A New Approach to Momentum: A Novel Framework to Evaluate Momentum in Sports

by

© J. T. P. Noel

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science

Supervisors: Dr. Amilcar Soares and Dr. Vinicius P. da Fonseca

Department of Computer Science

Memorial University of Newfoundland

December 2023

St. John's Newfoundland

Abstract

Momentum often faces skepticism as a discredited phenomenon in sports. While some contradictory evidence has been found, we provide new insights by quantifying this phenomenon. Momentum literature often relies on proving the dependence or independence of sequential outcomes. However, we argue that this approach is not appropriate due to the large body of literature showing that sports are highly subject to randomness. If sports are subject to randomness, we should focus on what leads to winning rather than not winning itself. Here, we engineer momentum-based features that quantify a team's linear trend of play in several underlying performance indicators and compare the predictive power of these features to more traditional frequency-based features when only using a small sample of recent games to assess team quality. We developed a complete data pipeline that allows us to compare the effects of momentum on multiple sports. We found evidence of momentum in the NHL and the five major European football/soccer leagues; however, we could not find evidence of momentum in the NBA. The differences between these sports indicate that momentum could be a sport-specific phenomenon. We also found that the combination of momentum-based and frequency-based feature sets, along with more powerful machine learning techniques such as random forest, led to very promising results. In the future, we believe that by combining these two feature sets with proper hyperparameter tuning and feature selection, better pre-game prediction models can be created that accurately capture both short-term and long-term quality of play.

Acknowledgements

I would like to thank Dr. Amilcar Soares and Dr. Vinicius P. da Fonseca for their supervision. Their expertise was invaluable to the completion of this thesis and my personal increase in knowledge.

I would like to also thank my family and friends for their continued support throughout my program.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, we have seen a paradigm shift in the world of sports relating to how teams treat the available data [1]. Professional teams that traditionally relied only on their scouts' expertise and their executives' business know-how, more often turn to statisticians, computer scientists, and data scientists for help making daily decisions. This is mainly due to the growth of sports analytics in recent years. Insightful analysis of in-game data allows the upper management of sports teams to make informed decisions untainted by human bias. By eliminating the human element from the equation, decisions can be made based on empiricism, allowing decision-makers to create more successful teams and increase the revenues and profits seen by the team owners [2].

While sports analytics may be based on eliminating human bias, there is a plethora of phenomena that players and coaches provide anecdotal evidence on that are worth exploring empirically. One such phenomenon that is very common in sports is momentum. In most cases, momentum refers to the idea that consecutive positive or negative outcomes will lead to a greater chance of positive or negative outcomes in the near future.

While momentum has been a long-held belief by coaches, players, and fans [3], a large segment of the academic work surrounding the topic refutes its existence [4, 5, 6, 7] or argues its effects are marginal [8]. However, in recent years, scattered amongst the refutations is a growing body of work that shows evidence of momentum in sports [9, 10, 11, 12, 13]. Most of the academic work regarding momentum has focused on the dependence or the independence of sequential outcomes, using statistical run tests to determine whether or not repeated success or failure are dependent events.

In this work, our goal is to neither confirm nor refute these claims but to attempt to shift the focus of momentum research away from the dependence on sequential outcomes. We believe that due to demonstrable randomness in sports [14], momentum is better captured with the trends of several in-game performance indicators. Then, using these indicators and machine learning (ML) models, we can examine how momentum may affect sports.

In the rest of this Chapter, we first state the problem we are trying to solve in Section 1.1, then outline our research goals and question in Section 1.2. Our unique contributions to the momentum and sports analytics field are described in Section 1.3. In Section 1.4, we provide a co-authorship statement for the two papers under review that are the major outcomes of this thesis. Finally, in Section 1.5, we outline the rest of this thesis.

1.1 Problem Formulation

As previously stated, momentum is often a controversial topic in sports analytics. However, we believe that it currently may be approached improperly. The ultimate goal of sports analytics is to make informed decisions that maximize the likelihood of a positive outcome. It is, however, essential to understand that just because the probability of a positive outcome is maximized, victory is not guaranteed. It is always possible to perform actions that result in winning most of the time and still lose a match or a championship. This can be backed up by the literature on the randomness of outcomes in sports, particularly ice hockey [14]. For this reason, we believe that measuring momentum using methods that work with the dependence or independence of sequential events, such as the Wald-Wolfowitz test for randomness, is not an appropriate approach. We believe momentum is better captured by the increase or decrease in the overall quality of play and that we should move away from the binary view of winning and losing when addressing it. While several papers find evidence of momentum, it is also usually viewed in a binary fashion (i.e., is this team affected by momentum or not) and lacks a method of quantification that shows the degree to which momentum affects them [5, 9, 4, 15, 6, 8].

1.2 Research Goals and Questions

Instead of approaching momentum with a binary view, we believe momentum should be quantified by the increase or decrease of underlying performance indicators over a short period of games. Therefore, the goals of this work are as follows:

- Creating a method of quantification for momentum in sports based on the increase or decrease in the overall quality of play.
- Using this quantification method along with Machine Learning (ML) models to determine the predictive power of momentum and compare that to the predictive power of traditional frequency-based features.

Therefore, this thesis is attempting to answer the following research questions:

- How can momentum be quantified in terms of performance indicators?
- Does momentum affect the outcome of sports game outcomes?

1.3 Contributions

Our work advances research in sports analytics and momentum through the following contributions.

- Using a team's recent games, we engineer momentum-based features that use individual performance indicators from each game and the passing of days between these games to develop a linear line of best fit. This allows us to determine the play trend over a recent sample of games. These features, therefore, quantify momentum and allow us to measure a team's momentum pre-game using the slope of such a line of best fit, which in turn also allows us to project the team's performance in the upcoming game should these trends continue. A complete overview of our final proposed pipeline is shown in Figure 4.1 (Chapter 4).

- We compare the predictive power of these proposed momentum-based features against the predictive power of traditional frequency-based features. We found evidence that momentum affects outcomes in ice hockey (Section 3.3 of Chapter 3 and Section 4.4 of Chapter 4) and soccer (Section 4.4 of Chapter 4) but does not affect basketball (Section 4.4 of Chapter 4).

- We find that ensemble-based tree models, such as random forest, favor models that include both momentum-based and frequency-based features (Section 4.5 of Chapter 4). This indicates that more robust models may benefit from combining the feature sets and, consequently, create better future pre-game prediction models.

All these contributions and the results achieved by this thesis are detailed in Chapters 3 and 4. We first developed a prototype framework focused mainly on Hockey game events. This work was titled *The Use of Momentum-Inspired Features in Pre-Game Prediction Models for the Sport of Ice Hockey.* was published in the *International Journal of Computer Science in Sports (IJCSS)* [16]. The published version of this paper can be seen in Chapter 3. Later, we generalized the framework for multiple sports, and the results were included in a manuscript titled *A Comprehensive Data Pipeline for Comparing the Effects of Momentum on Sports Leagues.* which was published with *MDPI Data* [17]. The published version of this paper can be seen in Chapter 4.

1.4 Co-authorship Statement

Chapter 4 is our first manuscript published as **Noel, J.T.P., Prado da Fonseca, V., Soares, A. (2024). The Use of Momentum-Inspired Features in Pre-Game Prediction Models for the Sport of Ice Hockey. *International Journal of Computer Science in Sport (IJCSS), 23*(1), 1-21.** `https://doi.org/10.2478/ijcss-2024-0001`. [16]. This was our first work, which looked at introducing the momentum-based features and how they were engineered and then applied to an NHL dataset. I performed the experiments and wrote all the required code for the manuscript. My supervisors, Dr. A. Soares and Dr. V. Prado da Fonseca guided me through design concepts, statistical tests, and ML techniques. They both provided great feedback and editorial input for the manuscript before its initial submission and in the following peer review rounds.

Chapter 5 is our second manuscript which is published as **Noel, J.T.P., Prado da Fonseca, V., Soares, A. (2024). A Comprehensive Data Pipeline for Comparing the Effects of Momentum on Sports Leagues. *MDPI Data 9*(2), 29.** `https://doi.org/10.3390/data9020029`. [17]. I designed the basic concept of the work with constant feedback from my supervisors, Dr. A. Soares and Dr. V. Prado da Fonseca. I coded and performed all the experiments. I performed the data analysis but with substantial feedback from my co-supervisors. I drafted and prepared the manuscript with subsequent editorial input from my two supervisors.

1.5 Thesis Outline

This work is a thesis by articles and, therefore, is structured as follows. In Chapter 2, we provide an in-depth literature review and show how our work's unique aspects advance the literature in sports analytics, more specifically in the field of momentum research. In Chapter 3, we present our first published work on momentum which explores the use of momentum-based features in the NHL [16]. Next, in Chapter 4 we present a second published work on momentum, this time exploring different sports and outlining a complete data pipeline [17]. Finally, in Chapter 5, we conclude this thesis by articles and provide a list of future works that can be performed.

Chapter 2

Literature Review

The concepts of momentum and sports are, in many ways, inseparable. While there has been a considerable amount of research in the field, the amount of work that pertains to ML is sparse. It could be argued that the main two field research areas are psychological and statistical. While this paper is based on ML, we will use all the information we can find, including data from other disciplines to paint a clearer picture of the current state of scientific literature regarding momentum.

It is first essential to define the differences between the two main momentum fields. The statistical field attempts to prove or disprove the notion that prior events lead to positive or negative individual/team outcomes in the future. The psychological field usually attempts to understand the cognitive changes an individual experiences while going through a series of positive or negative events and how these effects can change the individual's outlook on themselves and their abilities. This will often be referred to as psychological momentum.

We should also note the difference in the terms positive and negative momentum. Positive momentum is the phenomenon felt after a series of positive outcomes; it would be believed that this momentum would lead to more favorable results in upcoming events. The negative momentum is the phenomenon felt after a series of adverse outcomes, and likewise, this type of momentum would lead to less favorable results in upcoming events.

Below we will look at the prior work in this field, determine what their contributions to the research topic were, and try and determine what we can learn and incorporate into this paper. To best achieve this, we will first, in section 2.1 identify the aspects of the literature we are most interested in analyzing. We only include here what we believe to be relevant concepts that help show how this thesis uniquely contributes to the field of momentum research. Then, we will discuss these works in depth in Section 2.2. Lastly, in Section 2.3, we will compare this thesis with some of the prior works discussed. While we will discuss psychology-based works in section 2.2, as they can still inform us on the theoretical idea of how players or teams would build and maintain momentum, we will not include them in Section 2.3 due to the fact they are, in most cases, not looking to answer the question of whether momentum exists.

2.1 Analyzed Aspects

The aspects used for the comparison are as follows. First, we determine if the paper is ML inclusive, that is, does the paper use machine learning to explore the data? The second aspect is a meta-analysis of the Sport the paper focuses on. Here, we determine if the paper explores the entirety of the sporting event, including the outcome, and not just a tiny part of it, such as simple events (i.e., game actions) such as shooting. Next, we determine if the paper used some sort of shifting window. In other words, does the paper use a small sample of recent events to determine something about future events? Lastly, we identify if the research yields a proposed method of quantification for momentum in sports.

2.2 Papers on Momentum Analysis in Sports Analytics

The real beginning of mainstream momentum research started in 1985 with "The Hot Hand" [4]. This paper used a controlled experiment that involved participants shooting free throws and claimed that sequences of made or missed free throws were statistically independent events and therefore, players, coaches, and fans were perceiving momentum where there was none. The foundation of this school of thought is that an individual has a true shooting percentage and that given a sufficient sample size they will eventually return to this number despite streaks of made or missed shots. The nature of this paper was psychological as its main claim was that humans perceive streaks in random sequences. However, this paper served as the backbone of a large amount of the momentum research that followed and is still discussed heavily today. Thus, the "Hot Hand Fallacy" was born.

After "The Hot Hand" more research was focused on momentum, if it existed, and what effects it may have if it did. The work of Taylor and Demick [18] proposed a multidimensional model for momentum in sports. The authors proposed that momentum was developed through a series of changes, and termed this series of changes the "momentum chain". Through these series of changes, an individual could develop positive or negative momentum. While this paper was psychological in nature it can be viewed as an early attempt at quantifying momentum.

The work of Vergin [5] looked at winning streaks in Major League Baseball (MLB) and the National Basketball League (NBA) and found that when using statistical tests such as the chi-square goodness-of-fit test and the Wald-Wolowitz test, wins and losses in these leagues were statistically independent events. Thus, concluding that streaks of similar outcomes in the league were created by random chance. The paper uses a constant probability for each team to win which is represented by the team's end-of-season win percentage. This can be viewed as a shortcoming of the study and that the quality of a team should be their win percentage up to that point in time rather than judging a team's quality on games that have yet to transpire. When looking at our analyzed aspects, this paper only applies to the meta-analysis of sport, as it attempted to use team quality to determine if the outcomes of games were dependent or independent events.

