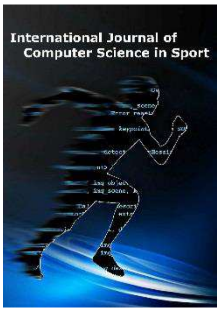
International Journal of Computer Science in Sport

Volume 23, Issue 1, 2024

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2024-0001



The Use of Momentum-Inspired Features in Pre-Game Prediction Models for the Sport of Ice Hockey

Jordan T.P. Noel[1], *Vinicius Prado da Fonseca*[1], *Amilcar Soares*[2]

[1] *Department of Computer Science, Memorial University of Newfoundland and Labrador, St. John's, Canada*

[2] *Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden*

Abstract

We make a unique contribution to momentum research by proposing a way to quantify momentum with performance indicators (i.e., features). We argue that due to measurable randomness in the NHL, sequential outcomes' dependence or independence may not be the best way to approach momentum. Instead, we quantify momentum using a small sample of a team's recent games and a linear line of best-fit to determine the trend of a team's performances before an upcoming game. We show that with the use of SVM and logistic regression these momentum-based features have more predictive power than traditional frequency-based features in a pre-game prediction model which only uses each team's three most recent games to assess team quality. While a random forest favors the use of both feature sets combined. The predictive power of these momentum–based features suggests that momentum is a real phenomenon in the NHL and may have more effect on the outcome of games than suggested by previous research. In addition, we believe that how our momentum-based features were designed and compared to frequency-based features could form a framework for comparing the short-term effects of momentum on any individual sport or team.

KEYWORDS: MOMENTUM, ICE HOCKEY, NHL, PREDICTION, APPLIED MACHINE LEARNING

Introduction

Since the early 2000s, the National Hockey League (NHL) has undergone a significant change regarding the use of analytics (Nandakumar & Jensen, 2019). Teams now have an amplified reliance on advanced analytics, leading them to establish in-house analytics departments staffed with data scientists, statisticians, and computer scientists that focus on aspects from player performance all the way to ticket pricing (Mondello & Kamke, 2014). While the acceptance of advanced analytics has grown in the NHL, the league still trails most other major leagues in accepting and using the information in daily decision-making (Goldman, 2022). This could be because analytics can often challenge or refute phenomena that players, coaches, and fans often view as tangible and real. An example of such a phenomenon is momentum. Although momentum is frequently considered a discredited concept in analytics, anecdotal evidence is reported by many players and coaches. In the context of this study, momentum refers to the notion that successive achievements or failures can influence the likelihood of future success.

Academic research in momentum often can be broken into several categories starting with individual momentum and team momentum. Moreover, momentum research can be further classified into intra-game momentum (within a match) and inter-game momentum (between matches). Most of the cited literature in our paper revolves around inter-game momentum. The most influential piece of research around momentum is a paper written by Gilovich, Vallone, and Tversky (1985). This research found that when shooting free throws in a controlled setting, the outcome of each free throw was independent of the previous outcomes. This means that groups of converted or missed shots were simply created by chance and did not indicate a shooter was "heating up" or "cooling down." Thus, the “Hot Hand Fallacy” concept was born, stating that humans often suffer from a cognitive bias in which they perceive streaks in random sequences where none exist. Although the findings of Gilovich et al. (1985) are generally undisputed, contradictory results have emerged in this field of study. For instance, the research conducted by Arkes and Martinez (2011) demonstrated evidence of positive momentum in winning streaks when considering factors such as team quality and recent opponents. This implies that winning increases the likelihood of winning the next game, with no distinction between home and away teams. Research conducted by Pelechrinis and Winston (2022) argued that only looking at free throws did not accurately depict what happens during your average shot in a game. Instead, using actual NBA event data they found strong evidence of the hot hand effect in certain players when accounting for variables such as shot distance from the rim and defensive coverage, suggesting that momentum can influence individual performance.

While a large amount of research has been published regarding momentum in other sports, limited research has been conducted on the concept of momentum in ice hockey. Kniffin and Mihalek (2014) examined a group of two-game series in the NCAA; when accounting for team quality, the outcome of the first game did not affect the outcome of the second game. Similarly, Steeger, Dulin, and Gonzalez (2021) reported comparable findings, where the wins and losses of only one team provided evidence to reject the null hypothesis of random wins and losses in the 2017 NHL season. This limited amount of research is unsurprising, as ice hockey is often one of the least researched sports. For example, literature regarding pre-game prediction in the NHL is often scarce, and in our research, we encountered only four relevant studies based on pre-game prediction: Weissbock and Inkpen (2014), Weissbock, Viktor, and Inkpen (2013), Remander (2021), and Pischedda (2014). Our paper contributes to this area of research by introducing a novel approach to pre-game prediction in the NHL and expanding the overall body of literature. However, the most frequently cited paper we found in the NHL prediction field is the work of Gu, Foster, Shang, and Wei (2019). They concluded that a goalie's save percentage is critical in determining game outcomes, suggesting that goalies can single-handedly win or lose a game. Nonetheless, a goalie's save percentage can vary considerably randomly from game to game.

Previous studies have suggested that the upper limit of accuracy in pre-game NHL prediction is approximately 62%. This finding was initially demonstrated by Weissbock and Inkpen (2014). They came to this conclusion by simulating an NHL season 10000 times using Monte Carlo simulations and assigning random strengths to each team at each iteration. They then experimented with different trade-offs between skill and luck in the simulated outcomes, such as outcomes being determined 100% by skill or by 50% skill and 50% luck. The standard deviation of the win percentages obtained in the simulations was then compared to the standard deviation of win percentages for all NHL teams between the 2005 and 2011 seasons. They found that the standard deviation of the win percentages in the 2005 to 2011 seasons was most like a simulated season with a trade-off of 24% skill and 76% luck. With this information, they derive that if 24% of outcomes are determined by skill and 76% of outcomes are determined by luck, they can mathematically determine that the upper bound for prediction in the NHL is 62% as 24 + (76/2) = 62.

The conclusions of this approach have been supported by Pischedda (2014), who used an actual pre-game prediction model based on machine learning and could not reach the 62% ceiling, achieving an accuracy of 61.54% over 517 matches. Weissbock and Inkpen (2014), in the same paper in which they proposed this ceiling, created a pre-game prediction model that could not reach the 62% ceiling, achieving an accuracy of 60.25% over 720 games. In a more recent thesis by Remander (2021), they attempted to predict outcomes of the 2016 through 2019 NHL seasons and achieved an accuracy of 55.4% over 5043 games. However, we do not believe that the approach of Weissbock and Inkpen (2014) is infallible as it is based on simulation and not real-world game outcomes. Their method also relies on using the standard deviation of win percentages; in theory, they could have used a different metric, such as mean absolute deviation, or they could have compared the complete distributions with tests such as the Kolomogorov-Smirnov test. Changing the measurement to compare the distributions could lead to a different theoretical ceiling. Therefore, concluding that a ceiling exists in NHL pre-game prediction is fair, but the actual ceiling could differ from 62%.

