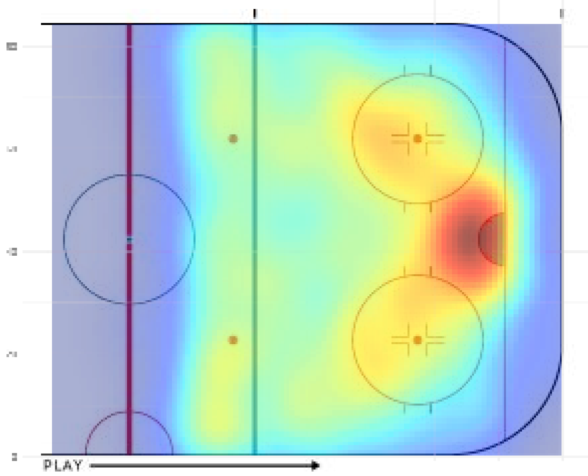
The Pass Is Right:
An Analysis of Decision Making To Improve Scoring Chance in Women's Hockey

Tej Seth, University of Michigan
Santiago Casanova, Instituto Tecnológico Autónomo de México



I. Introduction

A. This study analyzes key factors that could improve in-game pass/shoot decisions to better capitalize on offensive opportunities in women’s hockey. To determine the optimal course of action and to better understand situational scenarios, we developed an array of statistical models and extrapolated variables regarding pass completion probability and expected goals that aid decision making.

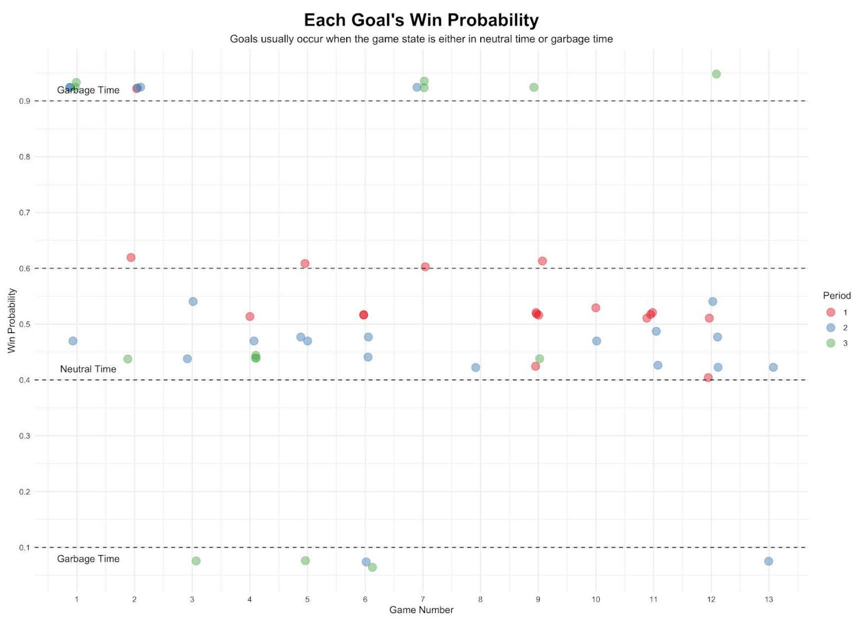
II. Data

A. 50,884 plays from NCAA, Olympic and NWHL Women's Hockey, ranging between 2018 and 2021 provided for the Big Data Cup 2021

III. Models

A. Win Probability

1. To understand the flow of the game, we created a win probability model based on score differential, total time remaining, skater differential, time left in the period and whether it was first, second or third period. Of the 10 games to finish with a winner in regulation, the model had correctly calculated 9 of those winners with a higher win probability at the end of the 3rd period.
2. With score differential being the driving factor in the model, it was also important to note that goals usually occurred when the game was in neutral time (between 40%



and 60% win probability) or when the game was in garbage time (>90% or <10% win probability).

B. Pass Completion Probability

1. An important factor in determining whether it is optimal for a player to pass or shoot is the probability of said pass being completed. Our Pass Completion Probability

model used pass distance, on-ice skater differential, x coordinates for both players involved, and pass type, specifically, tape-to-tape or indirect.