When we look at individuals who play the games such as players or coaches, we find that the belief in momentum and momentum's effects are very strong. The work of Jones and Harwood[19] looked to examine psychological momentum from the players' point of view. What they found was that players have specific triggers that they believe lead to positive or negative psychological momentum and that players and coaches will go as far as to create strategies in order to maintain positive momentum once it has been obtained or overcome negative momentum when they are in a slump.

As analytics began to change the game of hockey [20] more papers began to examine the sport. The work presented by Shmanske and Lowenthal [21] used NHL game data to attempt to study overtime play in the league. For reference, an NHL game cannot end in a tie. If the score is tied after three periods the game will then enter a next goal wins overtime period. Using several variables for the home and away team, as well as multiple regression, the authors attempted to quantify the statistical significance of certain aspects of the game when heading into this overtime period. They found that if the home team had scored the last goal of regular time, then they were more likely to score the overtime goal and win the game. This means that the momentum from the last goal had carried over into the overtime period, however, this effect was not present for the away team. Therefore, the work shown by Shmanske and Lowenthal [21] made use of none of the aspects we are looking to analyze.

The work presented by Leard and Doyle [22] studied the effects of fighting, home advantage, and momentum in the sport of hockey. Using game-level data they constructed statistical models which accounted for aspects of the game such as fighting, home advantage, and current win streak. The authors found that when looking specifically at winning streaks of two or three games there was enough evidence to make the claim that these streaks were statistically significant to the outcome of the current game. Through this, they meet our qualification for a meta-analysis of their respective sport. The authors do, however, acknowledge that these streaks may be accounting for other unseen aspects of the game in some way.

The research presented by Arkes and Martinez [9] in 2011 looked to determine if there was any measurable proof of momentum in the NBA. The authors found that when they adjusted their quality metric for an individual team based on a short sequence of recent games there was statistically significant evidence that these teams were being affected by their recent results. They also find that these recent results could demonstrate team quality and thus, be used as an indication as to what team would win. Given the author's approach in this work, they meet the qualification for two of our analyzed aspects, those being meta-analysis and short-term window analysis. Interestingly, the authors also found that in their database both the home and away teams were affected the same by momentum. The paper also identifies what it believes to be three things that affect the outcome of the game. These are momentum, team strength, and randomness. This is one of the only relevant research studies that recognizes the randomness of sports and the fact that often the "better team" can lose.

The paper by Kniffin and Mihalek [6] explored a meta-analysis of hockey and momentum. Their paper focused on groups of two-game series from NCAA college hockey. They found that when accounting for team quality the outcome of the first game had no effect on the outcome of the second game and that these were two statistically independent events. The authors also found that a team's margin of victory in the first game had no effect on the second game. A potential shortcoming of this study was the measure of team quality which was the team's win percentage at the end of the current season. This is a similar technique to what we saw in work by Vergin [5] which studied the MLB and NBA.

In 2014 a paper was submitted to the MIT SLOAN Sports Analytics Conference which looked at the 83,000 shots in the 2012-2013 NBA season and found that players who were experiencing a streak were more likely to take lower percentage shots and the opposing team would also tend to guard the player more closely. When all this was put together the authors estimated that the "Hot Hand" could increase the likelihood of a player making a shot by 1.2 to 2.4 percentage points [10]. While this percentage increase may seem small, every percentage point matters in sporting events due to the fact a team's margin of victory is often small. This paper also denotes that all baskets made are not the same and this affects how a player builds momentum, if a player makes three of his last four simple layups he will likely not experience the hot hand. However, if he makes three of his last four more difficult three-point attempts he may potentially experience the effects of the hot hand.

There has also been some work in the area of predicting momentum shifts in NBA games. The work shown by Larow, Mittl, and Singh [15] used ML, specifically, Logistic Regression and SVM to attempt to predict streaks in play using a sliding window of the last 20 in-game events. This makes their work one of the first we reviewed that made use of our ML analyzed aspect. While the authors experimented with other window sizes this is the size used for the majority of the paper. The authors defined a streak as a period of play where a team scores eight or more points while holding their opponents to zero points. In doing this, the authors found success in predicting if a streak would occur obtaining a true positive rate of 0.787 and a true negative rate of 0.860.

One of the few statistical analyses for individuals in hockey was put forth by Vesper [23] in 2015. Using a number of variables that expressed the quality of a given shot along with regression models, the authors attempted to determine if individual players experienced the hot hand when shooting the puck. What they found was that after a player scored, their shot quality dropped and they became less likely to score a goal. The authors further explain that this negative hot hand may be due to a number of unseen changes, such as defensive adjustment from the opposing teams.

When it comes to European football or soccer as it is usually referred to in North America, there exists very little work on momentum. However, work by Parsons and Rohde [24] examined games from the English Premier League (EPL), in hopes of re-examining what is understood about the concept of the hot hand using fixed effects regressions to control for time-invariant heterogeneity in conjunction with traditional non-parametric techniques such as the Wald-Wolowitz test which has been used in past momentum-based studies such as the study by Vergin [5]. The authors proceeded to examine both across-game and within-game momentum, their findings neither accepted nor rejected the theory of the hot hand but added that evidence of momentum is largely context-dependent.

There has been some more recent research in the area of psychological momentum such as that conducted by Den et al. [25]. This study focused on 22 athletes competing in a rowing-ergometer tournament. The study's purpose was to find a connection between short-term momentum and long-term momentum. The authors found that individuals who had positive outcomes in the first two races were less perceptible to the negative momentum if they encountered a less desirable outcome in the third race than individuals who had negative outcomes in the first two races. The overall claim of the paper was that long-term and short-term momentum are interconnected and that positive long-term momentum can hinder the effects of short-term negative momentum. It is also worth mentioning that this paper was more focused on the athlete's perception of themselves and their performance and rather than the outcome of the sporting events.

There have been a large number of papers that have challenged the belief in the

"Hot Hand Fallacy" particularly in the mid to late 2010s. One of these papers that challenged the belief of the hot hand fallacy was a paper by Green and Zwiebel [13] in 2018. The authors analyzed data from the MLB and found what they considered to be strong evidence for the hot hand in several statistical categories. The authors claim that due to the fact baseball allows for minimal defensive changes to stop a hot player the evidence of the streaks becomes more clear. This stands in contrast to a sport such as basketball where there is a large number of ways for a defense to deal with a hot player. This was well addressed by the authors in [10] as discussed before.

Recently, the authors in [8] released a paper that looked at the 2019 NHL season in an attempt to find evidence of momentum at play. This paper was built upon a meta-analysis of hockey in an attempt to determine if momentum had an effect on outcomes. The 2019 NHL season is of interest particularly because the St. Louis Blues who sat at the bottom of the league about halfway through the season performed very well in the second half of the season and won the Stanley Cup. In their research, the authors used entropy to determine if wins and losses were dependent events. The authors found that there was no evidence of momentum and that wins and losses were independent events in all cases except for the Anaheim Ducks, which led them to conclude that if momentum exists its effects are likely exaggerated.

In the NBA, during All-Star weekend some players take part in a three-point shooting competition where they must shoot 20 balls from five different places behind the three-point line. Fans will often view this as an event where the hot hand is relevant and in [11], the authors looked to see if there was any evidence of this. Using 34 years of NBA three-point contest data and a statistical strategy that the authors claim accounts for bias, they found that there is considerable evidence of the hot hand in the three-point contest. The authors finally claim that the hot hand should be viewed as a bias and not as a fallacy, due to the fact that they provide evidence of the measurable effects of the phenomenon. They believe that it is better understood as a bias as people often overestimate or underestimate the effect it is having.

Recent work from Ding et al. [7] sought to determine if NHL goalies get hot during the playoffs by using a multilevel logistic regression analysis. The authors used a shifting window of the last "x" number of shots a goaltender faced or a time-based window, in order to feed data to the model. This is one of the few relevant research studies that made use of a moving window as in our own work. What the authors found was the opposite of "The Hot Hand" effect, they found that saves were an indication of worse performance ahead.

The most recent paper reviewed was published in 2022 [12]. The paper argued that most hot hand research was studies that took place in a vacuum, so they looked to study the hot hand in real game situations. The authors found that when considering in-game basketball situations and accounting for several variables, certain players do display the "hot hand". However, your average player does experience a drop in performance after several successful attempts, bringing their performance back to an equilibrium.

2.3 A Summary of Momentum Literature

We summarize the discussion in the previous section by presenting Table 2.1. This table compares this thesis with the discussed previous works in the field of momentum. As can be seen, this thesis is the only work that blends elements of machine learning, meta-analysis, short-term windows, and the quantification of momentum together.

Work	**Machine Learning Inclusive**	**Meta-analysis of Sports**	**Short-term Window Analysis**	**Quantification of Momentum**
R.C. Vergin (2000) [5]		✓		
S. Shmanske & F. Lowenthal (2007) [21]				
B. Leard & J.M. Doyle (2011) [22]		✓		
J. Arkes & J. Martinez (2011) [9]		✓	✓	
K. M. Kniffin & V. Mihalek (2014) [6]		✓		
A. Bocskocsky et al. (2014) [10]				
W. Larow et al. (2014) [15]	✓		✓	
A. Vesper (2015) [23]				
S. Parsons & N. Rohde (2015) [24]		✓		
B. Green & J. Zweibel (2018) [13]				
G.M. Steeger et al. (2021) [8]		✓		
JB. Miller & A. Sanjurjo (2021) [11]				
L. Ding et al. (2021) [7]			✓	
K. Pelechrinis & W. Winston (2022) [12]		✓		
This thesis	✓	✓	✓	✓

Table 2.1: A comparison of previous works and ours.

As we can see the research related to sports momentum is large and includes contributions from several different scientific fields. There seems to be a large amount of contradictory evidence for or against the statistical significance of momentum. However, it is worth noting that most of the relevant research studies that provided some statistical proof of momentum have been more recent.

These differences in findings could be due to any number of things such as models that do not take into account significant aspects of the sport in question. It could also be possible that evidence of momentum is situational and can only be measured in certain players or teams.

As previously mentioned, the amount of research that employs the use of some form of ML is slim, as is the amount of research that explores the sport of hockey. Most papers are based on more traditional statistical methods and more popular American sports such as basketball or baseball.

We should also note that the majority of momentum research focuses only on men's leagues. This is likely due to the lack of game event data for some of the major women's leagues. It is essential to identify this trend in the currently available research.

Due to the fact that sporting events include a significant amount of randomness that cannot be accounted for, our critique of the field would be that momentum research is too heavily tied to the outcome of the event. In previously discussed work by Arkes and Martinez [9], the authors outline their belief in three factors that determine the outcome of the game, those things being momentum, team strength, and randomness. I believe that team strength does not play as big of a role as some may believe. If we study winning percentages in the NHL regular season between the years of 2010-2021 (inclusive) and only use games that occur after each team's first 20 games in each season (this is done to ensure a reliable sample size to calculate a win percentage). We find that the team with the higher win percentage only wins the game 56.6% of the time, with a variance of 1.1%, and a range from 16.7% to 87.5%. It is possible that a more reliable metric to determine momentum is the standard of play upheld over a period of time, rather than the results that were obtained. We attempt to investigate this idea along with others in the coming sections.

To the best of our knowledge and based on all criteria discussed in this literature review, we place our work as unique in the field of sports analytics with a focus on momentum analysis as the first one to provide a methodology to quantify momentum, as well as making the use of ML, provide a meta-analysis of sports and using a short-term window.

Chapter 3

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

3.1 Introduction

Since the early 2000s, the National Hockey League (NHL) has undergone a significant change regarding the use of analytics [26]. 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 [27]. 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 [28]. 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 [4]. 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. [4] are generally undisputed, contradictory results have emerged in this field of study. For instance, the research conducted by Arkes and Martinez [9] 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 [12] 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 [6] 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 [8] 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 [29], Weissbock, Viktor, and Inkpen [30], Remander [31], and Pischedda [32]. 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 [33]. 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 [29]. 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 [32], 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 [31], 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 [29] 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 [14] and further supported by Gilbert and Wells [34], 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. [14] 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 [35] 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 [36]. 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 [37]. However, as was described by Osborne [38], 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 [29] 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.

3.2 Methods

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

3.2.1 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 [39]. 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 3.1 provides a small sample illustrating the structure of the raw data.

Table 3.1: 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	...

3.2.2 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 3.2 provides an example showcasing the structure of the game events data.