Previous research, such as the studies conducted by Lopez, Matthews, and Baumer (2018) and further supported by Gilbert and Wells (2019), has extensively explored the randomness of NHL game outcomes. These studies demonstrated that NHL game outcomes exhibit a considerable degree of randomness, and often, the team with superior performance indicators before the game would end up losing. Lopez et al. (2018) also show that in a tournament setting such as the NHL playoffs, there only exists a 19% chance that the team with the best performance indicators will win the tournament due to this randomness. However, these studies were performed on a pre-game basis. What about instances of in-game performance indicators? Research conducted by Daniel Kari (2020) found that even when using several popular in-game performance indicators for NHL games and a convolution neural network, they could not achieve a high level of predictive power in NHL games, as they could only predict the correct game-winner with an accuracy of 61.6%. It is worth noting that Kari's study did not include team-specific performance indicators such as goals, save percentage, or shooting percentage; including goals in the model would lead to 100% accuracy, as the team with the most goals will always win the game. At the same time goals scored is very closely tied to shooting and save percentage as they respectively represent the percentage of shots that result in a goal or the percentage of shots a goalie saves. As the goal of the study was to look at the effects of randomness in the NHL and compare how certain performance indicators often used to forecast a team’s future success would act in in-game prediction they opted to not include these performance indicators as they would not yield very interesting or useful results. Based on these prior studies, there is strong evidence that NHL game outcomes are subject to a high level of randomness.

This randomness greatly affects our ability to make accurate predictions about game outcomes. For instance, in 2022, the sports betting sector achieved a record gross revenue of USD 7.5 billion (American Gaming Association, 2023). If there were currently a highly accurate method for predicting NHL game outcomes, this multi-billion-dollar industry would undoubtedly adopt it for setting initial odds, as the methods bookmakers utilize to establish these odds are integral to the profitability of their business (Hubáček, Šourek, & Železný, 2019). However, as was described by Osborne (2020), NHL betting underdogs emerge victorious in 41.1% of games.

Several different factors could cause the perceived randomness of outcomes in the NHL. There may be undiscovered performance indicators in the NHL that could provide better insights into game outcomes. Alternatively, the game itself may simply possess a high degree of randomness, whereby "luck" plays a significant role in determining results. Weissbock and Inkpen (2014) suggest that this randomness may have something to do with the parity of the NHL, which is created by the hard-ceiling salary cap. This means that teams are only allowed to spend up to a certain amount of money on player salaries, making creating “super-teams” funded by wealthy ownership groups impossible. However, parity alone does not account for all the observed randomness in the league. We, therefore, believe that momentum may be a missing piece of the puzzle and could potentially help account for some of this randomness.

To date, there is a lack of papers proposing a quantification method for team momentum as a performance indicator and an absence of studies attempting to address inter-game team momentum using machine learning techniques. We perceive this as a gap in the existing momentum literature we aim to address. Instead of looking at whether wins and losses were dependent or independent events, we used momentum-based features to create machine learning models and compare their predictive power against models trained with the same algorithm that only used more traditional frequency-based features as well as a model that used a combination of both feature sets. We hypothesize that these momentum-based features can potentially find evidence of momentum in ice hockey where previously none has been found. We also believe these features can provide a new momentum-based framework for creating performance indicators in the NHL and other sports. Additionally, these features can help determine the impact of momentum on NHL game outcomes, shedding light on some of the perceived randomness in the league. Furthermore, the comparative analysis of our three feature sets can serve as a foundation for future investigations into the effects of momentum on different sports.

Methods

The organization of this section follows the flow of data through our proposed pipeline, with a final summary presented in Figure 1.

Raw Event Data

Our raw database was created using the Python module *hockey_scraper*, which utilizes the NHL's API to retrieve event data for NHL games (Shomer, 2019). Our study collected all event data for regular-season NHL games from 2011 to 2020, resulting in a dataset comprising 10,602 NHL games. The raw data utilized in this research encompasses various game events recorded by the NHL, including shots, misses, goals, blocks, takeaways, giveaways, and more. The specific events we track and utilize will be discussed in the subsequent section on game events extraction. Each event entry contains information regarding its timing, such as the associated game, the period in which it occurred, and the remaining time in that period. Furthermore, situational details are provided, including the teams involved, the players participating, and the score at the time of the event. This situational information holds significance for our analysis as we incorporate "5v5" and "close" values throughout the paper. The term "5v5" in a value signifies that the event count only includes occurrences when both teams had five players and a goalie on the ice, thereby eliminating potential bias arising from power plays and penalty kills. On the other hand, the term "close" in a value indicates that the event count only encompasses instances within a closely contested game scenario. Specifically, we define this as a score differential of one or less in the first two periods or a score differential of zero in the third period or overtime. This definition addresses the notion that specific teams may adopt a more defensive-oriented approach when holding a substantial lead. While the raw event data consists of 56 columns of information per event entry, a significant portion is irrelevant to our objectives. Table 1 provides a small sample illustrating the structure of the raw data.

Table 1. A small example of what the scraped data from hockey_scraper looks like. It shows several columns that we use in our calculations.

Game_Id	Event_Team	Event	Strength	Period	…
201200001	MTL	SHOT	5x5	1	…
201200001	TOR	BLK	5x5	1	…
201200001	TOR	PENL	5x5	1	…

Game Events Data Extraction

We performed in-game event extraction using the raw events dataset, specifically focusing on counting events by type for each team within individual games. The extracted events included blocks, faceoffs, giveaways, goals, hits, misses, penalties, shots, and takeaways. These events play a crucial role in subsequent interval-based extraction. The process of extracting game events data involved traversing the raw events data and tallying the occurrence of each event type for a given team in a specific game. This grouping of events was facilitated by utilizing the unique game ID. To illustrate, we examined the data to determine the number of shots taken by the home team in a particular game by counting the instances of the "SHOT" event attributed to the home team. For instance, if the home team had 25-shot events, it indicated they had taken 25 shots in that game. Another example would be calculating the number of blocks for each team by summing up the "BLOCK" events associated with each team. Table 2 provides an example showcasing the structure of the game events data.

Table 2. A small example of what our game events look like after events have been extracted from the raw dataset.