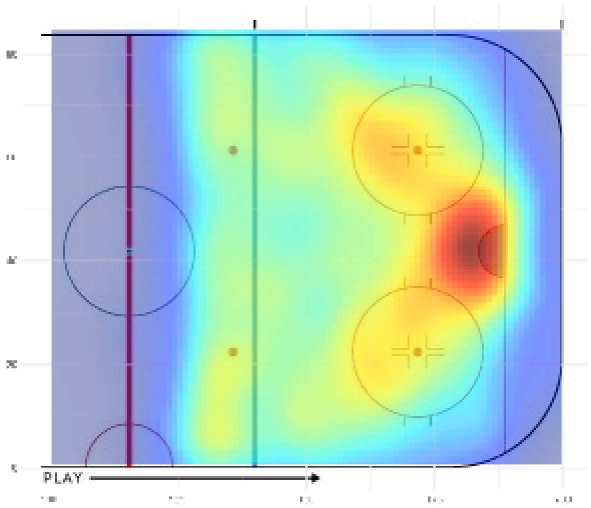
2. We utilized “area under the ROC curve” (AUC) as the performance evaluating metric for our model, as it is an effective metric when evaluating imbalanced classification problems. Our final AUC value on unseen data was 0.744, which suggests sufficient discrimination, as the rule of thumb deems values ≥0.7 acceptable (Hosmer & Lemeshow).

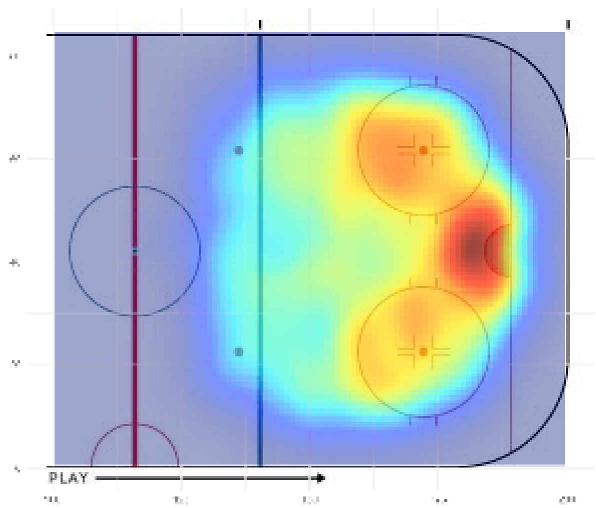
C. Expected Goals

1. As Figure 1 above shows, there is an observable relationship between goals scored and win probability at the time of the shot. Accordingly, we utilized this relationship as a feature when estimating expected goals. . We also utilized shot distance and shot angle following David Sumpter’s principles for expected goals in soccer to provide the model with the first of the two most important references for prediction: location and situation.
2. Following the same AUC evaluation metric, we found an AUC value of 0.822, which suggests excellent discrimination on unseen data.

D. Expected Pass/Shot Rate

1. For the final component in our decision analysis, we developed a model to predict the action a player might take at any point during a hockey game . To do so, we trained an extreme gradient boosting algorithm using the x coordinate of the player with the puck, the expected goals (xG) at that moment and position, score differential, on-ice skater differential, and the time remaining in the period. This model yielded outstanding results, as the AUC value on unseen data was 0.947. This accuracy in estimation is observable when we look at a shot density map (Figure2) next to an expected shot density map (Figure 3), albeit more condensed.





IV. Methodology

A. Based on the models we created and the values we were able to obtain from them,, we created three metrics to develop decision-making criteria. Since we do not have information regarding the players on the ice, exclusive of the puck handler, a hypothetical pass option for shot evaluation was not possible. With this constraint in mind, we selected only the passes in our data for further analysis. Once the situational values were calculated - Expected Goals (xG_1), Pass Completion Probability (CP) and Expected Goals After the Pass (xG_2) - we created the following metrics:

1. Pass Risk: The xG percentage risked by passing the puck instead of shooting at any given moment. Denoted by μ

$$\mu = xG_1 - (CP) * xG_2$$

2. Pass Reward: Whether the xG increased or decreased after the pass. A negative pass reward would indicate a decrease.

B. The decision formula follows an established set of conditions in order to maximize the scoring chance. If we break down the logical structure, the formula suggests a shot if:

1. xG is greater than 12%
2. Pass reward is negative ($xG_1 > xG_2$) AND $xG_1 > 7.7\%$*
3. $\mu > ß$ AND $xG_1 > 7.7\%$

Where ß is a constant adjustable to team play style/preference.