Table 3.2: 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	...
201200003	TBL	FL	20	29	10	...

3.2.3 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 [30] 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 [40], 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 [40], 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.

3.2.4 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.

3.2.5 Momentum-Based Features

Various definitions of momentum have been proposed in the scientific literature over the years. For instance, Arkes and Martinez [9] 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. [8] do not explicitly provide a definition of momentum but differentiate it from 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 [9], 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.

3.2.6 Features Used

Table 3.3 and 3.4 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 [32] and Weissbock et al. [30]. However, we also considered inputs from the public domain, considering numerous NHL advanced analytics sites, including MoneyPuck [40] and Natural Stat Trick [41]. 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 3.3: 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.	Slope-based, Average-based
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

Table 3.4: Features continued

Feature	Description	Deviations
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
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 [32] 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.

3.2.7 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 [42]. 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

[43], 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 [44]. For our machine learning algorithms, we selected models from the Sklearn library [42] to evaluate our pipeline. The chosen algorithms include logistic regression, random forest [45], and support vector machines (SVM), also known as SVM [46]. 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.

3.2.8 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 [47]. 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 (*).

3.2.9 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.

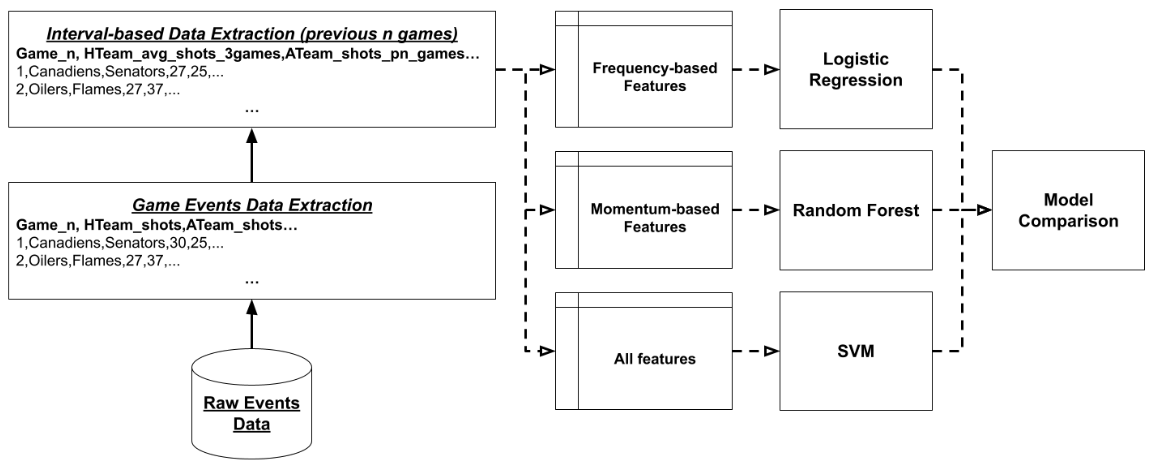

Figure 3.1: Data pipeline for our methodology.

3.3 Results

3.3.1 Logistic Regression

Figure 3.2 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 3.5, 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.

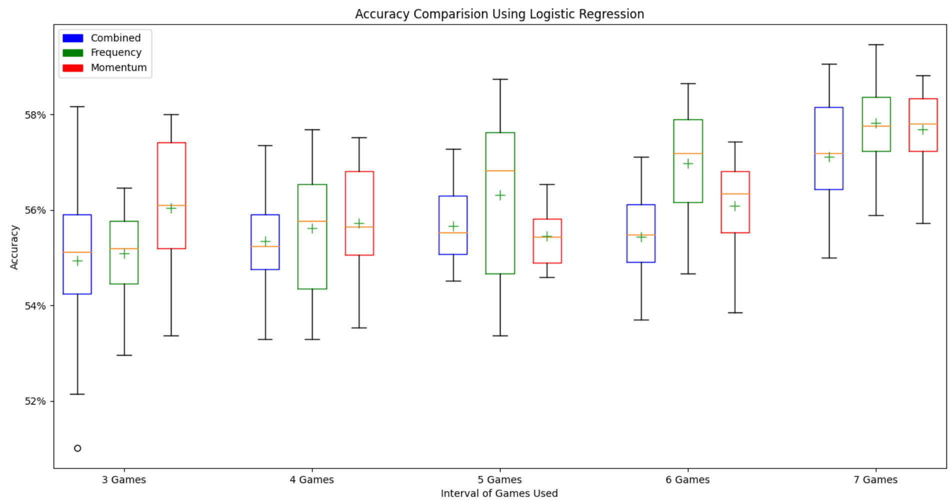

Figure 3.2: Comparison of the models using the logistic regression algorithm and game intervals 3 through 7.

Table 3.5: 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 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 Statistical value	28.5	78.5	41.5	17.5	85
Momentum – Combined Statistical value	15	33.5	52.5	28	34
Combined – Frequency Statistical value	86	57.5	60.5	16	20

3.3.2 Random Forest

Figure 3.3 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 3.6.

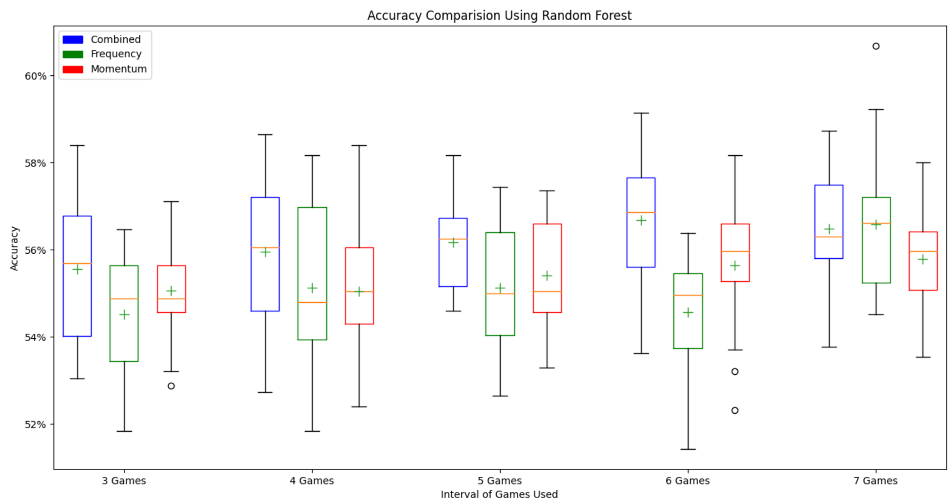

Figure 3.3: Comparison of the models using the random forest algorithm and game intervals 3 through 7.

Table 3.6: 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 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

3.3.3 SVM

Figure 3.4 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.

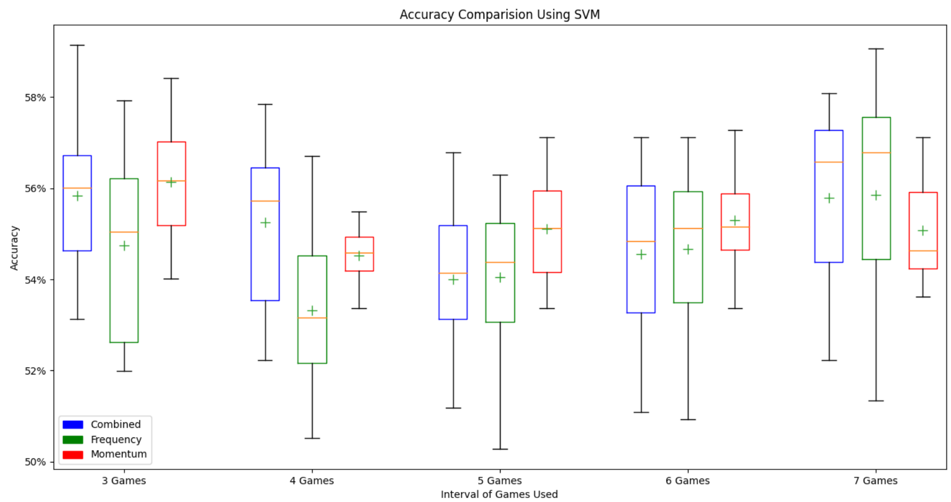

Figure 3.4: Figure 4. Comparison of the models using the SVM algorithm and game intervals 3 through 7.

Table 3.7: 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 [46].

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 [29] 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 [29] 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 [48].

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 3.2. 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 3.2. These averages fall within a four to five-percent margin of the proposed 62% ceiling by Weissbock and Inkpen [29]. 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 [49]. 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 [50].

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 [12], 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. [51], 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 [52], 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 [6] or Steeger et al. [8]. 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 [26].

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.

3.4 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.

Chapter 4

A Comprehensive Data Pipeline for Comparing the Effects of Momentum on Sports Leagues

4.1 Introduction

By the year 2027, the revenue for the global sports market is estimated to reach 623.63 Billion USD. This is a marked increase from the revenue of 486.61 Billion USD seen in 2022 [1]. This significant revenue growth also increases financial incentives for management groups to attempt to create successful sports franchises. One of the ways to improve a franchise is through the team's quality of play, which has given rise to

[1] Research and Markets: `https://www.researchandmarkets.com/reports/5781098/sports-global-market-report`

the use of sports analytics. Sports analytics are built on the idea that statistics, data science, and machine learning can improve a team's quality by allowing management to make informed decisions with the use of information about the in-game play that the average scout, coach, or person in management may not notice by simply watching games.

The acceptance of sports analytics by fans, players, or people in management has often been slow and met with skepticism. In other cases, teams have embraced analytics and prospered because of it. Such was the case for the 2002 Oakland Athletics, who managed to create a competitive MLB team at a fraction of the cost of their more wealthy competitors; this event became more colloquially known as "Moneyball" [53]. Due to this success, the field of sports analytics has grown to the point that many sports teams have in-house analytics departments. There is also a third-party market in which companies work with professional sports teams to provide in-depth statistical analysis or give them the tools to do so internally. Examples of such companies include Wyscout, [2] StatsBomb, [3] and Sportslogiq.[4]

Recently, a growing body of literature in sports analytics is interested mainly in studying the claims of many coaches, players, and fans that momentum impacts the outcome of the game [54]. These papers often conclude that momentum, as it is traditionally defined, does not exist in team sports [5, 4]. But in recent years, several

[2]Wyscout: `https://www.hudl.com/en_gb/products/wyscout`

[3]StatsBomb: `https://statsbomb.com/`

[4]Sportslogiq: `https://www.sportlogiq.com/`

papers have come to the opposite conclusion, stating that momentum impacts team sports [55, 56]. In reviewing the literature on data engineering in sports analytics, we found a gap in the research. That gap is in providing a comprehensive approach to building a generic data pipeline using game events, from which valuable features can be extracted from event data. In this work, we try to fill this gap as follows. First, we propose a way to extract distinct sets of features from the event data to use later to evaluate their predictive power; this allows us to evaluate the effect (or lack thereof) of momentum on a given sport. We define momentum as the increase or decrease in a team's overall quality of play over a small sample of games. Previous literature has often defined momentum differently. For instance, Arkes and Martinez [9] focused on momentum over multiple games and defined the effect of momentum as a "situation in which a team has a higher probability of winning or success had the team been playing well in the last few games". Fry and Shukairy [57] studied momentum more from an in-game perspective and defined it as the idea that "certain positive (negative) events during a game cause a team to do better (worse) subsequent to that event". Taylor and Demick [18] explore momentum from a psychological approach and define it as "a positive or negative change in cognition, physiology, affect, and behavior caused by a precipitating event or series of events that will result in a shift in performance". These definitions differ slightly, but all hinge on the idea that success or failure increases the likelihood of future success or failure. Our definition of momentum focuses on momentum over multiple games and is similar in some way to these previous examples, but rather than focusing on catalysts, we focus on the overall trends of teams, whether these trends are positive or negative. This means that we aren't dependent on all prior events being alike (positive or negative) as we are more concerned with the overall trend of the team. In our definition of momentum, the quality of play is defined as the features extracted from the in-game events. We assume that if momentum indeed plays a significant role in the outcome of a given game, our momentum-based features should be able to be incorporated into a model such that it can more accurately predict the outcome of the next game than a more traditional frequency-based approach. In this work, we provide a practical data pipeline that uses raw sports event data that can be transformed to create multiple feature sets (e.g., frequency, momentum, and combined) and then compare the predictive power of each of these feature sets using several different machine learning algorithms and performance metrics. We test our pipeline on data from the National Hockey League (NHL), National Basketball Association (NBA), and five major first-division European football/soccer leagues (e.g., Premier League, Bundesliga, La Liga, Serie A, and Ligue 1). Our standardized pipeline can be replicated in any sport for which event data is available.