Game_Id	Home_Team	Away_Team	Home_Shots	Away_Shots	Home_Blocks	…
201200001	MTL	TOR	25	29	15	…
201200002	BOS	PIT	35	26	9	…
201200002	TBL	FL	20	29	10	…

Interval-Based Data Extraction

We utilized the game events data to extract pre-game team quality assessments by considering intervals of the previous three to seven matches for both the home and away teams. Although not the direct inspiration for our work, Weissbock, Viktor, and Inkpen (2013) briefly mentions using a recent number of games to evaluate team quality. It's important to note that this data extraction was performed pre-game, meaning we only extracted data from games preceding the specific game we intended to model. To illustrate the procedure of interval-based data extraction, let's consider an example. Suppose we are working with an interval of the last three games, and we want to create an instance for game number 40, where we aim to obtain the average number of goals for both the home and away teams. To achieve this, we calculate the average number of goals for both teams (home and away) using only their 37th, 38th, and 39th games. During the interval-based data extraction, we excluded certain games from the datasets. This included all overtime games, as they are more likely to produce outliers in individual statistical categories due to the longer game duration and the recent change from 5v5 overtime to 3v3 play, which affects the number of players on the ice. Additionally, we removed the first 20 games played by each team in every season, as teams tend to struggle with consistency during this initial phase, which could potentially skew the training phase of our machine-learning algorithm. Although we experimented with removing 15 and 25 games, removing 20 games yielded the best predictive power in our algorithms. This step aligns with standard practices in other public pre-game prediction models, such as the one developed by MoneyPuck.com (Tanner, n.d), which also removes the first 20 games of each season during training. In addition to the data from the NHL’s play-by-play, we incorporated expected goals (xG) data from MoneyPuck.com (Tanner, n.d), which was valuable to our dataset. Overall, this approach created five datasets, each corresponding to different time windows of 3 to 7 games, totalling 5 setups. Each dataset consisted of 5,730 instances.

Frequency-Based Features

We utilized frequency-based features to capture teams' performance over several previous games. These features were categorized into three types: sum-based, average-based, and percentage-based features. Sum-based features are created by totalling a specific event's occurrences throughout the selected game window. For example, if we considered a four-game interval and wanted to create a sum-based feature for shots taken by a team, we would sum up the total number of shots taken by the team in their last four games. As the name suggests, average-based features were based on the average value of a particular event over the chosen interval of games. Instead of summing the occurrences, we calculated the average value. For instance, we could compute the average number of goals scored by a team in their last four games. Percentage-based features were calculated using data from all the games within the specified interval. These features represented the percentage or proportion of a specific event out of all instances recorded in the given window of games. For example, we could determine the percentage of face-offs won by a team in their last four games. By incorporating these frequency-based features, we aimed to capture the patterns and trends in team performance over a certain period, providing valuable insights into their playing style and effectiveness in various aspects of the game.

Momentum-Based Features

Various definitions of momentum have been proposed in the scientific literature over the years. For instance, Arkes and Martinez (2011) define a "momentum effect" as a situation where a team is more likely to win or achieve success if they have been performing well in their recent games. On the other hand, Steeger et al. (2021) do not explicitly provide a definition of momentum but differentiate it from a winning or losing streak by suggesting that streaks are observed sequences of wins or losses that may or may not be related, while momentum implies a dependence between similar events. Our research defines momentum as a consistent increase or decrease in the overall quality of play over a specific number of previous games. Our definition of momentum considers it as something that can be estimated using engineered features that determine the trend of a team's quality of play for the upcoming game. Although our definition aligns more closely with Arkes and Martinez (2011), we exclude the aspect of winning from our definition due to the random nature of the NHL, where game outcomes are subject to chance.

In this study, we have extracted two types of momentum-based features. The first type is the slope-based feature. This feature focuses on the interval of recent games, specifically the last three games for each team. Using a selected statistic, such as the number of shots, we plot the number of shots taken in each of these recent games on a two-dimensional space. The y-axis represents the number of shots taken in a game, while the x-axis represents the passing of days between games, starting from zero for the earliest game in the interval. We then determine the linear line of best fit for these plotted points, and the slope of this line becomes the value of the slope feature. The second type of momentum-based feature is a projection feature that utilizes the same line of best fit created for the slope feature. In this case, we use the equation of the line of best fit to project a given statistic for the team in the game for which we are predicting the outcome. By employing the function $y = mx + b$, where x represents the number of days that have passed since the first game in our interval of games, m denotes the slope of the line of best fit, b represents the intercept on the line of best fit, and y represents our actual projection, we obtain a feature that provides insight into how the team is expected to perform if their recent trends continue.

Features Used

Table 3 presents the features utilized in this study, briefly describing each feature and their respective deviations. The deviations correspond to the four feature types discussed: sum-based, average-based, slope-based, and projection-based. While a comprehensive understanding of these features is not essential for comprehending the paper, their inclusion is crucial for ensuring the replicability of this study. Prior works influenced the selection of these features in NHL pre-game prediction, particularly the studies conducted by Pischedda (2014) and Weissbock et al. (2013). However, we also considered inputs from the public domain, considering numerous NHL advanced analytics sites, including MoneyPuck (Tanner, n.d) and Natural Stat Trick (Natural stat trick, n.d). These sites offer performance indicators that have not yet been widely adopted in the academic space but have demonstrated their effectiveness as predictors of game outcomes, such as expected goals.

Table 3. Features used by the ML models along with a description and the deviations from the base calculation of the feature.

Feature	Description	Deviations
Wins	The number of times a team won the game	Sum-based
Losses	The number of times a team lost the game.	Sum-based
Goals (For & Against)	When the puck crosses the line, enters the net, and is awarded to the team as a goal.	Sum-based, Average-based, Slope-based, Projection-based
Goals 5v5 (For & Against)	When a Goal is scored by a given team in a 5v5 situation.	Sum-based, Average-based, Slope-based, Projection-based
Goals 5v5 Close (For & Against)	When a Goal is scored by a given team in a 5v5 close situation.	Sum-based, Average-based, Slope-based, Projection-based
Shots (For & Against)	When the puck is sent toward the opposing net by a player on a given team. This has to result in the goalie stopping the puck or a goal being scored, misses do not count.	Sum-based, Average-based, Slope-based, Projection-based
CORSI	The shot attempt differential for a team. That is the number of attempted shots for, minus the attempted shots against.	Sum-based, Average-based
CORSI%	The percentage of the shot attempts a given team had.	Slope-based, Average-based
Fenwick %	The shot attempt differential for a team represented as a percentage. However, this calculation does not include shots that were blocked.	
CORSI 5v5	The shot attempt differential in 5v5 scenarios.	Sum-based, Average-based
CORSI 5v5 Close	The shot attempt differential in 5v5 close scenarios.	Sum-based, Average-based
Face-offs (For & Against)	When two players meet to “face-off” for the puck before the start of play. The puck is dropped and whoever moves it to a teammate first is considered to have won the faceoff. This is the number of face-offs won.	Slope-based, Projection-based
Face-off Percentage	The percentage of face-offs won.	Percentage-based
Hits (For & Against)	When a player body checks an opposing player.	Sum-based, Average-based, Slope-based, Projection-based
Penalty Minutes (For & Against)	The number of minutes a team had a player in the penalty box, due to an infraction performed on the ice.	Sum-based, Average-based, Slope-based, Projection-based
Blocks (For & Against)	When an opposing player takes a shot which is in turn blocked by a player on the team before the puck can reach the net.	Sum-based, Average-based, Slope-based, Projection-based

Table 3. continue ...

