the same approach. In one research study, players are ranked according to how much their actions impact their team's chance of scoring the next goal.[3] Additionally, in another study done on NHL data, researchers attempt to predict outcome based on pre-game data.[3] Lastly, one study was done to predict NHL playoffs using Google's PageRank feature.[4]

2   Approach

Our approach to address the problem iterated through three steps over multiple phases. The first step was to prepare the data. Once the data was prepared, we performed a number of visualizations around the data to get an understanding of relationships between features and the end of period goal differential. Once we had an understanding of the relationships between the features, we ran them through different machine learning classifiers based on the results we were seeking. We did not seek a binary prediction, but a goal differential which is a bit more challenging. Furthermore, as we settled on our models we experimented with different normalization approaches and ran a grid search on our classifier to help fine tune the results. We performed forward stepwise selection to determine the best subset of features to be used to predict the outcomes of each period, and used the best subset per period to train and predict the models for the individual periods.

Several classifiers will be tested including a gradient boost, random forest, and KNN. The classifier that performed the best will be used to present the results.

When our results were output we output confusion matrices to help us interpret how our model was performing relative to the other potential outcomes of the model.

3 Data

The data for this work was obtained from Kaggle, comprised of 3.6 million plays over 11,400 games over the course of 3 full seasons. The plays including the standard stats recorded by the NHL goals, penalties, shots on goal, blocked shot, hits, faceoffs, takeaways, giveaways, and missed shots. This data was synchronized with 8.9 million shifts for the same games (a shift is defined as a consecutive time on ice for a player, typically 20-40 seconds, and a player might have 40 shifts over the course of a single game). This data was then split by home and away, the artificial split we used as the easiest way to differentiate teams. Concordantly, the predicted outcome, final period goal differential, was also given as Away minus Home. Thus to determine final goal differential for the away team, simply take the negative of the result. The resultant feature matrix was 37589x19: 37,589 periods for which we had valid statistics and 19 statistical measurements, originally given as raw counts to be later normalized and scaled depending on the statistic in question. A sample of the data can be found in Table 1.

A small handful of periods had to be rejected from having missing statistics, or obviously incorrect statistics. This generally happened in overtime periods for some unknown reason. Additionally, we noted a minor statistical error, probably attributed to typographical errors at the input level, in which a players shift length was given as 1-2 seconds, followed by a regular 30 second shift. The number of these was small enough that they were included regardless, reasoning they would have little effect on the overall average shift length feature, and indeed in the final outcome itself.

After preparing the data, a number of visualizations were performed to better understand the relationships.

Figure 2 contains a Seaborn violin plot. In this particular plot the difference between the home and away value of each attribute against the change in goal differential at the beginning and ending of the period. The penalty differential and takeaway differential appeared to have a strong relationship which makes sense. If you are on a powerplay more than your opponent in a period, you will have more goals. Same argument can be made for takeaways. Nonetheless, some attributes had more impact on the outcome than others.

The same analysis was done attribute by attribute using a Seaborn joint plot. A sample plot can be found in Figure 3.

4 Results

A number of multi-class classifiers were tested including KNN, Random Forest, and Gradient Boost. No classifier performed particularly well relative to our 60% accuracy target with results being output in the 35-45% ac-

periodID	period initial goal differential	blocked shots away	blocked shots home	faceoff away	faceoff home	...	period final goal differential
2010020001_1	0	7	6	9	5	...	1
2010020001_2	1	8	13	5	8	...	0
2010020001_3	1	6	3	9	7	...	0

Table 1: NHL game data prepared sample
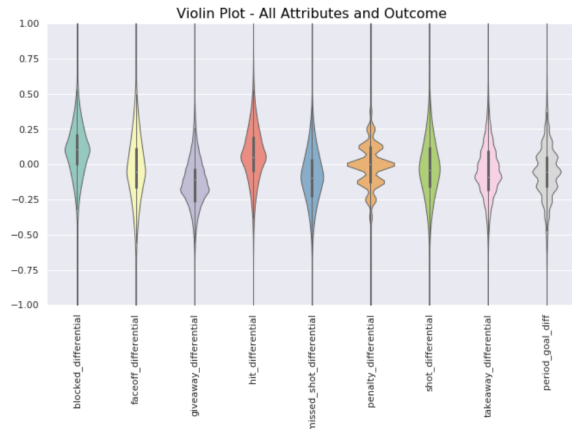
Figure 2: Violin plot of difference between home and away attributes vs. Outcome

curacy range for the first three periods with increasing accuracy for each period. These results can be found in Figure 4, Figure 5, and 6.