This paper is organized as follows. Section 4.2 reviews several papers that discuss and propose methodologies to evaluate the effects of momentum in distinct sports and papers that propose sporting event data pipelines. Section 4.3 outlines our data pipeline in detail with information about how data was retrieved, processed, engineered, and finally used for comparisons in a machine learning model. Section 4.4

presents the results that we achieved in our research. Finally, Section 4.5 shows our conclusions and future works.

4.2 Literature Review

In our review of previous literature, we found only limited research in the area of data engineering pipelines for sporting event data; however, adjacent fields of study in sports can provide us with valuable information and potential ways to approach the momentum evaluation problem.

The work of Carson Leung and Kyle Joseph[58] proposed a way to mine data from college football games to predict the game's outcome. An approach to storing and mining the available statistical information was outlined, but their work focuses on predicting the games rather than defining and generalizing their mining approaches. This work also relied heavily on traditional statistical approaches to predicting outcomes rather than a machine learning-based approach.

The work conducted by Wang et al., [59] proposed a deep reinforcement learning approach to individual play retrieval from sporting data. They do this by proposing a way to measure and quantify the similarity between plays. They then use algorithms (some based on reinforcement learning) to determine how to split games to discover quality candidates within a given game. Lastly, using deep metric learning, they prune games that likely only contain poor candidates to improve the efficiency of their search. This work was, however, intended for situations where a person would be more interested in retrieving alike plays for comparison rather than using these plays to predict the outcome of games.

Very recently, the work of Wongta and Natwichai [60] shows how to create a data pipeline that ingests game analytics efficiently. Their work focuses on minimizing data flow issues in the pipeline from a much larger scope than what we are looking at in this work and includes research into how certain approaches can affect data transfer rates and how that, in turn, may affect the central processing unit (CPU) in a given system. Therefore, this work provides valuable information to those actively parsing data as the game takes place and are greatly concerned with the speed of the process, such as those third-party analytics companies we previously mentioned. This work was similar to their previous work [61], which focused on the multiple issues arising in such an end-to-end sports data pipeline and how to remedy them.

It should also be noted that a large amount of research focuses on predicting game outcomes in a given sport, which also usually gives a brief overview of the pipeline used to create a game-predicting model. This can be seen in the work by Thabtah, Zhang, and Abdelhamid [62], who predicted outcomes in the NBA; a similar work shown by Pischedda [32] predicted outcomes in the NHL; and the work of Rodrigues and Pinto [63] focused on European football/soccer. Differently from these works, our work intends to create an approach more focused on feature engineering and a data pipeline such that it could be used as a potential guide for the comparison of feature sets and the effects of momentum in the sports analytics research space.

The extensive body of work regarding momentum in sports has covered many disciplines. In this work, we are more concerned with statistics, machine learning, and data science research. The most influential paper in this space is likely the work of Gilovich et al., [4] whose work gave rise to the “hot-hand fallacy”. The study presented by Gilovich et al. showed that consecutive made or missed shots in basketball resulted from random sequences and not from a momentum effect on the player shooting the ball. This conclusion has been backed up several times in works such as those by Vergin [5], who found no evidence of momentum in winning streaks from Major League Baseball (MLB) or the NBA when using the Wald-Wolfowitz or chi-square goodness-of-fit tests. The work of Koehler and Conley [64] also found no evidence of the hot hand in NBA shooting competitions. However, as the years have passed, several papers provide contradictory evidence to the claim that perceived streaks are the result of random sequences. The work of Ritzwoller and Romano [55] found evidence that some players exhibit shooting patterns in basketball that do not indicate randomness. The work of Green and Zwiebel [13] found evidence of the hot hand in ten separate statistical categories from MLB. The work presented by Miller and Sanjurjo [56] argues that there exists a bias in such experiments of conditional dependence and that when this bias is accounted for, the conclusions of such studies are often reversed. It is essential to highlight that our goal is not to validate or refute any prior work on momentum but to shift the field’s discourse away from the dependence or independence of sequential events and toward viewing momentum from a feature engineering perspective and measuring the predictive power of such features.

The random nature of outcomes in sports has been measured in several ways over the years. The work of Lopez, Matthews, and Baumer [14] showed that the best team does not win the game frequently in North American sports. The work of Wunderlich, Seck, and Memmert [65] shows that goal scoring in the English Premier League (EPL) is often affected by the randomness of the sport. This insinuates that often, a team may lose games while, in theory, outplaying their opponent. Because of this demonstrable randomness, we believe a more appropriate approach to assessing momentum is quantifying the increase or decrease in a team or player's quality of play over a short period of time rather than focusing on the dependence of sequential outcomes.

Like the pieces of literature above, our work looks to answer if momentum exists in sports. However, we approach momentum from a different perspective by quantifying it using a team's linear play trend in several performance indicators rather than focusing on the dependence or independence of sequential outcomes. This allows us not to depend on the outcome of the game solely but to see how teams are trending in performance indicators that typically lead to success and use those trends to predict the outcomes of future games. While this approach is not without fault, we believe it can capture the essence of momentum more effectively than winning or losing streaks in a set of outcomes. As others have shown, sports are subject to randomness and are not determined by pure skill. Therefore, we should focus on what leads to winning/losing rather than the act of winning/losing. This, however, doesn't mean to say that there isn't a lot of value in the literature we have cited. Rather than refuting or affirming previous findings, we look to provide a new perspective on momentum in sports.

4.3 Materials and Methods

Figure 4.1 outlines the data pipeline proposed in this paper. Our pipeline starts with a raw dataset that includes individual events over games we intend to characterize. We then summarize individual games in our game event extraction, which groups events to determine key performance indicators in a game. We create the distinct features of our frequency-based, momentum-based, and combined feature sets using our game events. These feature sets are then fed into machine learning algorithms. The performance of these models is then evaluated with chosen performance metrics. After, the models are compared to each other, aiming to determine which groups of features have more predicting power.

4.3.1 Data Sources

The NHL dataset was sourced from a Python module called hockey scraper created by Harry Shomer[5]. This module scraped all NHL event data from the NHL's public API for 2011 to 2020. For this dataset, we only included games that ended in the regular

[5] Hockey Scraper: `https://hockey-scraper.readthedocs.io/en/latest/`

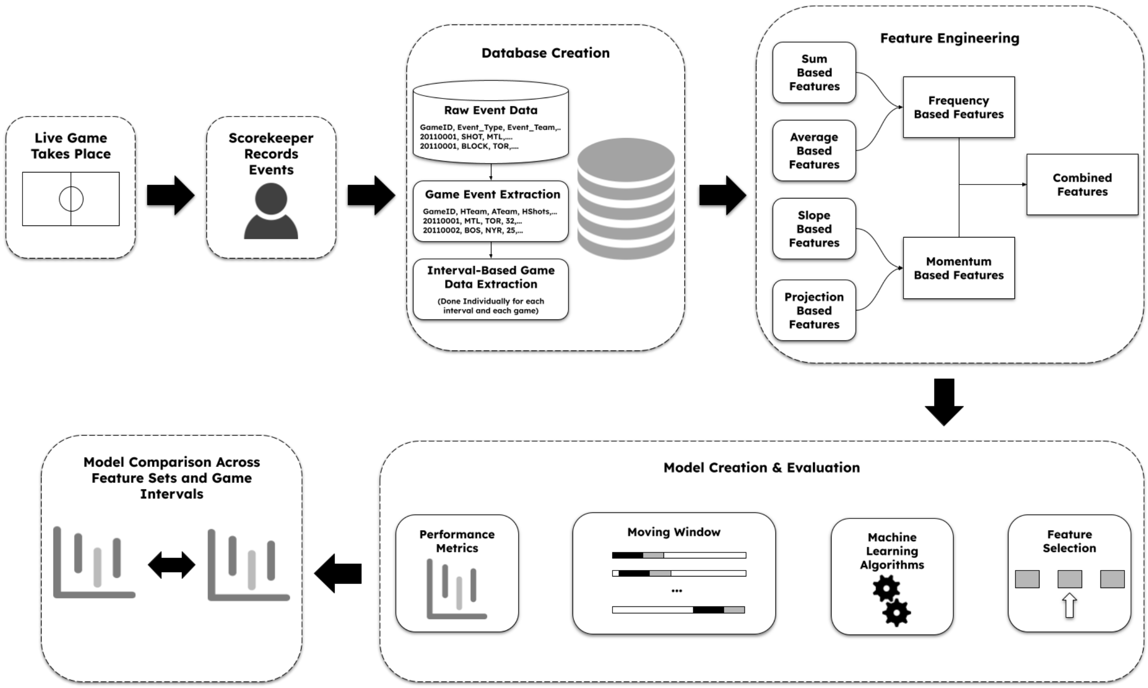

Figure 4.1: A diagram of our complete data pipeline.

time. Overtime games are often close and do not help models distinguish between the binary targets of a home win or away win. We also removed each team's first 20 games from each season as they seem to struggle with consistency up to this point in the season. Both of these steps are often used in the training set for NHL pre-game prediction models, and we were inspired to do this by the work done by Peter Tanner on MoneyPuck.com[6]. We combined the scraped event data with expected goals data, also sourced from MoneyPuck.com, to ensure we had a solid dataset to build our models upon.

The NBA dataset was sourced from kaggle[7]. This dataset held all play-by-play

[6] MoneyPuck: `https://moneypuck.com/`

[7] NBA play-by-play data 2015-2021: `https://www.kaggle.com/datasets/schmadam97/`

data for the NBA from 2015-2020. The NBA has a similar format to the NHL; each team plays 82 games a season, and ties after regulation are settled through overtime. We also found that removing overtime games and each team's first 20 games yielded similar results, and therefore, we performed these steps as we did for the NHL dataset.

The European football/soccer dataset came from the public dataset released by Pappalardo et al., [66] using data collected by Wyscout. This dataset contained data from only the 2017 season; however, it contained this data for each of the five major first-division European leagues, those leagues being the first division in Spain, Italy, England, Germany, and France. We could view the data from these five competitions as five seasons of data. Due to the fact these leagues do not include overtime periods, we did not have to remove any games for that purpose. However, we removed each team's first seven games as the number of games per team per season in the five major leagues ranges from 34 to 38. This is nearly the same percentage of games removed per season from the NHL and NBA datasets.

4.3.2 Game Event Extraction

Using the raw data as our main source, we created individual game events that captured all the main events in a given game. This was done by grouping events based on their unique game ID. Each dataset had its way of uniquely identifying games. The NHL uniquely identified games with pairs of game IDs and the season in

`nba-playbyplay-data-20182019`

which the game took place. The Soccer dataset took the same approach, while the NBA dataset used the home team, away team, and date to identify each game.

We created a unique game ID for each game, so we would only require a single column to uniquely identify a game. We combined the season and game ID strings for the NHL and Soccer datasets to create a single game ID string. This would work as follows. If the game had the ID "1115" and was being played in the 2015 season, our created game ID would be "20151115". While in the NBA, we used the season and a counter variable that incremented every new game. For example, the first game of the 2015 season would be game ID "201500001" and the second game would be "201500002" and so on.

These unique game IDs allowed us to create the game events by querying the raw datasets. For example, if we were looking to determine the number of shots the home team took in a given game, we would count the number of "SHOT" events that contained the unique game ID in the game_id column and contained the home team in the event_team column. This allowed us to summarize the outcome of each game from the event data and discern statistics such as the score, number of shots, shooting percentage, and more.

Table 4.1 displays a small amount of the information that is held in each game event and shows how they are organized in the database.

Table 4.1: Example of the data contained in each game event for the NHL.

Game ID	Away Team	Home Team	Away Goals	Home Goals	Away Shots	...
20150001	MTL	TOR	3	2	25	...
20150002	CAL	EDM	1	3	30	...
20150003	MIN	PIT	2	0	22	...
...	...	...	...	...	...	...

4.3.3 Interval-Based Game Data Extraction

One of the fundamental parts of our pipeline is the game intervals. These intervals represent the number of games used to assess team quality in our models. For each game in our dataset, we would use the previous "n" games to gather statistics representing each team's quality. For example, if we use an interval of three games, we will only use each team's last three games when creating the features that represent the quality of those teams for a given game we are trying to predict the outcome of. Specifically, this paper uses intervals of three to ten games.

For each game and each game interval, this extraction must take place. The data from interval-based extraction is used to calculate the values for our features and is, therefore, done concurrently with our feature set extraction to ensure that only one extraction of interval-based data takes place, therefore making the parsing of the data more computationally efficient.

