Problem Chosen	2024 MCM/ICM Summary Sheet
C	

Summary

This study delves into the quantification and impact of psychological momentum (PM) in Men's Elite Tennis, utilizing a novel model to numerically evaluate players' psychological states from match scoring patterns. Building upon Adler's and Taylor and Demick's foundational work, momentum is understood as a complex interplay of cognitive, affective, physiological, and behavioral changes triggered by specific events, leading to alterations in competitive performance. However, existing models, while providing a thorough conceptual framework, lack in quantitatively analyzing psychological momentum, presenting a gap this research aims to bridge. Our goal is to develop a model that quantifies a player's psychological state through match scoring patterns, offering a new lens to predict and understand momentum shifts in tennis.

We build our momentum model using Machine Learning Linear Regression. It is a pioneering method for quantifying psychological momentum, revealing a strong correlation between momentum metrics and match outcomes. The graphical representation of momentum quantification displays a clear pattern of momentum surges preceding winning streaks, indicating the pivotal role of psychological momentum in enhancing performance and influencing match dynamics.

The correlation analysis is carried out using hypothesis test and Pearson Coefficient, which further validates the impact of psychological momentum, underscoring the relationship between match play swings and momentum shifts. Additionally, the model's performance, as detailed by the confusion matrix and feature importance analysis, emphasizes the critical role of point sequences in momentum dynamics, offering strategic implications for players and coaches.

Moreover, the predictive model based on the Random Forest algorithm, validated using the 2023 Wimbledon dataset, provides deep insights into momentum's intricacies, highlighting the importance of consecutive point wins and strategic play in securing momentum.

The implications of our findings for player strategy and coaching are profound. The emphasis on consecutive point wins as a momentum driver suggests a strategic focus on consistency and minimizing errors. The challenges in predicting neutral momentum states open avenues for developing strategies to exploit these critical junctures, potentially altering match outcomes. This analysis provides a data-driven foundation for strategic decision-making, encouraging aggressive play balanced with risk management.

Our model's generalizability across various sports underscores its potential in sports psychology and coaching, offering a versatile tool for predicting match outcomes based on observable variables and psychological momentum. Its adaptability and accuracy across different sports demonstrate the model's utility in enhancing understanding and strategic planning in competitive settings.

Contents

1 Introduction

In professional tennis, the concept of momentum and its impact on the flow of play captivates players, coaches, and fans alike. Often perceived as an elusive force, momentum, generated through a series of events, plays a pivotal role in determining match outcomes, despite being outside direct player control.

Building on the studies by Adler[1], Taylor and Demick [10], momentum is defined as a dynamic shift in cognition, affect, physiology, and behavior triggered by certain events, leading to corresponding changes in performance and competitive outcomes. Taylor and Demick's Multidimensional Model of Momentum in sports offers a comprehensive framework for understanding momentum's development through a "momentum chain," enhancing our grasp of this phenomenon in a sporting context.

However, the model's qualitative approach falls short in providing a quantitative analysis of a player's psychological momentum during a match. Moreover, the subjective nature of momentum, encompassing factors like affective changes and variations in self-efficacy, poses challenges for measurement. Taylor and Demick caution against intrusive methods such as athlete interviews during competitions, which may disrupt the natural competition flow and momentum-related cognitive and emotional processes.

This research aims to quantitatively assess psychological momentum in Men's Elite Tennis by developing a model that evaluates a player's psychological state based on match scoring patterns. By numerically visualizing psychological momentum, our model captures the match's flow of play, enabling predictions of momentum shifts from one player to another. We argue that these shifts are not merely random but are influenced by the player's psychological momentum at those moments.

1.1 Literature Review

Research on psychological momentum (PM) in sports has evolved significantly, with seminal studies paving the way for a deeper understanding of its effects on performance. Vallerand et al's [11] exploration of the antecedents-consequences (AC) model marked a pivotal advancement in distinguishing between perceptions of PM and its underlying causes and effects. This model proposes that perceptions of PM, shaped by personal and situational factors, directly influence performance outcomes, with these relationships moderated by individual and contextual variables. Their empirical work, employing tennis scenarios to manipulate PM perception, lends substantial support to the AC model, demonstrating how specific conditions can enhance PM perceptions and subsequently affect expectations of match outcomes.

Taylor & Demick [10] introduced a theoretical model delineating a "momentum chain," which begins with an event that triggers changes across cognitive, physiological, affective, and behavioral domains, ultimately impacting performance. This model uniquely accounts for the competitive interplay, suggesting that momentum gained by one player inherently implies a loss of momentum for the opponent, thereby directly influencing the competition's immediate outcome. Their preliminary research provides initial validation of the model, emphasizing the necessity of simultaneous positive and negative momentum shifts among competitors for a notable impact on outcomes.