In search of a conclusive result, we included shift data and overtime results into our features and data. Shift results didn't do much to sway our results. However, when running our classifier for overtime, the results were much more favorable with accuracy scores of approximately 75%. Details of the accuracy score and confusion matrix can be found in Figure 7

To achieve our results with the gradient boosting classifier, score differentials were
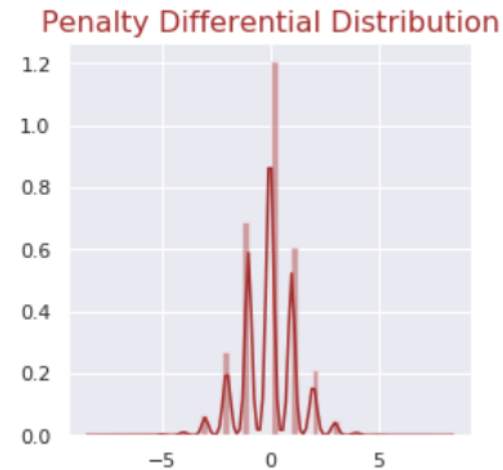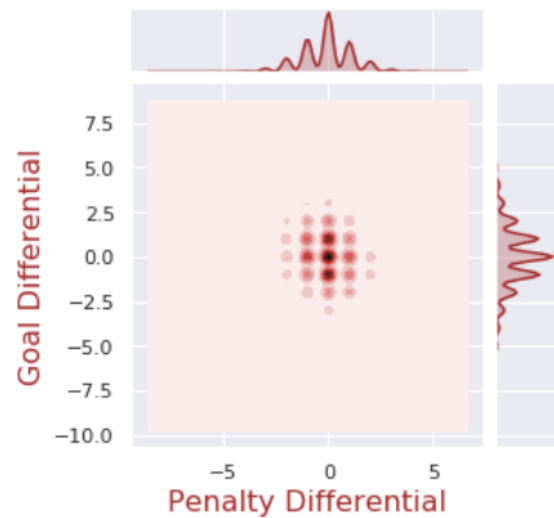
Figure 3: Joint plot of penalty differential and goal differential for the period

Period 1
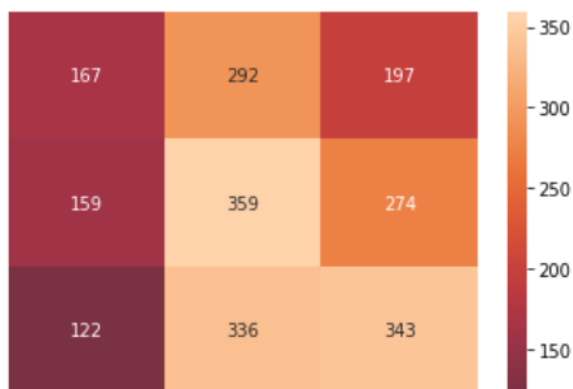Gradient Boost Acc: 0.3863939528679413

Figure 4: Period 1 accuracy score and confusion matrix heat map.

Period 3
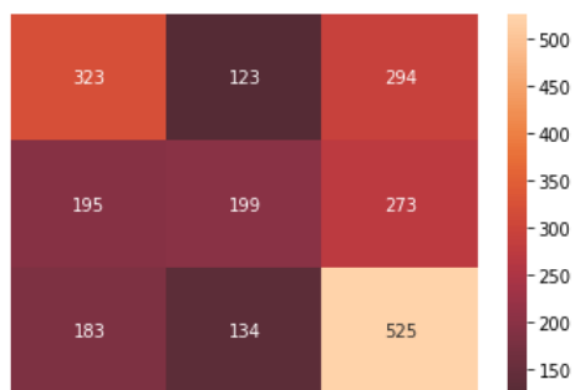Gradient Boost Acc: 0.4655402401067141

Figure 6: Period 3 accuracy score and confusion matrix heat map.
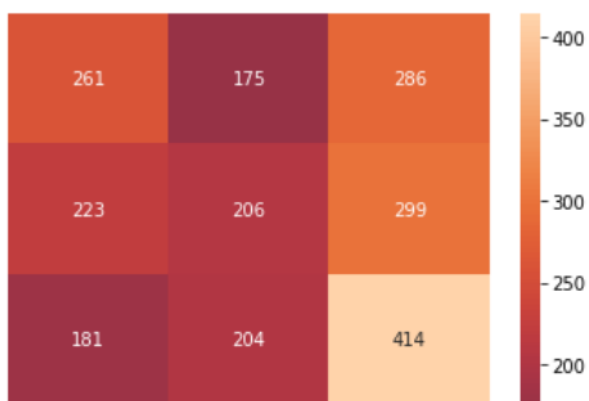
Period 2
Gradient Boost Acc: 0.3917296576256114

Figure 5: Period 2 accuracy score and confusion matrix heat map.