Giveaways (For & Against)	When a player gives the puck away to an opposing player.	Sum-based, Average-based, Slope-based, Projection-based
Takeaways (For & Against)	When a player takes the puck away from an opposing player.	Sum-based, Average-based, Slope-based, Projection-based
xG (For & Against)	The expected number of goals the team should have scored based on a machine learning algorithm that takes into account the quality of each shot and trains on historical shot data.	Sum-based, Average-based, Slope-based, Projection-based
xG 5v5 (For & Against)	The xG for a team in 5v5 scenarios.	Sum-based, Average-based, Slope-based, Projection-based
xG 5v5 Close (For & Against)	The xG for a team in 5v5 close scenarios.	Sum-based, Average-based, Slope-based, Projection-based
Power Play Opportunities	A power play is when a team has an extra player on the ice due to a penalty that is being served by the opposing team.	Slope-based, Projection-based
Power Play Goals	The number of goals a team scores on the power play	Slope-based, Projection-based
Power Play Percentage	Power play percentage represents the percentage of power plays that result in a goal for the team.	Percentage-based
Penalty Kill Opportunities	A penalty kill is when a team has fewer players on the ice due to a penalty that is being served by the team.	Slope-based, Projection-based
Penalty Kill Goals	The number of goals that are scored on a team during the penalty kill.	Slope-based, Projection-based
Penalty Kill Percentage	Penalty kill percentage is the percentage of time that these situations do not result in a goal for the opposing team.	Percentage-based
Shooting Percentage	The percentage of shots from a team that results in a goal.	Percentage-based
Save Percentage	The percentage of shots that are taken against a team that results in the goalie stopping the puck.	Percentage-based
PDO	A team’s shooting percentage plus their save percentage. Sometimes referred to as a “luck” statistic.	Percentage-based
Turnover to Giveaway Ratio (For & Against)	The ratio of turnovers to giveaways.	Slope-based, Projection-based

Two momentum-based features in our analysis deviate from the slope and projection-based approach. These features include the team's current streak, which indicates the number of consecutive games they have won or lost. The streak is represented as an integer, where a positive value indicates a winning streak and a negative value indicates a losing streak. Additionally, we incorporate a rest calculation feature that captures the number of days between a team's most recent game and the upcoming game for which we predict the outcome.

It is worth noting how the features were represented before their use by the machine learning models. Pischedda (2014) demonstrated that their models exhibited greater predictive power when the features were represented as the home team's value minus the away team's value. We obtained similar results upon testing it with our data, leading us to adopt this approach in presenting the features to the machine learning models. For instance, if the home team was projected to have 30 shots in a game and the away team was projected to have 25 shots, we would calculate the difference (30 - 25 = 5) and store the value of five in the projected shots column for that specific game.

Model Evaluation

It is important to note that in this section, the term "model" refers to a specific combination of a feature set (momentum-based, frequency-based, or combined), an interval of games (3-7), and a machine learning algorithm (Logistic Regression, Random Forest, Support Vector Machine). The models were evaluated using a moving window approach. The games were sorted chronologically, and we trained the model on a group of games and tested it on the subsequent group of games, starting from the game immediately after the last game in the training set. Specifically, we used a training size of 2460 games and a testing size of 1230 games, resulting in a train/test moving window of 3690 games. This setup allowed for approximately two seasons of training data and one season of testing data, as each NHL season consists of roughly 1230 games. We observed that using less than two seasons of training data yielded inconsistent results during the experimental design. Therefore, we ensured that our moving window always included two seasons of data for training purposes. We created 20 train-test sets and utilized these sets to evaluate our models. The objective was to predict the winner of the games in the testing set while fitting the model to the training set. To measure the performance of the models, we employed the accuracy metric from the Sklearn module (Pedregosa et al., 2011). This metric was applied to each of the 20 train-test splits and calculated the percentage of correctly identified game outcomes out of the 1230 games in the testing set. The accuracy metric is computed by dividing the number of correctly identified outcomes by 1230, representing the total number of games in the respective testing set.

We consider the 20 individual results obtained from the moving window approach to indicate each model's performance. Additionally, this number of results meets the recommended minimum requirement for conducting the Wilcoxon signed-rank test (Wilcoxon, 1945), a nonparametric test used to assess the statistical significance of differences in result distributions. By performing this test, we can determine whether the observed differences are statistically significant. Specifically, if the test yields a p-value less than 0.05, we can reject the null hypothesis, indicating that the two sets of results are not drawn from the same distribution. The Wilcoxon signed-rank test was conducted using the SciPy Python module (Virtanen et al., 2020). For our machine learning algorithms, we selected models from the Sklearn library (Pedregosa et al., 2011) to evaluate our pipeline. The chosen algorithms include logistic regression, random forest (Ho, 1995), and support vector machines (SVM), also known as SVM (Cortes & Vapnik, 1995). These algorithms were employed to assess the performance of our models. The models mostly relied on their default hyperparameters, except for random forest, which had its estimators set to 500, and the seed set to 1415 (the first four digits after the decimal in Pi) to ensure the results were replicable. Logistic regression also had its max iterations set to 10000 to ensure convergence was achieved.

Model Comparison

To compare the performance of our models, we utilized box plots for visualizing the data. These box plots were created with the use of the Matplotlib module in Python (Hunter, 2007). Each box plot represents the distribution of 20 individual results obtained by the moving window approach previously discussed. All calculations regarding minimums, maximums, medians, and interquartile range are performed by the Matplotlib module at the time the plot is created. We have also included the mean in each box plot denoted by a green plus sign. Each section of our methodology corresponds to a specific machine learning algorithm: logistic regression, random forest, and support vector machines (SVM). For each interval of games ranging from three to seven, we examined and compared the accuracy distributions achieved by the momentum-based, frequency-based, and combined feature sets. Additionally, we provided a table displaying the p-values and test statistics resulting from the Wilcoxon signed-rank test. In the p-value table, statistical significance is indicated by the presence of a star (*).