The game IDs that were previously discussed can help improve this process. If IDs are sequential, that is, the IDs increase in the order of game occurrence, we can extract the 'n' most recent games by getting all games involving a given team that has a game ID less than the ID of the game we are modeling, from there we get the 'n' games with the highest game ID. If the game IDs are not sequential, a better approach would be to retrieve the most recent games using their dates.

4.3.4 Feature Engineering

This work considers three feature sets in our pipeline: frequency-based, momentum-based, and combined. The frequency-based and momentum-based features have different sub-features, which we will discuss below. The combined feature set represents the combination of both the frequency-based and momentum-based feature sets. The features that were used by the models can be seen in Tables 4.6, 4.7, and 4.8. The features extracted from the data were largely reflections of the event types recorded in each dataset. However, in the case of the NHL and NBA, more advanced performance indicators were extrapolated using the data. The advanced metrics used from each of these datasets were influenced by performance indicators discussed in other research, such as work in ice hockey by Johansson, Wilderoth, and Sattari [20] and work in basketball by Kubatko, Oliver, Pelton, and Rosenbaum [67]. Some popular public sources were also consulted, such as Natural Stat Trick[8] and Basketball-Reference[9].

[8]Natural Stat Trick: `https://www.naturalstattrick.com/`

[9]Basketball-Reference: `https://www.basketball-reference.com/`

4.3.4.1 Frequency-Based

Two separate frequency-based feature types are considered in our pipeline, named sum-based and average-based features. These features represent a primarily static and traditional approach to pre-game prediction in sports. These features are created by first querying data from each team's previous 'n' games as defined by our game interval. From there, we select a statistic for which we are creating these features, such as shots or blocks. Lastly, we extract features from the chosen statistic. Sum-based features are the sum of the statistics over the previous 'n' games as defined by the game interval, i.e., the total number of shots over the last three games. At the same time, average-based features are the average statistical value achieved over the previous 'n' games, i.e., the average number of shots over the last three games.

Table 4.2 shows statistics achieved by a given team over the last three games. We use this table to illustrate how to create sum-based and average-based features to predict the outcome of the team's upcoming game on January 7th. If we were to calculate the sum-based feature for the shots statistic on this set of recent games, we would add 25, 42, and 35 together, giving us a sum-based shot feature of 102. If we were now to calculate the average-based goal feature, we would divide the total number of goals by the number of games in the interval. So, we would divide 7 by 3, giving us an average-based goal feature of 2.33.

Table 4.2: Example of statistics over a team's last three games.

Game Date	Number of Shots	Number of Goals
January 1st	25	0
January 3rd	42	3
January 6th	35	4
January 7th	?	?

4.3.4.2 Momentum-Based

This work defines two separate momentum-based feature types, slope-based and projection-based. Slope-based features are created as follows: first, we query data from each team's previous 'n' games. We then select a statistic, such as shots, blocks, goals, etc., to create momentum-based features. Using the queried data from the recent games and our selected statistic, we create a two-dimensional space to plot our data, where the x-axis represents the passing of days between games and the y-axis represents the statistical value they obtained for each given game. For example, if we had chosen the number of shots for our statistic, we would have plotted the number of shots for each game on the y-axis with the x-axis representing the passing of days between games starting at $x = 0$ for the first/earliest game in the interval. With these points, we calculate the linear line of best fit, and the slope of such a line would be our slope-based feature. Projection-based features are essentially just an extension of our slope-based features. Using the equation of the line of the best fit, we calculate the slope-based feature. We add the x value, which is the number of days that have passed since the first day in our game interval, and we project the statistical value for the upcoming game if the given trends continue.

We feel that an example can best show how these features are calculated. Suppose we are trying to calculate the slope-based and projection-based features using a three-game interval for the team in Table 4.2. We have chosen the number of shots to be the statistic we are using for our calculations, and we are calculating features for the upcoming game on January 7th. Using this data, we plot the data in the two-dimensional space as previously stated. Starting with the first game in our interval, which is the game on January 1st, we would plot $x = 0$, $y = 25$. From there, we would count the days between the game on the 1st and the 3rd; there are two days between these games, so next, we would plot $x = 2$, $y = 42$. Lastly, we would note that January 6th is five days after January 1st, so we plot $x = 5$, $y = 35$. With the points we have plotted, we can calculate the linear line of best fit; this line of best fit is represented by the equation $y = 1.657x + 30.13$. From this equation, we can see that our slope-based feature would be 1.657. Now, using the line of best fit, we can calculate our projection-based feature. Using the fact that January 7th is six days after our first game in the interval on January 1st, we plug $x = 6$ into our previously stated line of best-fit equation. This gives us roughly 40.08, which would be our projection-based feature and represents how many shots the team should have if recent trends continue.

This example can be seen in Figure 4.2.

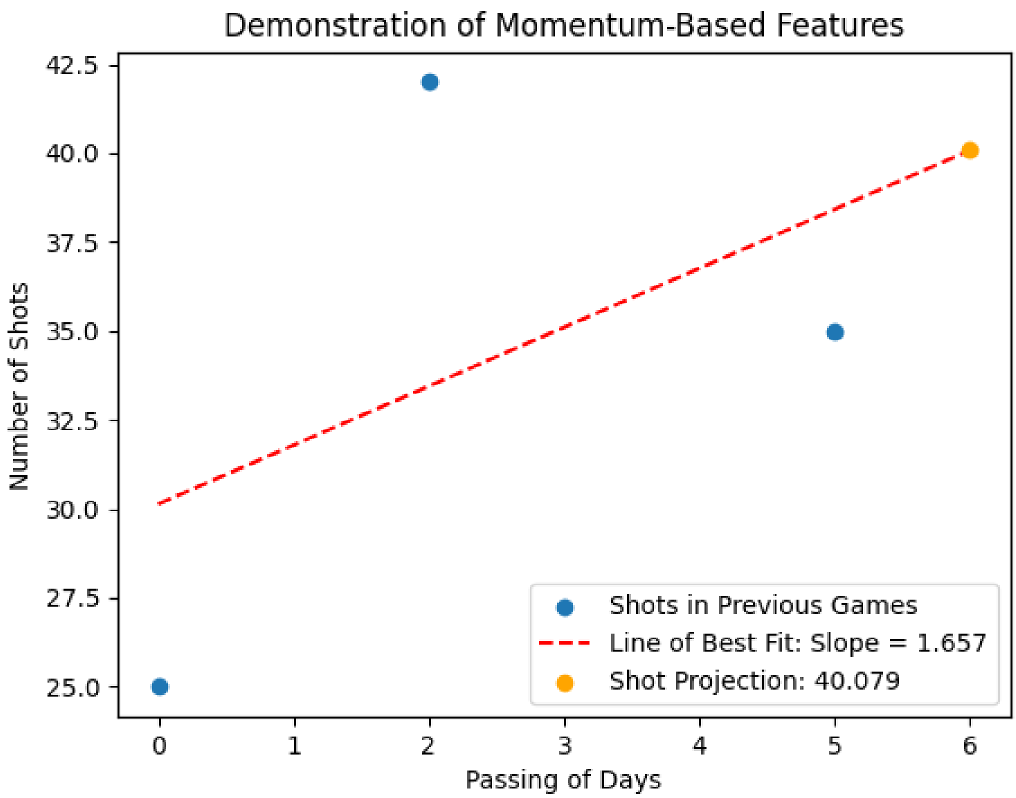

Figure 4.2: Example of how momentum-based features are calculated with the use of the data in Table 4.2.

4.3.5 Model Evaluation/Comparison

We use a moving window to generate the data folds rather than evaluating our models using a standard cross-validation approach. This is because sports events are, in many ways, temporal data, and therefore, we do not want future events to be predicting past events. While this approach may slightly decrease our predictive power because we will be training on significantly fewer games than we would when using a traditional 80-20 split and cross-validation, we believe it keeps the integrity of the temporal data.

Our moving window first sorts all games in order of their occurrence. From there, we determine the number of games the model should train on and the number of games it should test on; these numbers are different for each sport we will discuss. Next, we divide all games into 20 equal train-test splits, such that a model is trained on a group of games and then tested on a group of games that begin immediately after the last game in the training set, thus ensuring that the future is not predicting the past. Therefore, these train-test splits move from left to right in the window of time.

As we alluded to, the train-test size is different for every sport due to the difference between the number of games in a season for each sport. The training and testing windows in the NHL and NBA are the same size, as teams in both leagues have 82-game seasons. The training and testing size used for the NHL/NBA is 2460 and 1230, respectively. We need to use much smaller training and testing sizes in the soccer dataset as teams play significantly fewer games, with only 34-38 games for each team in a given season and fewer teams playing in the league. The training and testing sizes used for the soccer dataset are 760 and 380, respectively. These sizes were chosen to represent roughly two seasons of training and one season of testing in all three sports.

We adjusted the feature representation before our data was fed into our machine-learning algorithms. Rather than having two values for each feature, one for the home team and one for the away team, we created one singular feature by subtracting the away feature from the home feature. For example, if we were creating our goals feature and the value for the home team was 10, and the value for the away team was 7, we would subtract 7 from 10, giving us 3 for our goals feature. Our inspiration for this type of feature representation came from the work of Pischedda [32], who found that this increased the predictive power of their NHL game prediction model. We came to the same conclusion when testing our models and using this feature representation.

We have included feature selection approaches in our pipeline. We tested multiple approaches, such as recursive feature selection and sequential feature selection. However, we obtained the best results with the SelectBestK method. This method is handed some features to select (k) as a parameter and then selects the k best-scoring features from the complete feature set per a chosen statistical measurement. We used this method to select the best half of the available features for each model per their scoring based on mutual information. We perform feature selection on each training set in the moving window. This means we perform it 20 separate times, once for each training and testing set, which can result in slightly different features being chosen each time. We chose to do this because selecting features based on the entire dataset may introduce a bias into our moving window that unfairly increases the performance of our models.

We used three separate machine learning approaches for our models: logistic regression, random forest [45], and linear discriminant analysis (LDA). While we attempted to use more powerful techniques such as XGBoost [68] or neural networks, all gradient boosting methods yielded inferior results with our data. When testing neural networks, we found it was impossible to produce a neural network that wasn't biased toward one feature set or the other, and the logistics of having a network for each model and feature set was beyond the scope of what we are proposing here as our focus is the data pipeline itself. However, we should note that any machine-learning techniques could be employed in this pipeline.

To evaluate our models' predictive power, we used the accuracy metric from Sklearn [42]. After our models have predicted the outcome of each game in the testing set, the accuracy metric determines what percentage of those predictions were correct. Our data manipulation and model evaluation were done using the Python programming language. For our data manipulation, we relied heavily on the Pandas module [69], while our machine-learning algorithms and evaluation metrics were all imported through Sklearn [42]. Lastly, we created data visualizations that show the average accuracy obtained at each game interval for each machine-learning technique and each sport. These visualizations were created using the Matplotlib and Seaborn modules [47, 70].

We should also note the hyperparameters used with our machine-learning techniques. Logistic regression used all default hyperparameters except for max iterations set to 10000. Random forest used a random state of 1415 to ensure reproducibility and had the number of estimators set to twenty percent of the training window size. Lastly, LDA used all default hyperparameters as well.

4.4 Results

Below, we present the results obtained from each machine-learning technique we employed. In all figures of this subsection, we report the average accuracy of the 20 equal train-test splits in the NHL, NBA, and football/soccer datasets. In subsection 4.4.1, we show the results obtained when using logistic regression, followed by subsection 4.4.2, which shows the results obtained when using random forest. Lastly, subsection 4.4.3 shows the results obtained when using LDA.

4.4.1 Logistic Regression

In Figure 4.3 and Table 4.3, our logistic regression analysis unveils intriguing findings. It becomes evident that the performance of various feature sets differs across different sports leagues. In the NHL, for instance, the momentum-based models exhibit superior performance when game intervals are shorter. In contrast, frequency-based features and combined features consistently outperform momentum-based features for the NBA across all game interval durations. Turning our attention to soccer, we observe that the momentum-based model excels when game intervals range from six to seven. While frequency-based and combined features perform similarly to each other throughout all intervals. Logistic regression consistently delivers the highest average accuracy across all three sports, underscoring its effectiveness in our analysis.