The 7.7% xG cutoff was chosen as exploratory analysis indicated that there was a “level change” in the average observed goals scored close to that value (Figure 4).

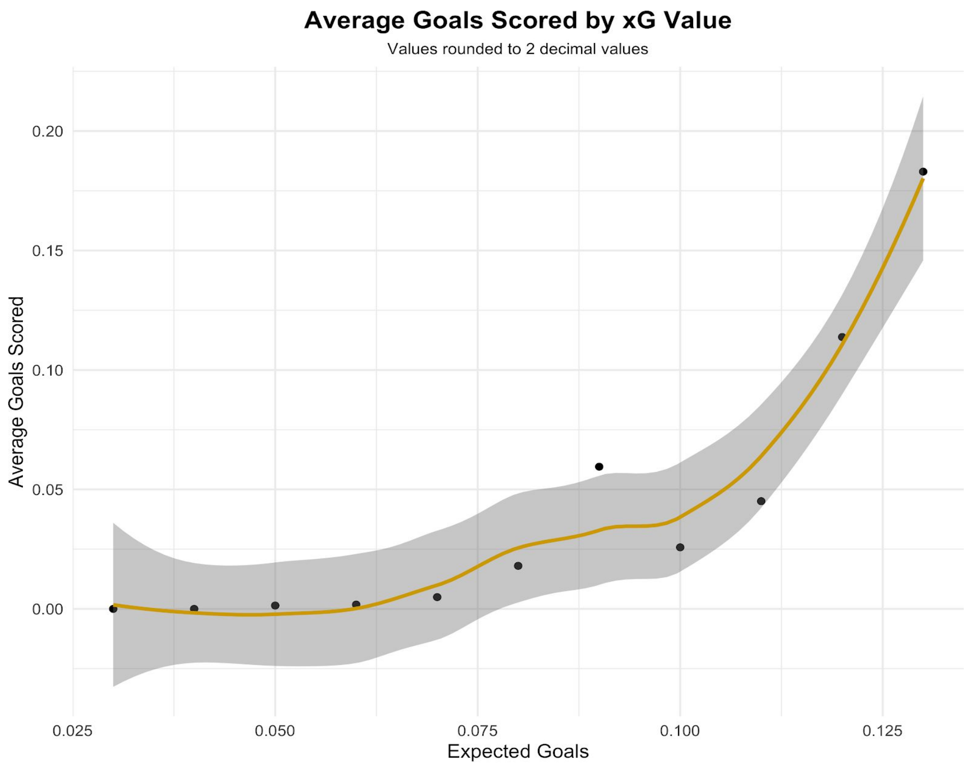


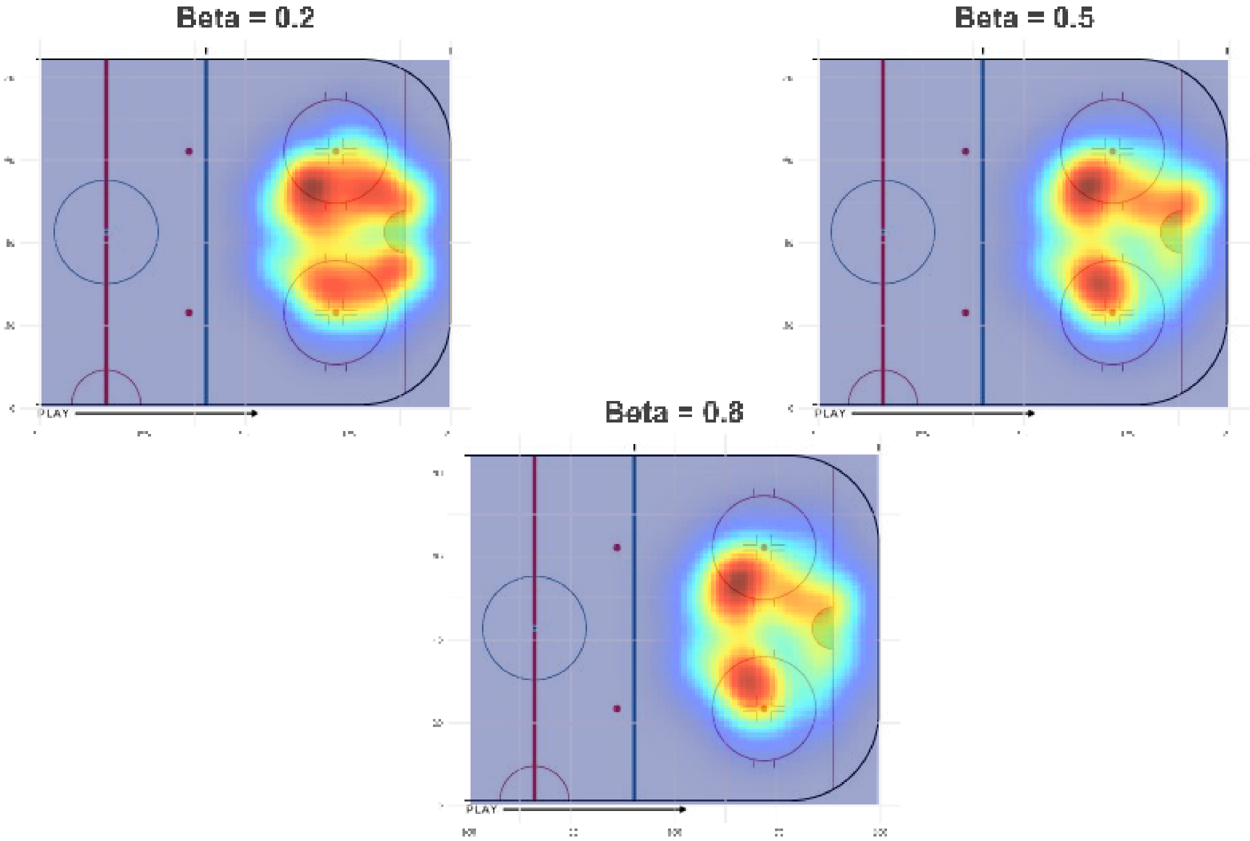
Figure 4

The ß parameter is adjustable from small values for puck security/shot volume styles of play, where the goal is to seize as many scoring opportunities as possible, to large values used in a conservative offense that looks for the right shot opportunity.

V. Results

A. When comparing suggested shot density against plays that resulted in a pass during a game, we get the "area of improvement" plays that the decision making formula deems shot-worthy. In the following maps we can observe the difference in suggestions with a low, balanced and high ß value. (Figures 5,6,7)

When compared to the actual and predicted shot density maps, it is clear that the areas highlighted here correspond to under-utilized attacking zones where the formula believes there is value in shooting more often, even when utilizing conservative ß values.
As the ß value increases, therefore a higher pass risk is tolerated, shots with a less-than-ideal view of the net are deprecated and substituted by passes back to the point and shots with a head-on look.



Let's take a look at this play during the 2019 WWC Gold Medal Game between team USA and team Finland. Annie Pankowski gets a pass right in our established “hotspot” and is suggested a shot by the model. The play resulted in a pass and a lost possession

VI. Appendix

A. Sources

1. David Sumpter xG Methodology as seen on the “Friends of tracking” YouTube channel.
2. Hosmer & Lemeshow (2013). Applied logistic regression. P.177
3. IIHF Rulebook: https://blob.iihf.com/iihf-media/iihfmvc/media/downloads/rule%20book/iihf_official_rule_book_2018_ih_191114.pdf
4. 2019 IIHF Ice Hockey Women’s World Championship Gold Medal Game: USA vs Finland via TSN

B. Code Repository:

1. https://github.com/SCasanova/Big-Data-Cup-2021