Entire Data Pipeline

The methodology employed in our study is summarized in Figure 1. The raw event data is initially processed to extract and organize game event summaries, enabling access to previous games of both home and away teams. Subsequently, an interval-based procedure is employed to collect data from the previous "n" games for both home and away teams. This interval-based data generates distinct feature sets such as frequency-based, momentum-based, and all features, serving as input feature vectors for the machine learning models. The models are then trained using the same algorithm but with different feature sets, and their performances are compared in the final stage of model evaluation and comparison.

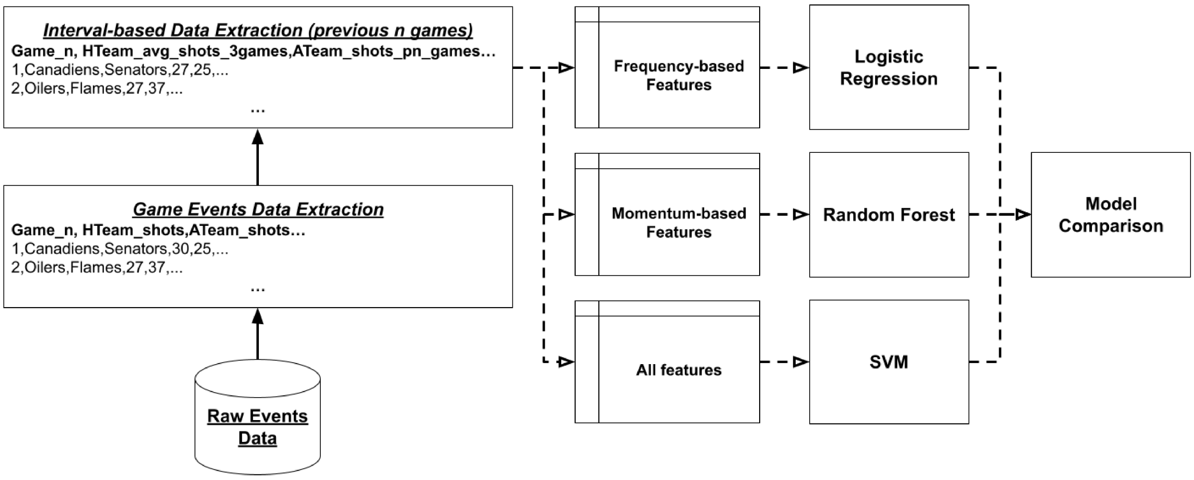


Figure 1. Data pipeline for our methodology.

Results

Logistic Regression

Figure 2 provides an overview of our findings, indicating that momentum-based features outperformed both the frequency-based and combined feature sets when utilizing a three-game interval. The statistical analysis using the Wilcoxon signed-rank test, as presented in Table 4, confirms the significant differences in performance between these feature sets at this game interval. At an interval of four games, all feature sets have very similar performance. However, as we extend the interval to five or six games, the advantage momentum-based features initially saw quickly diminishes, and they exhibit worse performance than the frequency-based features; this difference in performance is also confirmed as statistically significant in Table 4. Interestingly, momentum-based features perform similarly to frequency-based features when employing a seven-game interval.

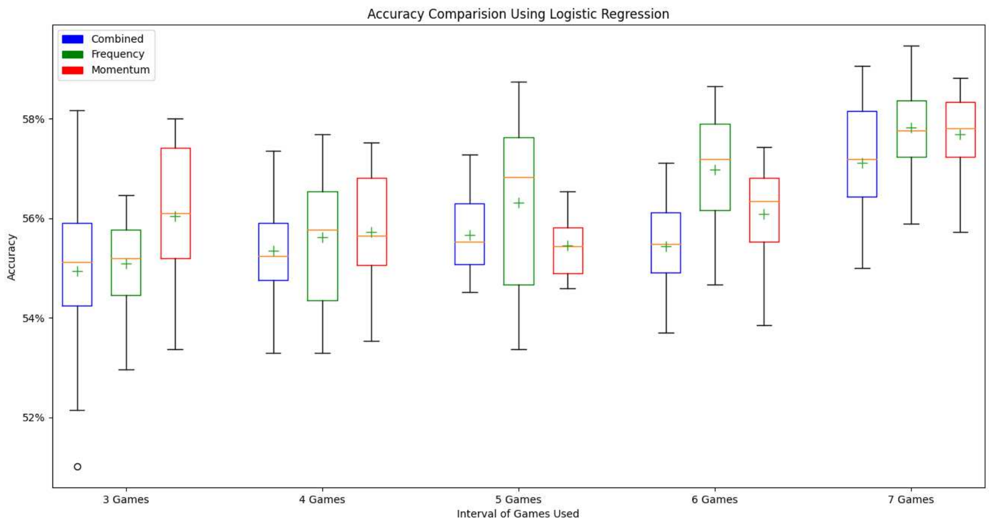


Figure 2. Comparison of the models using the logistic regression algorithm and game intervals 3 through 7.

Table 4. The p-values and statistical value of the test achieved when comparing the distributions of two given feature sets for a given interval of games when using logistic regression. Statistically significant p values are denoted by a star *. The statistical value represents the sum of the ranks of the differences above or below zero, whichever is smaller.

Models	**Interval of Games**				
	3	**4**	**5**	**6**	**7**
Frequency – Momentum p-value	0.00315*	0.760	0.0313*	0.00181*	0.687
Momentum – Combined p-value	< 0.001*	0.0415*	0.0867	0.00700*	0.0140*
Combined – Frequency p-value	0.717	0.368	0.105	< 0.001*	0.00253*
Frequency – Momentum	28.5	78.5	41.5	17.5	85

Statistical value					
Momentum – Combined Statistical value	15	33.5	52.5	28	34
Combined – Frequency Statistical value	86	57.5	60.5	16	20

Random Forest

Figure 3 illustrates that the frequency-based and momentum-based feature sets exhibit similar performance across intervals of three to five games. However, it is essential to note that this does not diminish the value of momentum-based features in this dataset, as the combination of frequency-based and momentum-based features yields the best results across intervals of three to six games. This observation may be attributed to the potency of the random forest algorithm, which has the ability to assign weights to features and potentially prioritize a subset of the combined feature set for classification purposes. The statistically significant differences between the distributions generated by the combined feature set and the momentum-based or frequency-based feature sets are evident in all game interval feature set comparisons but two, as demonstrated in Table 5.