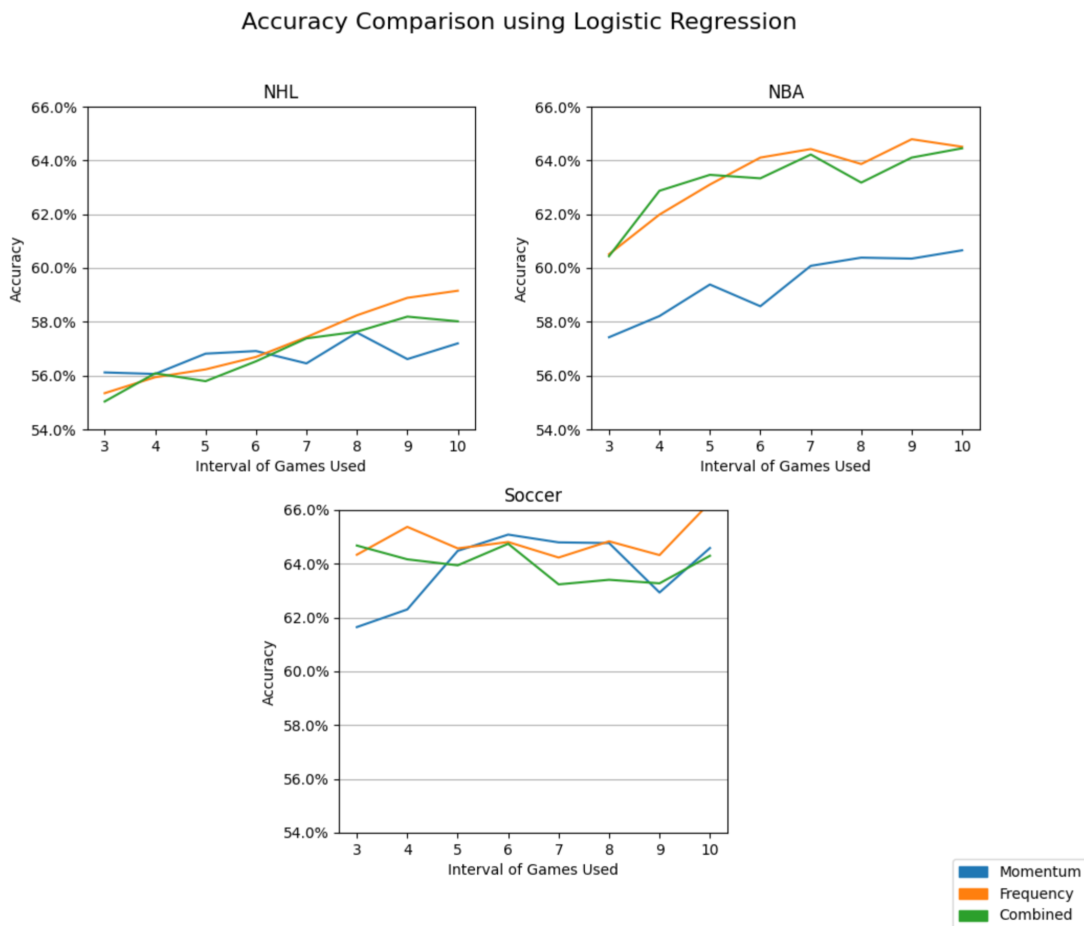

Figure 4.3: A comparison of the accuracy achieved among different sports and feature sets using logistic regression.

Table 4.3: Average accuracy obtained from the 20 train-test splits of the moving window for each league and feature set when using logistic regression. The highest average accuracy for each league at each game interval is represented in bold font.

League	**Feature Set**	**Interval of Games Used**							
		3	4	5	6	7	8	9	10
NHL	Frequency	0.5535	0.5594	0.5623	0.5669	**0.5743**	**0.5824**	**0.5889**	**0.5916**
	Momentum	**0.5612**	0.5606	**0.5682**	**0.5691**	0.5645	0.5760	0.5661	0.5720
	Combined	0.5504	**0.5608**	0.5580	0.5653	0.5738	0.5763	0.5819	0.5802
NBA	Frequency	**0.6051**	0.6199	0.6310	**0.6411**	**0.6442**	**0.6387**	**0.6479**	**0.6451**
	Momentum	0.5742	0.5821	0.5939	0.5858	0.6008	0.6039	0.6035	0.6066
	Combined	0.6044	**0.6287**	**0.6346**	0.6333	0.6422	0.6318	0.6411	0.6445
Soccer/Football	Frequency	0.6434	**0.6538**	**0.6457**	0.6481	0.6423	**0.6484**	**0.6433**	**0.6628**
	Momentum	0.6165	0.6230	0.6448	**0.6509**	**0.6480**	0.6477	0.6293	0.6459
	Combined	**0.6468**	0.6417	0.6394	0.6475	0.6324	0.6341	0.6328	0.6430

4.4.2 Random Forest

In Figure 4.4 and Table 4.4, some interesting observations emerge as we present the results obtained for the random forest model. Momentum-based models, it appears, do not exhibit strong performance across any of the sports. In contrast, the combined and frequency-based feature sets consistently yield better results across all three sports. Notably, in both the NHL and soccer datasets, the combined features display an edge over the frequency-based feature set, suggesting that combining frequency-based and momentum-based features may offer a promising approach to enhance prediction accuracy. This can also be seen to a lesser extent in the NBA.

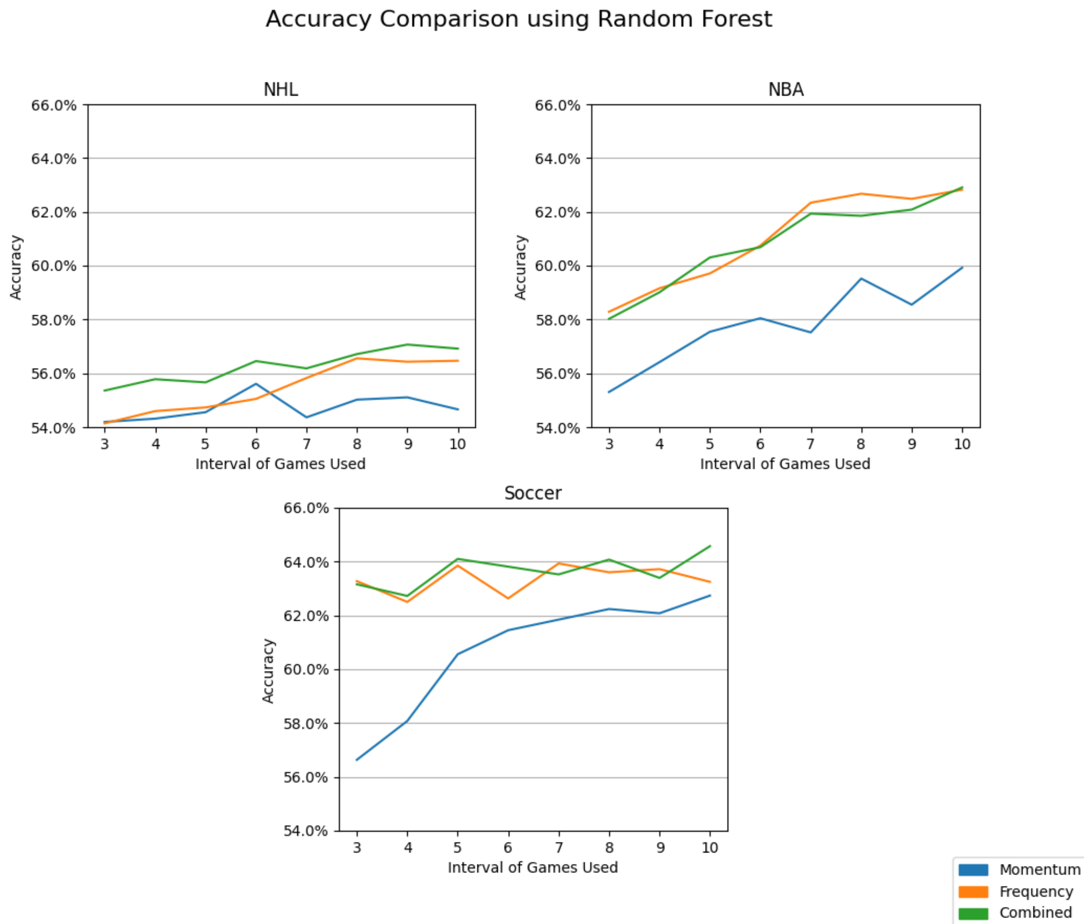

Figure 4.4: A comparison of the accuracy achieved among different sports and feature sets when using random forest.

Table 4.4: Average accuracy obtained from the 20 train-test splits of the moving window, for each league and feature set when using random forest. The highest average accuracy for each league at each game interval is represented in bold font.

League	Feature Set	Interval of Games Used							
		3	4	5	6	7	8	9	10
NHL	Frequency	0.5414	0.5459	0.5473	0.5505	0.5582	0.5656	0.5643	0.5647
	Momentum	0.5419	0.5431	0.5455	0.5561	0.5436	0.5502	0.5511	0.5466
	Combined	**0.5536**	**0.5578**	**0.5566**	**0.5645**	**0.5618**	**0.5671**	**0.5707**	**0.5692**
NBA	Frequency	**0.5829**	**0.5916**	0.5971	**0.6074**	**0.6234**	**0.6267**	**0.6248**	0.6282
	Momentum	0.5530	0.5641	0.5754	0.5805	0.5752	0.5952	0.5855	0.5992
	Combined	0.5802	0.5901	**0.6030**	0.6069	0.6193	0.6185	0.6208	**0.6291**
Soccer/Football	Frequency	**0.6328**	0.6250	0.6385	0.6263	**0.6393**	0.6360	**0.6372**	0.6325
	Momentum	0.5663	0.5808	0.6056	0.6145	0.6184	0.6224	0.6208	0.6274
	Combined	0.6316	**0.6272**	**0.6410**	**0.6381**	0.6353	**0.6408**	0.6339	**0.6458**

4.4.3 Linear Discriminant Analysis

Utilizing linear discriminant analysis (LDA) reveals a consistent preference for the frequency-based feature set, a trend evident in Figure 4.5 and Table 4.5. However, a couple of notable results are present, the NHL dataset mirrors the initial advantage for momentum-based features observed with short game intervals when using logistic regression. We consistently observe a preference for frequency-based features in the NBA dataset across all but two game intervals. In the soccer dataset, the frequency-based features excel at all intervals. However, all feature sets seem to perform well in the soccer dataset.

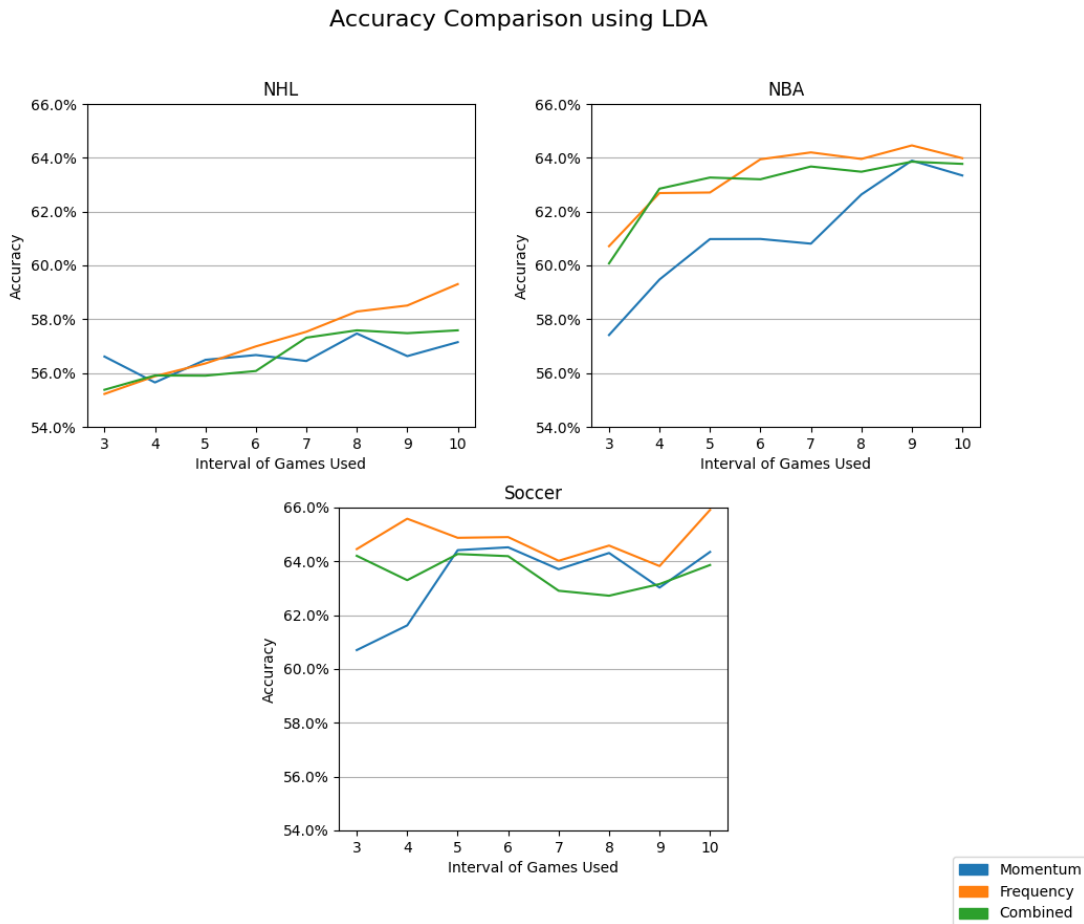

Figure 4.5: A comparison of the accuracy achieved among different sports and feature sets when using linear discriminant analysis.