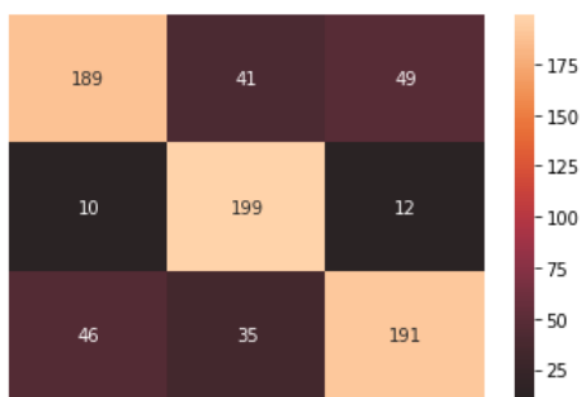Period 4
Gradient Boost Acc: 0.75

Figure 7: OT accuracy score and confusion matrix heat map.

binned to $\{-1, 0, 1\}$. A MinMaxScaler was applied to all attributes to normalize the data and help with interpretation and comparison of models. A GridSearch was done to help identify and understand the impact of each of the parameters on the model. Once a model was found, different iterations were performed using different features of the data based on our initial visualization analysis of features that correlated well. We performed forward stepwise selection to determine the best subset of features to be used to predict the outcomes of each period, and used the best subset per period to train and predict the models for the individual periods.

5 Discussion

The ability to predict the outcome for overtime is considered a significant contribution. Perhaps predicting a goal differential in terms of a number of goals was too arduous for the duration of this research. However, there are a number of reasons why the overtime period was as accurate as it was.

The overtime period is a win, lose, or draw scenario. There less classes of goal differential. Furthermore, the score is always tied going into overtime so the conditions of the game are completely different than the the second and third periods. It is understood that the first period starts off as a tie a well. However, the features for the game in the first period are not yet established. The dynamic of the game is not yet established.

In terms of predicting goal differential, we may have had better results if we attempted

to calculate a goal differential and allowed non-integer values to be a result. It is understood that you can't have a partial goal differential. However, a fractional goal differential extrapolated over an entire season could have a significant impact on a teams placement in the standings.

When performing the forward stage-wise feature selection, it was interesting to see that different features proved more important when trying to predict the outcomes of each period. This can be interpreted to mean that depending on the phase of the game, there are different strategies that can be employed to obtain a desired outcome. Furthermore, there were some features that were important across all the periods, meaning that there are some general strategies that are independent of the game situation. For example, the penalty differential was an influential feature across all periods, meaning that it is a sound strategy to avoid taking penalties and putting your team at a disadvantage in all phases of the game. While penalties were important across all the periods, stats like blocked shots were most important in the third period, which can be attributed to a more defensive style of play used when a team is trying to preserve a lead. Likewise, the number of hits proved important in the first period and overtime periods, which likely means that using a more aggressive style of play to obtain possession of the puck will lead to more positive outcomes in these periods. Leveraging this type of feature analysis could be used by teams to adjust their strategies in real time to try to obtain the desired outcome.

References

[1] Predators Hockey, USA Today, 2015
https://usatftw.files.wordpress.com/2015/11/ap_jets_predators_hockey_77540060.jpg

[2] , Oliver Schulte, Zeyu Zhao, Mehrsan Javan, Philippe Desaulniers, *Apples-to-Apples: Clustering and Ranking NHL Players Using Location Information and Scoring Impact*
http://www.sloansportsconference.com/wp-content/uploads/2017/02/1625.pdf

[3] , Josh Weissbock and Diana Inkpen, *Combining Textual Pre-game Reports and Statistical Data for Predicting Success in the National Hockey League*
https://link.springer.com/chapter/10.1007/978-3-319-06483-3_22

[4] , Nathan Swanson, Donald Koban, Patrick Brundage, *Predicting the NHL playoffs with PageRank*
https://www.degruyter.com/view/journals/jqas/13/4/article-p131.xml