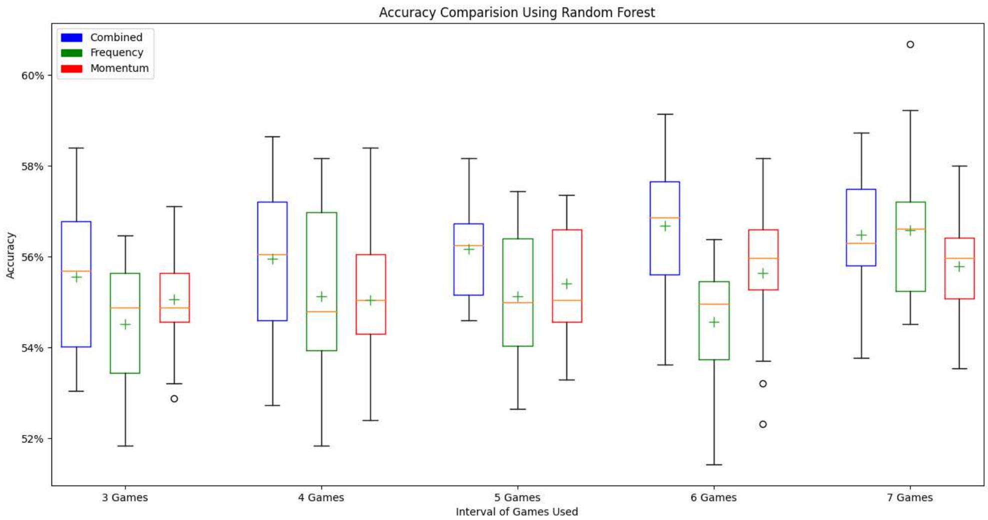


Figure 3. Comparison of the models using the random forest algorithm and game intervals 3 through 7.

Table 5. The p-values and statistical value of the test achieved when comparing the distributions of two given feature sets for a given interval of games when using random forest. Statistically significant p values are denoted by a star *. The statistical value represents the sum of the ranks of the differences above or below zero, whichever is smaller.

Models	**Interval of Games**				
	3	**4**	**5**	**6**	**7**
Frequency – Momentum p-value	0.105	0.840	0.234	0.00137*	0.0153*

Momentum – Combined p-value	0.112	0.00486*	0.0119*	0.00365*	0.0107*
Combined – Frequency p-value	< 0.001*	0.00639*	0.00102*	< 0.001*	0.985
Frequency – Momentum Statistical value	61	90	65.5	15.5	41.5
Momentum – Combined Statistical value	55.5	32.5	32.5	30.5	38.5
Combined – Frequency Statistical value	3	34.5	22	0	104.5

SVM

Figure 4 shows that the momentum-based and combined feature sets deliver the best performance when utilizing a three-game interval and that the performance increase seen over the frequency-based features is statistically significant for both the momentum-based and combined feature sets. The performance at the three-game interval is also the best performance out of all the intervals. However, it should be noted that SVM exhibits inconsistent performance, as indicated by the relatively large interquartile ranges depicted in Figure 4.

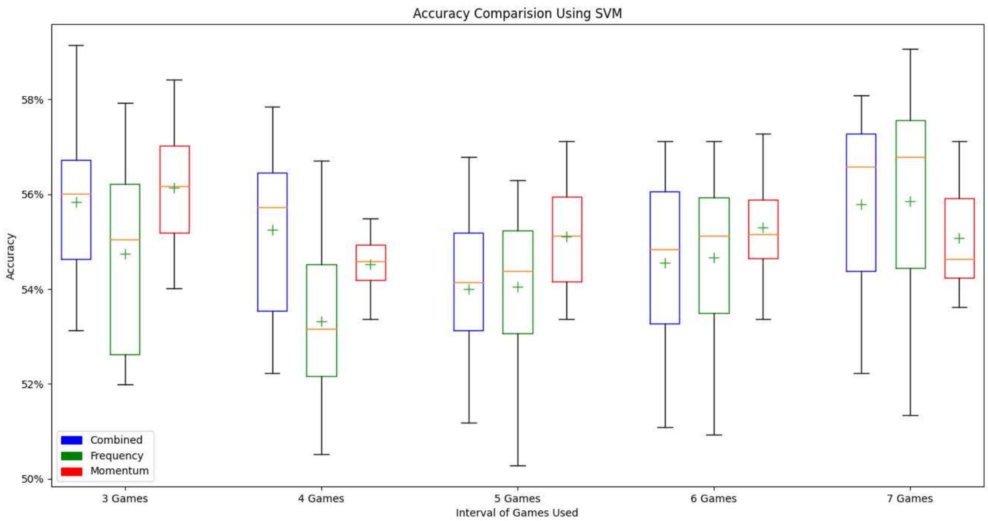


Figure 4. Comparison of the models using the SVM algorithm and game intervals 3 through 7.

Table 6. The p-values and statistical value of the test achieved when comparing the distributions of two given feature sets for a given interval of games when using SVM. Statistically significant p values are denoted by a star *. The statistical value represents the sum of the ranks of the differences above or below zero, whichever is smaller.

Models	Interval of Games				
	3	4	5	6	7
Frequency – Momentum p-value	< 0.001*	0.0282*	0.0406*	0.122	0.0583

Momentum – Combined p-value	0.105	0.143	0.0279*	0.123	0.0362*
Combined – Frequency p-value	< 0.001*	< 0.001*	0.856	0.304	0.648
Frequency – Momentum Statistical value	11	40.5	38.5	50	54.5
Momentum – Combined Statistical value	60.5	65	35	63	43
Combined – Frequency Statistical value	8	0	64.5	69.5	92

Our results from logistic regression show it is the most reliable predictor of NHL games among the algorithms we tested, and it should be considered and compared in future research on NHL game prediction. Additionally, our findings indicate that momentum-based features have greater predictive power than traditional frequency-based features in short intervals of three games. This observation did not come as a surprise. While frequency-based features have been established and proven effective in NHL prediction over the years, we believed that momentum-based features would outperform them in very short game intervals due to momentum's potential marginal and temporary effects.

When examining our results from the random forest algorithm, we observe that the combination of all features consistently produces the highest predictive power across all but one interval. This suggests that momentum-based features are not a universal solution for pre-game prediction in the NHL. Instead, it highlights the need for further research into these momentum-based performance indicators to identify the ones that possess the most significant predictive power. This can contribute to developing more comprehensive pre-game prediction models that consider a team's performance over the entire season while incorporating momentum-based indicators from recent games.

Our results from SVM show that the momentum-based and combined feature sets outperformed the frequency-based feature set at a three-game interval and that this was the best performance seen over all intervals for SVM. We should note that the benefit seen here when using the three-game interval and momentum-based features is similar to what was seen in logistic regression. Aside from this, we acknowledge that our SVM model did not yield meaningful results as the accuracy never seems to reach an average of 56% again, which aligns with our previous findings in which SVM would always choose the home team to win (Noel, 2021). This suggests that SVM may not be a suitable pre-game predictor for NHL games without appropriate feature selection and hyperparameter tuning, as parameters such as the regularization parameter (C) can affect the model's ability to generalize (Cortes & Vapnik, 1995).