Table 4.5: Average accuracy obtained from the 20 train-test splits of the moving window, for each league and feature set when using LDA. The highest average accuracy for each league at each game interval is represented in bold font.

League	Feature Set	Interval of Games Used							
		3	4	5	6	7	8	9	10
NHL	Frequency	0.5522	0.5588	0.5636	**0.5699**	**0.5753**	**0.5829**	**0.5851**	**0.5931**
	Momentum	**0.5661**	0.5565	**0.5649**	0.5667	0.5645	0.5747	0.5663	0.5715
	Combined	0.5538	**0.5591**	0.5590	0.5608	0.5732	0.5759	0.5748	0.5759
NBA	Frequency	**0.6071**	0.6269	0.6271	**0.6394**	**0.6415**	**0.6396**	**0.6446**	**0.6398**
	Momentum	0.5741	0.5948	0.6098	0.6098	0.6081	0.6264	0.6390	0.6334
	Combined	0.6007	**0.6285**	**0.6327**	0.6320	0.6368	0.6348	0.6385	0.6377
Soccer/Football	Frequency	**0.6446**	**0.6559**	**0.6488**	**0.6490**	**0.6402**	**0.6459**	**0.6383**	**0.6591**
	Momentum	0.6070	0.6162	0.6442	0.6452	0.6371	0.6431	0.6303	0.6435
	Combined	0.6421	0.6330	0.6427	0.6419	0.6291	0.6272	0.6316	0.6387

4.5 Discussion

Upon examining our results, several compelling patterns come to light. Notably, in the context of the NHL dataset, we discern potential evidence of momentum's influence across all three machine learning techniques employed. Both logistic regression and LDA reveal that momentum-based features and at times combined features hold an early predictive advantage over the frequency-based feature set. This finding carries notable significance, suggesting that a team's performance trend, as opposed to their aggregated and averaged performance over the same period, better predicts outcomes when we only consider a small sample of recent games to access team quality. This insight implies that momentum can wield predictive power, particularly for outcomes over shorter spans of games such as three or four. It is essential to clarify that this does not imply that momentum-based features or momentum itself are universally superior predictors; instead, it underscores that momentum-based features can enhance more intricate models by offering valuable insights into short-term performance.

In the case of random forest and the NHL dataset, we observe that a combination of momentum-based and frequency-based features exhibits the highest predictive efficacy. Given that random forest is a more powerful technique than logistic regression and LDA, this outcome suggests that the model places significant weight on specific features from the combined feature set. This finding emphasizes the potential advantages of employing combined features when the goal is to predict the game outcome, provided that feature selection and hyperparameter tuning are rigorously performed, particularly in the context of more complex machine learning techniques.

In the NBA dataset, our analysis reveals a minimal effect attributable to momentum across all three machine learning approaches applied in this study. At times we see advantages when using the combination of both frequency-based and momentum-based features but these advantages are rare and short-lived. Several factors may contribute to this lack of evidence, including the specific features we utilized, the level of hyperparameter tuning, or the inherent characteristics of the NBA itself. It is worth acknowledging that the NBA, as shown by Rockerbie [71], may exhibit distinctive attributes, such as lesser degrees of team parity and randomness, which can potentially impact the predictability of outcomes.

When examining the outcomes derived from our European football/soccer dataset, we uncover a particularly intriguing trend, mainly when focusing on the performance of momentum-based features, notably within the context of logistic regression. In contrast to the NHL, where momentum-based features demonstrated an advantage over shorter intervals of three or four games, the soccer dataset presents a distinctive pattern. Here, the initial performance of momentum-based predictors is underwhelming, but they exhibit exponential growth in predictive power, reaching their peak at game intervals of six to eight. Notably, logistic regression showcases an accuracy peak at an interval of six to seven games. We also see the combined feature sets outperform the other feature sets the majority of the time when using random forest. Though the margins of this advantage are small, this recurring trend emphasizes the potential benefits of leveraging combined feature sets with more sophisticated machine-learning techniques.

The primary objective of this study was to demonstrate the feasibility of comparing momentum's impact within a given sport by utilizing both frequency-based and momentum-based features. We contend that we have effectively elucidated and substantiated the rationale behind this methodology. Furthermore, we have illustrated this approach's practical implementation by contrasting momentum's effects across three distinct datasets.

The outcomes presented in this study strongly suggest that team-based momentum is a tangible phenomenon, quantifiable through the use of momentum-based features. We believe these findings lay the foundation for future predictive models, which could employ a combined feature approach to capture short-term and long-term success in sports analytics more comprehensively.

On a more practical level, this work helps us differentiate the level to which momentum affects different sports, thus allowing us to contextualize other literature on the topic in their respective sports. For instance, studies that find no evidence of momentum in the NBA are no surprise, given our findings here. This work could also be adapted to examine how momentum affects individual teams differently. This, in theory, could give teams a better understanding of how an upcoming opponent may perform, given their predisposition to being affected by momentum or the lack thereof. For instance, if a team seems to be heavily affected by momentum, their upcoming opponents can put extra emphasis on their more recent games to judge the quality of the team.

In this work, we did not use neural networks due to the large scope of constructing multiple network architectures for each sport and feature set. However, we would like to explore how neural networks, particularly Long Short-Term Memories (LSTMs), interact with each feature set in future works. LSTMs are particularly interesting due to their advantages when using temporal data such as the datasets we have outlined in this manuscript. Ideally, we would like to construct a pipeline capable of engineering unique networks for each sport and feature set. Thus allowing us to further our understanding of the predictive power of momentum-based features.

It is essential to acknowledge certain limitations in our overall approach. Our models did not implement a rigorous feature selection technique, nor a hyperparameter selection technique of any kind, potentially introducing bias towards specific feature sets in certain sports. In future research, a practical strategy that combines momentum-based and frequency-based features from various game interval durations may prove instrumental in achieving more accurate predictive models, particularly in sports like the NHL, where momentum exhibits more substantial indications. Additionally, our approach to engineering momentum-based features has two potential drawbacks. The first is related to cases where we use as few as three games for our game interval. In such instances, the residuals may deviate significantly from the line of best fit. Currently, there is no immediate solution to this issue in our research. Nevertheless, in future endeavors, exploring the possibility of exclusively predicting games for which the line of best-fit results in an acceptable value for the residual sum of squares (RSS) may be worthwhile. Another essential concern is using momentum-based features with many recent games, such as 20 or more. In such cases, it is possible that the slope of the momentum-based feature could approach zero. However, employing a more significant number of games may not align with the conventional concept of momentum, which is typically viewed as a short-term phenomenon. As the slope feature approaches zero, it could be harnessed with the residual sum of squares (RSS) to gauge consistency in scenarios where larger game intervals are applied. Recognizing that these larger game intervals might also lead the projection-based feature to yield outcomes connected to those obtained through the average-based feature is essential. To avoid this, we recommend limiting the use of this method to no more than 10-12 recent games unless the objective is to measure consistency rather than momentum. The specific limit could vary, depending on the league's structure for which predictions are made. It is crucial to clarify that our intention is not to imply that momentum-based features, in isolation, are inherently superior to frequency-based features. Instead, we aim to illustrate the capacity to measure momentum in ways that extend beyond solely relying on sequential outcomes. Through these diverse approaches, we can gain insights into how momentum can influence different sports in distinctive manners, potentially paving the way for developing more robust pre-game prediction models.

4.6 Conclusion

In this work, we aimed to outline a data pipeline that could effectively compare and demonstrate the predictive power of momentum-based features among multiple sports leagues. We believe we have achieved that, and in its current state, this pipeline can be used to measure the effects of momentum in any sport for which event data is available and could also be adapted to view the effects of momentum on individual players. We also believe we have provided further insight into momentum by approaching the phenomenon from a trend of play perspective; we believe this view allows us to emphasize that momentum can be separate from game outcomes and that it is more important to focus on what historically leads to winning rather than winning itself. In the future, we would like to explore potential refinements of the pipeline to select the optimal features and hyperparameters for each model to explore the capabilities of the momentum-based features further, particularly when they are being combined with frequency-based features in more robust models such as random forest.

4.7 Abstract

Table 4.6: The statistics that were used to create the features for the NHL model.

Statistic	Sum-based	Average-based	Slope-based	Projection-based
Wins	✓			
Loses	✓			
Goals For	✓	✓	✓	✓
Goals Against	✓	✓	✓	✓
Goals For 5v5	✓	✓	✓	✓
Goals Against 5v5	✓	✓	✓	✓
Goals For 5v5 Close	✓	✓	✓	✓
Goals Against 5v5 Close	✓	✓	✓	✓
Shots For	✓	✓	✓	✓
Shots Against	✓	✓	✓	✓
CORSI	✓	✓	✓	✓
CORSI 5v5	✓	✓	✓	✓
CORSI 5v5 Close	✓	✓	✓	✓
Faceoffs	✓	✓	✓	✓
Hits For	✓	✓	✓	✓
Hits Against	✓	✓	✓	✓
Penalty Minutes For	✓	✓	✓	✓
Penalty Minutes Against	✓	✓	✓	✓
Blocks For	✓	✓	✓	✓
Blocks Against	✓	✓	✓	✓
Giveaways For	✓	✓	✓	✓
Giveaways Against	✓	✓	✓	✓
Takeaways For	✓	✓	✓	✓
Takeaways Against	✓	✓	✓	✓
xG For	✓	✓	✓	✓
xG Against	✓	✓	✓	✓
xG For 5v5	✓	✓	✓	✓
xG Against 5v5	✓	✓	✓	✓
xG For 5v5 Close	✓	✓	✓	✓
xG Against 5v5 Close	✓	✓	✓	✓
Power Play Opportunities For	✓	✓	✓	✓
Power Play Opportunities Against	✓	✓	✓	✓
Power Play Goals For	✓	✓	✓	✓
Power Play Goals Against	✓	✓	✓	✓

Table 4.7: The statistics that were used to create the features for the NBA model.

Statistic	Sum-based	Average-based	Slope-based	Projection-based
Wins	✓			
Loses	✓			
Points For	✓	✓	✓	✓
Points Against	✓	✓	✓	✓
Field Goal Percentage For	✓	✓	✓	✓
Field Goal Percentage Against	✓	✓	✓	✓
Free Throw Percentage For	✓	✓	✓	✓
Free Throw Percentage Against	✓	✓	✓	✓
3-Point Field Goal Percentage For	✓	✓	✓	✓
3-Point Field Goal Percentage Against	✓	✓	✓	✓
Rebounds For	✓	✓	✓	✓
Rebounds Against	✓	✓	✓	✓
Offensive Rebounds For	✓	✓	✓	✓
Offensive Rebound Percentage For	✓	✓	✓	✓
Offensive Rebounds Against	✓	✓	✓	✓
Offensive Rebound Percentage Against	✓	✓	✓	✓
Defensive Rebounds For	✓	✓	✓	✓
Defensive Rebound Percentage For	✓	✓	✓	✓
Defensive Rebounds Against	✓	✓	✓	✓
Defensive Rebound Percentage Against	✓	✓	✓	✓
Assists For	✓	✓	✓	✓
Assists Against	✓	✓	✓	✓
Steals For	✓	✓	✓	✓
Steals Against	✓	✓	✓	✓
Blocks For	✓	✓	✓	✓
Blocks Against	✓	✓	✓	✓
Turnovers For	✓	✓	✓	✓
Turnovers Against	✓	✓	✓	✓
Effective Field Goal Percentage For	✓	✓	✓	✓
Effective Field Goal Percentage Against	✓	✓	✓	✓
True Shooting Percentage For	✓	✓	✓	✓
True Shooting Percentage Against	✓	✓	✓	✓

Table 4.8: The statistics that were used to create the features for the European football/soccer model.