It is important to contextualize these results in the area of pre-game prediction in the NHL. If we consider the lower bound of NHL prediction as the percentage of times the home team wins the game, which is 55.4% based on our dataset after data manipulation, and we observe the upper bound as the theoretical 62% ceiling that Weissbock and Inkpen (2014) purposed. We observe a variance of only 6.6% between the upper and lower bounds. Therefore, even slight improvements in predictive power are precious in the context of NHL pre-game prediction, given the relatively narrow range of variation between the upper and lower bounds. A potential shortcoming of this range is that the 62% ceiling is that Weissbock and Inkpen (2014) derived the ceiling using data from the 2005 through 2011 seasons while our approach uses data from the 2011 season to the 2020 season. Therefore, we are assuming that the league has not changed dramatically enough since 2011 to lead to a massive shift in the predictability of outcomes. However, we would expect that in seasons that have more parity (skill difference between the best and worst teams is small), the ceiling will likely be lower, while in seasons that have less parity (skill difference between the best and worst teams is large) the ceiling will likely be higher as in the past researchers have related parity and predictability (Ben-Naim et al., 2006).

In saying this, our goal in this work was to establish a new framework for ice hockey performance indicators and examine the potential effects of momentum in NHL games. We hypothesized that these momentum-based performance indicators would be able to capture the effects of momentum, whereas previous studies have not due to the lesser dependence on outcomes. While our primary goal was not to create the most powerful model, we achieved an average accuracy of 57.8% using logistic regression, a seven-game interval, and frequency-based features, as shown in Figure 2. Our second most powerful model achieved an average of 57.7% using logistic regression, a seven-game interval, and momentum-based features, as shown in Figure 2. These averages fall within a four to five-percent margin of the proposed 62% ceiling by Weissbock and Inkpen (2014). However, it is essential to note that while their proposal considered using all team games to measure team quality, our approach utilized only seven games. In comparison, an NHL season consists of 82 games, meaning our three to seven-game intervals represent a small fraction of a team's season.

The implications of momentum-based features having greater predictive power than frequency-based features at a small interval of three games, as seen in our logistic regression and SVM results, extend to sports science and analytics. Firstly, our results suggest that momentum does have some influence on the outcome of NHL games with smaller samples of games. However, its impact may not be as substantial as anecdotal reports from players and coaches suggest. Nonetheless, momentum-based features can be valuable assets, particularly for coaches. In a league characterized by its relative randomness, coaches should focus on factors they can control rather than those they cannot. These momentum-based performance indicators can help coaches identify areas in opponents' games that can be exploited in upcoming matches. For instance, if the momentum-based projection-based feature indicates that a forthcoming opponent is likely to give up a high number of power plays, it would be wise for the coach to emphasize power play strategy with their team before that game.

Our findings from random forest have broader implications for more complex and robust models like neural networks or ensembles. Such models can assign higher weights to specific features, thereby increasing predictive power (Ray, 2019). This, in turn, could have implications for sports betting by creating models that can potentially predict short-term success more accurately than existing models. However, we must exercise caution in attempting to outperform sports betting books at their own game. It is essential to recognize that bettors often succumb to the bias of favoring teams with apparent momentum and frequently lose due to this bias (Ötting et al., 2022).

While our study primarily examines team momentum rather than individual momentum, it is plausible that momentum-based features may also possess predictive power for individual players. This could impact a coach's player selection, determining who plays and who sits, especially when faced with players who exhibit similar frequency-based performance indicators. Momentum-based indicators can serve as tie-breakers in such cases. Exploring momentum-based performance indicators for individual players is an avenue we intend to explore in future research. It would be valuable to compare the results of such a study to the work of Pelechrinis and Winston (2022), who found that certain NBA players exhibited the effects of momentum more than others.

The contributions of our work are unique in that we have introduced a novel framework for performance indicators that quantifies momentum across various aspects of team performance, going beyond the traditional binary distinction of winning and losing. Although these performance indicators may only lead to marginal improvements in the predictive power of machine learning models, we have provided ample citations to support the notion that NHL games are subject to a significant amount of randomness. Therefore, it becomes crucial for teams to focus on maximizing their chances of success, understanding that losing is still possible even when making decisions that maximize the likelihood of a positive outcome. This idea can be backed up by the work of Schulte et al. (2017), who developed a Markov game model for evaluating actions in ice hockey. They found a strong correlation between the overall value of team actions and the likelihood of a team winning the game. This entails that the team who made the best decisions won most of the time but not all the time. This can also be seen in the work by Sprigings and Toumi (2015), who showed that past CORSI percentage and expected goal (xG) percentage were better predictors of future goal percentage than past goal percentage. This means that shot quality and shot attempt differential are better predictors of future goal scoring than past goal scoring, reaffirming the idea that there is a level of randomness present in the NHL and that teams should focus on the process (i.e., generating quality shot attempts) rather than the result (i.e., scoring goals). Therefore, our perspective on ice hockey should align with how one would approach a game of probability, such as blackjack, where the goal is to make informed choices to optimize the odds of winning.

While our results demonstrate that momentum has an impact on the outcome of NHL games, we do not view it as a contradiction to previous studies that found no evidence of momentum in ice hockey, such as the works by Kniffin and Mihalek (2014) or Steeger et al. (2021). Those studies primarily aimed to determine if teams experience winning or losing streaks solely due to momentum. In contrast, our findings suggest that momentum plays a role in game outcomes but is a small aspect of it and is not the sole determinant. Ice hockey is a complex sport with various factors at play, and our momentum-based performance indicators provide an additional perspective in understanding why teams win or lose.

However, our work has certain limitations. Our features measure momentum based on previous game performance indicators but evaluating our models' success primarily relied on predicting game outcomes accurately. While at the same time, we have acknowledged several times that momentum should be studied in way of underlying performance indicators and not just in streaks of wins or losses as teams should focus on maximizing the chance of success. In future research, we aim to develop models that focus on predicting other measures of team performance, such as the xG differential, which could provide insights into a team's share of scoring chances during a game. Exploring xG differentials in ice hockey, a statistic also popular in European football/soccer could allow for comparisons between the two sports.

Another aspect of our work that can be seen as a limitation is our use of a linear line of best fit in our slope-based and projection-based features. This is because short intervals of games may exhibit random fluctuations in a team's performance. As a result, the linear line of best fit may not always capture these fluctuations accurately, and it may not provide the most comprehensive representation of a team's recent performance trends, especially when using a minimal interval of games, such as three. However, despite this limitation, we utilized the linear line of best fit as it is the most intuitive and straightforward approach to quantify a team's recent performances. In future investigations, we are interested in exploring alternative methods for determining the line of best fit. By experimenting with different approaches, we aim to enhance our understanding and capture the nuanced dynamics of a team's performance over time.

Additionally, it is essential to note that the data provided by the NHL may contain some imperfections, despite our efforts to clean and ensure accuracy. Challenges arise when determining powerplay opportunities, mainly when simultaneous penalties occur. Scorekeeper bias has also been observed to affect NHL statistics, leading to potential inaccuracies in shot placements and shot credits (Thomas, 2015). To foster the growth of analytics in the NHL, public access to player and puck-tracking data would be necessary for transparency and trust in the provided data, as was previously recommended by Nandakumar and Jensen (2019).

Looking ahead, we believe this work could serve as a foundation for developing a framework that investigates the effects of momentum on individual teams, players, or even across different sports. By delving deeper into momentum-based analysis, we can gain further insights and expand our understanding of its implications in sports.

Conclusion

Our work sheds light on the impact of momentum in ice hockey and serves as a foundation for further research in this area by engineering features that show momentum can be approached from a trend of play perspective rather than looking at the dependence or independence of sequential outcomes.

Our work contributes to developing a more formal definition of momentum and offers momentum-based features that can be utilized to construct models for predicting pre-game outcomes in the NHL and other sports. We offer two main findings from this work: momentum-based features offer performance increases over frequency-based features at a small three-game interval with logistic regression or SVM, and random forest tends to see a performance increase when using the combined feature set. Overall, our findings imply momentum has a potential impact on game outcomes be it a rather small one. Of our two main findings, we believe the one regarding random forests to be the most promising. A data pipeline that uses sufficient feature selection, hyperparameter tuning, and a random forest model may better understand the finer details of pre-game prediction in ice hockey. Such a model could better pre-game prediction for anyone attempting to predict NHL game outcomes, from academics to teams to bookmakers setting initial odds. This highlights the need to strike a balance between the overestimation of momentum's effect by athletes, coaches, and fans and the potential underestimation of its significance by the analytics community. We can deepen our understanding of momentum's effects by conducting more research, particularly in ice hockey, where academic studies are limited.

Continued exploration of momentum in sports analytics can lead to valuable insights that benefit athletes, coaches, and fans alike. It is an area that warrants further investigation and can contribute to advancing our knowledge of the intricacies of sporting events.

References

American Gaming Association (2023). *2022 Commercial Gaming Revenue Tops $60B, Breaking Annual Record for Second Consecutive Year* [Press release]. Retrieved from https://www.americangaming.org/new/2022-commercial-gaming-revenue-tops-60b-breaking-annual-record-for-second-consecutive-year/

Arkes, J., & Martinez, J. (2011). Finally, evidence for a momentum effect in the NBA. *Journal of Quantitative Analysis in Sports, 7*(3).

Ben-Naim, E., Vazquez, F., & Redner, S. (2006). Parity and predictability of competitions. *Journal of Quantitative Analysis in Sports, 2(4).*

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Gilbert, D. E., & Wells, M. T. (2019). Ludometrics: luck, and how to measure it. *Journal of Quantitative Analysis in Sports, 15*(3), 225-237.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, *17*(3), 295-314.

Goldman, S. (2022, June 24). Analytics in the NHL: Where it's at and where it's going next. Retrieved May 5, 2023, from https://theathletic.com/3381342/2022/06/24/nhl-analytics-stats/

Gu, W., Foster, K., Shang, J., & Wei, L. (2019). A game-predicting expert system using big data and machine learning. Expert Systems with Applications, 130, 293-305.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).

Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. International Journal of Forecasting, 35(2), 783-796.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.

Noel, J (2021), *Evaluating Team's Momentum in Hockey Games.* University of Newfoundland & Labrador, St. John's, Canada. Retrieved from https://mun.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma993561131402511&context=L&vid=01MUN_INST:01MUN&lang=en&search_scope=MyInst_and_CI&adaptor=Local%20Search%20Engine&tab=Everything&query=any,contains,Jordan%20Noel&offset=0

Kari, D. (2020). Understanding how random chance affects the outcome of an ice hockey game.

Kniffin, K. M., & Mihalek, V. (2014). Within-series momentum in hockey: No returns for running up the score. *Economics Letters, 122*(3), 400-402.

Lopez, M. J., Matthews, G. J., & Baumer, B. S. (2018). How often does the best team win? A unified approach to understanding randomness in North American sport. *The Annals of Applied Statistics, 12*(4), 2483-2516.

Mondello, M., & Kamke, C. (2014). The introduction and application of sports analytics in professional sport organizations. Journal of Applied Sport Management, 6(2), 11.

Nandakumar, N., & Jensen, S. T. (2019). Historical perspectives and current directions in hockey analytics. Annual review of statistics and its application, 6, 19-36.

Natural stat trick. (n.d.). Retrieved May 3, 2023, from https://www.naturalstattrick.com/

Osborne, J. (2020). Retrieved from https://www.oddsshark.com/sports-betting/which-sport-do-betting-underdogs-win-most-often

Ötting, M., Deutscher, C., Singleton, C., & De Angelis, L. (2022). Gambling on Momentum. arXiv preprint arXiv:2211.06052.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. & others (2011). Scikit-learn: Machine

Shomer, H. (2019). Hockey_Scraper. Retrieved February 27, 2023, from https://hockey-scraper.readthedocs.io/en/latest/index.html

Sprigings, D., & Toumi, A. (2015, October 1st). *Expected goals are a better predictor of future scoring than Corsi, goals*. Hockey Graphs. https://hockey-graphs.com/2015/10/01/expected-goals-are-a-better-predictor-of-future-scoring-than-corsi-goals/

Steeger, G. M., Dulin, J. L., & Gonzalez, G. O. (2021). Winning and losing streaks in the National Hockey League: are teams experiencing momentum or are games a sequence of random events?. *Journal of Quantitative Analysis in Sports, 17*(3), 155-170.

Tanner, P. (n.d.). Moneypuck.com - NHL analytics, playoff odds, Power Rankings, player stats. Retrieved February 27, 2023, from https://moneypuck.com/

Thomas, A. C. (2015). A Statistician reads the sports pages: Pucksberry: What we can learn from hexagonal plots of NHL data. *Chance*, *28*(1), 43-50.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods, 17*(3), 261-272.

Weissbock, J., & Inkpen, D. (2014). Combining textual pre-game reports and statistical data for predicting success in the national hockey league. In *Advances in Artificial Intelligence: 27th Canadian Conference on Artificial Intelligence, Canadian AI 2014, Montréal, QC, Canada, May 6-9, 2014. Proceedings 27* (pp. 251-262). Springer International Publishing.

Weissbock, J., Viktor, H. L., & Inkpen, D. (2013, September). Use of Performance Metrics to Forecast Success in the National Hockey League. In *MLSA@ PKDD/ECML* (pp. 39-48).

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.