Statistic	Sum-based	Average-based	Slope-based	Projection-based
Wins	✓			
Loses	✓			
Draws	✓			
Goals For	✓	✓	✓	✓
Goals Against	✓	✓	✓	✓
Shots For	✓	✓	✓	✓
Shots Against	✓	✓	✓	✓
Shots On Target For	✓	✓	✓	✓
Shots On Target Against	✓	✓	✓	✓
Corners For	✓	✓	✓	✓
Corners Against	✓	✓	✓	✓
Corners On Target For	✓	✓	✓	✓
Corners On Target Against	✓	✓	✓	✓
Fouls For	✓	✓	✓	✓
Fouls Against	✓	✓	✓	✓
Yellow Cards For	✓	✓	✓	✓
Yellow Cards Against	✓	✓	✓	✓
Red Cards For	✓	✓	✓	✓
Red Cards Against	✓	✓	✓	✓
Passes For	✓	✓	✓	✓
Passes Against	✓	✓	✓	✓
Passes On Target For	✓	✓	✓	✓
Passes On Target Against	✓	✓	✓	✓
Free-kicks For	✓	✓	✓	✓
Free-kicks Against	✓	✓	✓	✓
Free-kicks On Target For	✓	✓	✓	✓
Free-kicks On Target Against	✓	✓	✓	✓

Chapter 5

Conclusion and Future Works

We believe that through the works described in chapters 3 and 4, we have achieved the initial research goal of creating a quantification method for sports based on the increase or decrease in the overall quality of play. This quantification method can be seen in the momentum-based features we propose, which are engineered using a team's linear trend of play over any desired performance indicator allowing us to extract both slope-based and projection-based features, thus quantifying the increase or decrease in their overall quality of play.

We have also achieved our second research goal of using this quantification method along with Machine Learning (ML) models to determine the predictive power of momentum and compare that to the predictive power of traditional frequency-based features. This can be seen in our comparison of the accuracy achieved amongst different game intervals, feature sets, and sports in chapters 3 and 4. This work led us to conclude that momentum can affect the outcomes of sporting events; however, this effect is small. While we found no evidence that momentum has an effect on outcomes in the NBA, we found that it does affect the outcomes of both the NHL and soccer/European football. We believe that due to the results seen in chapters 3 and 4, momentum plays a small role in game outcomes; however, it is not near the sole determinant of the outcome. Therefore, we believe the use of the combination of momentum-based features and frequency-based features is the way forward. As was seen in the use of random forest, more powerful machine learning techniques may favor the combination of the feature sets as it allows them to heavily weigh powerful features from each respective feature set. For this reason, as we have mentioned in the work above a pipeline that makes use of more powerful classification methods coupled with sufficient hyperparameter tuning and feature selection may create a more powerful pre-game prediction model for sporting events. As it will be able to capture both short-term trends of play and the standard of play over a period of time.

In the future, there is an opportunity to expand these momentum-based features by experimenting with features that quantify how well the trend line fits the games it is modeling or by experimenting with trend lines other than the linear line of best fit. Experimentation with combining frequency-based and momentum-based features together into a single feature is also an avenue worth exploring, such as combining the averaged-based and projection-based features for a given performance indicator into a single value.

These momentum-based performance indicators have opened up a new way to view momentum and how it affects pre-game prediction in sports. With the continued study of momentum and prediction in sports, we will be able to improve our understanding of how momentum affects teams and how to better quantify it.

Bibliography

[1] Louis Passfield and James G Hopker. A mine of information: can sports analytics provide wisdom from your data? *International journal of sports physiology and performance*, 12(7):851–855, 2017.

[2] Mahmoud M Nourayi. Profitability in professional sports and benchmarking: the case of nba franchises. *Benchmarking: An International Journal*, 13(3):252–271, 2006.

[3] Athalie J Redwood-Brown, Caroline A Sunderland, Antoinette M Minniti, and Peter G O'Donoghue. Perceptions of psychological momentum of elite soccer players. *International Journal of Sport and Exercise Psychology*, 16(6):590–606, 2018.

[4] Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, 17(3):295–314, 1985.

[5] Roger C Vergin. Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior*, 23(2), 2000.

[6] Kevin M Kniffin and Vince Mihalek. Within-series momentum in hockey: No returns for running up the score. *Economics Letters*, 122(3):400–402, 2014.

[7] Likang Ding, Ivor Cribben, Armann Ingolfsson, and Monica Tran. Do nhl goalies get hot in the playoffs? a multilevel logistic regression analysis. 2021.

[8] Gregory M Steeger, Johnathon L Dulin, and Gerardo O Gonzalez. 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, 2021.

[9] Jeremy Arkes and Jose Martinez. Finally, evidence for a momentum effect in the nba. *Journal of Quantitative Analysis in Sports*, 7(3), 2011.

[10] Andrew Bocskocsky, John Ezekowitz, and Carolyn Stein. The hot hand: A new approach to an old 'fallacy'. In *8th Annual MIT Sloan Sports Analytics Conference*, pages 1–10. Citeseer, 2014.

[11] Joshua B Miller and Adam Sanjurjo. Is it a fallacy to believe in the hot hand in the nba three-point contest? *European Economic Review*, 138:103771, 2021.

[12] Konstantinos Pelechrinis and Wayne Winston. The hot hand in the wild. *PloS one*, 17(1):e0261890, 2022.

[13] Brett Green and Jeffrey Zwiebel. The hot-hand fallacy: Cognitive mistakes or equilibrium adjustments? evidence from major league baseball. *Management Science*, 64(11):5315–5348, 2018.

[14] Michael J Lopez, Gregory J Matthews, and Benjamin S Baumer. 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, 2018.

[15] Whitney LaRow, Brian Mittl, and Vijay Singh. Predicting momentum shifts in nba games. 2014.

[16] Jordan TP Noel, Vinicius Prado da Fonseca, and Amilcar Soares. The use of momentum-inspired features in pre-game prediction models for the sport of ice hockey. *International Journal of Computer Science in Sport*, 23(1):1–21, 2024.

[17] Jordan Truman Paul Noel, Vinicius Prado da Fonseca, and Amilcar Soares. A comprehensive data pipeline for comparing the effects of momentum on sports leagues. *Data*, 9(2):29, 2024.

[18] Jim Taylor and Andrew Demick. A multidimensional model of momentum in sports. *Journal of Applied Sport Psychology*, 6(1):51–70, 1994.

[19] Martin I Jones and Chris Harwood. Psychological momentum within competitive soccer: Players' perspectives. *Journal of Applied Sport Psychology*, 20(1):57–72, 2008.

[20] Ulf Johansson, Erik Wilderoth, and Arsalan Sattari. How analytics is changing ice hockey. In *Linköping Hockey Analytics Conference*, pages 49–59, 2022.

[21] Stephen Shmanske and Franklin Lowenthal. Overtime incentives in the national hockey league (nhl) more evidence. *Journal of Sports Economics*, 8(4):435–442, 2007.

[22] Benjamin Leard and Joanne M Doyle. The effect of home advantage, momentum, and fighting on winning in the national hockey league. *Journal of Sports Economics*, 12(5):538–560, 2011.

[23] Andrew Vesper. Putting the "hot hand" on ice. *Chance*, 28(2):13–18, 2015.

[24] Stephanie Parsons and Nicholas Rohde. The hot hand fallacy re-examined: new evidence from the english premier league. *Applied Economics*, 47(4):346–357, 2015.

[25] Ruud JR Den Hartigh, Paul LC Van Geert, Nico W Van Yperen, Ralf FA Cox, and Christophe Gernigon. Psychological momentum during and across sports matches: Evidence for interconnected time scales. *Journal of Sport and Exercise Psychology*, 38(1):82–92, 2016.

[26] Namita Nandakumar and Shane T Jensen. Historical perspectives and current directions in hockey analytics. *Annual review of statistics and its application*, 6:19–36, 2019.

[27] Michael Mondello and Christopher Kamke. The introduction and application of sports analytics in professional sport organizations. *Journal of Applied Sport Management*, 6(2):11, 2014.

[28] Shayna Goldman. Analytics in the nhl: Where it's at and where it's going next, Jun 2022.

[29] Josh Weissbock and Diana Inkpen. 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*, pages 251–262. Springer, 2014.

[30] Joshua Weissbock, Herna L Viktor, and Diana Inkpen. Use of performance metrics to forecast success in the national hockey league. In *MLSA@ PKDD/ECML*, pages 39–48, 2013.

[31] Sebastian Remander et al. Comparison of classification methods in predicting nhl game event outcomes. 2021.

[32] Gianni Pischedda. Predicting nhl match outcomes with ml models. *International Journal of Computer Applications*, 101(9), 2014.

[33] Wei Gu, Krista Foster, Jennifer Shang, and Lirong Wei. A game-predicting expert system using big data and machine learning. *Expert Systems with Applications*, 130:293–305, 2019.

[34] Daniel E Gilbert and Martin T Wells. Ludometrics: luck, and how to measure it. *Journal of Quantitative Analysis in Sports*, 15(3):225–237, 2019.

[35] Daniel Kari et al. Understanding how random chance affects the outcome of an ice hockey game. 2020.

[36] *American Gaming Association*, 2023.

[37] Ondřej Hubáček, Gustav Šourek, and Filip Železný. Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2):783–796, 2019.

[38] Joe Osborne. Which sport do underdogs win in the most often?, Jun 2020.

[39] Harry Shomer. Hockey_scraper, 2019.

[40] Peter Tanner. Moneypuck.com.

[41] Natural Stat Trick. naturalstattrick.com.

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[43] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in Statistics: Methodology and Distribution*, pages 196–202. Springer, 1992.

[44] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

[45] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[46] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[47] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[48] Eli Ben-Naim, Federico Vazquez, and Sidney Redner. Parity and predictability of competitions. *Journal of Quantitative Analysis in Sports*, 2(4), 2006.

[49] Susmita Ray. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE, 2019.

[50] Marius Ötting, Christian Deutscher, Carl Singleton, and Luca De Angelis. Gambling on momentum. *arXiv preprint arXiv:2211.06052*, 2022.

[51] Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. A markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, 31:1735–1757, 2017.

[52] Dawson Sprigings and Asmae Toumi. Expected goals are a better predictor of future scoring than corsi, goals., 2015.

[53] Michael Lewis. *Moneyball: The art of winning an unfair game*. WW Norton & Company, Manhattan, New York, 2004.

[54] Joshua B Miller and Adam Sanjurjo. A visible (hot) hand? expert players bet on the hot hand and win. *Expert Players Bet on the Hot Hand and Win (December 15, 2017)*, 2017.

[55] David M Ritzwoller and Joseph P Romano. Uncertainty in the hot hand fallacy: Detecting streaky alternatives to random bernoulli sequences. *The Review of Economic Studies*, 89(2):976–1007, 2022.

[56] Joshua B Miller and Adam Sanjurjo. Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86(6):2019–2047, 2018.

[57] Michael J Fry and F Alan Shukairy. Searching for momentum in the nfl. *Journal of Quantitative Analysis in Sports*, 8(1), 2012.

[58] Carson K Leung and Kyle W Joseph. Sports data mining: predicting results for the college football games. *Procedia Computer Science*, 35:710–719, 2014.

[59] Zheng Wang, Cheng Long, and Gao Cong. Similar sports play retrieval with deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[60] Noppon Wongta and Juggapong Natwichai. Data pipeline of efficient stream data ingestion for game analytics. In *Advances in Internet, Data & Web Technologies: The 11th International Conference on Emerging Internet, Data & Web Technologies (EIDWT-2023)*, pages 483–490. Springer, 2023.

[61] Noppon Wongta and Juggapong Natwichai. End-to-end data pipeline in games for real-time data analytics. In *Advances in Internet, Data and Web Technologies: The 9th International Conference on Emerging Internet, Data & Web Technologies (EIDWT-2021)*, pages 269–275. Springer, 2021.

[62] Fadi Thabtah, Li Zhang, and Neda Abdelhamid. Nba game result prediction using feature analysis and machine learning. *Annals of Data Science*, 6(1):103–116, 2019.

[63] Fátima Rodrigues and Ângelo Pinto. Prediction of football match results with machine learning. *Procedia Computer Science*, 204:463–470, 2022.

[64] Jonathan J Koehler and Caryn A Conley. The “hot hand” myth in professional basketball. *Journal of sport and exercise psychology*, 25(2):253–259, 2003.

[65] Fabian Wunderlich, Alessandro Seck, and Daniel Memmert. The influence of randomness on goals in football decreases over time. an empirical analysis of randomness involved in goal scoring in the english premier league. *Journal of Sports Sciences*, 39(20):2322–2337, 2021.

[66] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

[67] Justin Kubatko, Dean Oliver, Kevin Pelton, and Dan T Rosenbaum. A starting point for analyzing basketball statistics. *Journal of quantitative analysis in sports*, 3(3), 2007.

[68] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[69] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[70] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[71] Duane W Rockerbie. Exploring interleague parity in north america: The nba anomaly. *Journal of Sports Economics*, 17(3):286–301, 2016.