Hughes et al.[5] extended this theoretical framework by creating 'momentum profiles' through

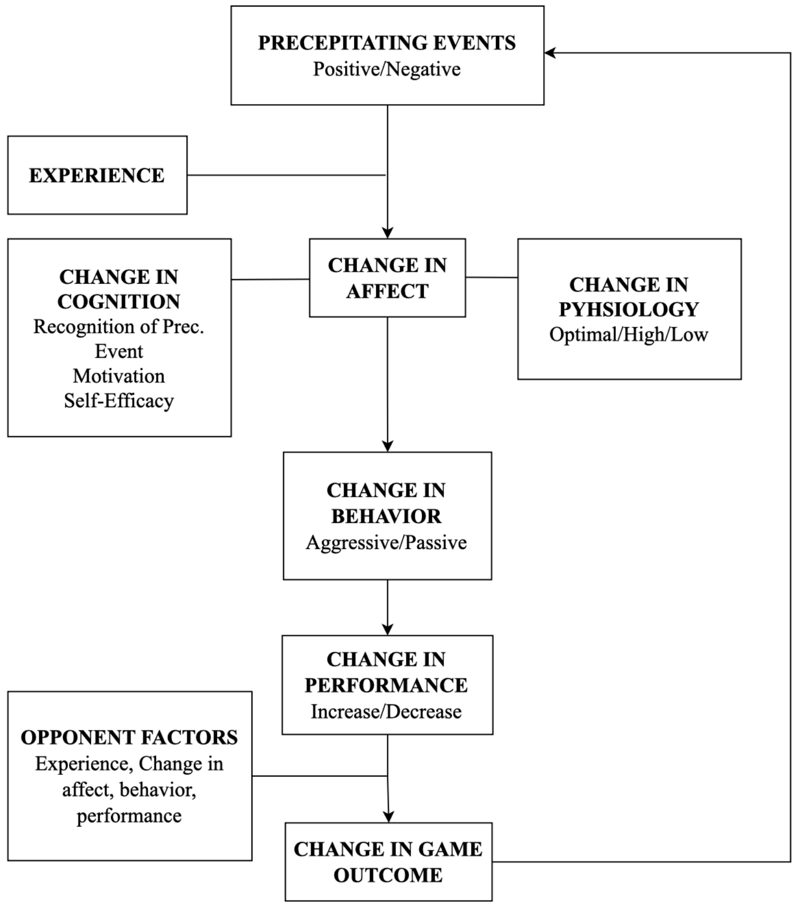

Figure 1: Multidimensional model of momentum in sports (adapted from Taylor & Demick, [10]).

notational analysis, identifying positive momentum with scoring streaks and negative momentum with errors. However, this approach may oversimplify the intricate psychological dynamics at play, underscoring the need for more nuanced analytical methods.

Cui's [3] findings on the significant impact of game location and rally length on winning percentages further complicate the landscape of PM research, indicating that both external factors and player performance under various conditions contribute to overall success.

Notably, Klaassen and Magnus [8] tested whether points in tennis are independent and identically distributed (iid) by modeling the probability of winning a point on service demonstrates that points in a game are not independent or identically distributed: securing the previous point positively influences the chances of winning the current point, and during "critical" points, it becomes harder for the server to win compared to less critical moments.

Contrastingly, research by Klaassen and Magnus [8] , and O'Donoghue & Brown [9], presents a divergent view, challenging the notion of PM in tennis by examining the independence and distribution of points, and questioning the perceived existence of momentum in match play.

Studies in other sports, like Ito & Miyamoto's investigation into Japanese professional baseball, underscore the potential of quantitative models to capture and predict the dynamics of sports competitions, illustrating the broader applicability and significance of PM research across different athletic domains.

Researches in other competitive sports also shed lights on the phenomenon of psychological momentum. Ito & Miyamoto [6] investigates the concept of momentum within a Japanese professional baseball game, using data from the 2015 season to develop a model that quantitatively analyzes teams' within-game momentum. By applying this model to 2016 season games, the research successfully predicted game streak changes, demonstrating the model's efficacy in capturing the dynamic nature of sports competitions and offering insights into the factors that influence game outcomes.

2 Method

Our methodology integrates a comprehensive bibliographic review to identify key factors influencing a player's psychological momentum in sports. These potential influencers were subjected to a machine learning (ML) regression analysis to assess their significance in determining immediate game outcomes. Following the approach pioneered by Ito & Miyamoto [6], we developed a numerical model to represent psychological momentum, categorizing it into nine distinct states ranging from strong negative to strong positive momentum. The "flow of play" is conceptualized as the variance in psychological momentum between the competing players. Moreover, to determine the correlation between momentum and the game features, we use hypothesis testing with Pearson Coefficient to validate the correlation of momentum with runs of success and swings in play.

Given the categorical representation of psychological momentum within our model, we opted for the Random Forest algorithm, renowned for its proficiency in managing multi-class classification challenges. This choice was informed by the algorithm's robustness in handling complex datasets and its ability to improve predictive accuracy through the aggregation of multiple decision trees, minimizing the risk of overfitting.

2.1 Feature Engineering

We divide the potential influencer into two categories: positive influencer and negative influencer.

Potential Positive Influencers

(1) *Consecutive point won*. Increasing consecutive points won can significantly boost a player's confidence and momentum, putting psychological pressure on the opponent. Higher numbers indicate stronger current game dominance, potentially leading to an increased rate of unforced errors from the opponent (Klaassen & Magnus, [8]; Knight & Donoghue, [9]).

(2) *Serving*. Serving is a crucial factor in tennis because it is the only moment a player has complete control over the point's initiation, allowing them to dictate the pace and style of play from the outset. Averaging over all players, the probability of serve win is estimated as 0.645 (with a standard error of 0.002) in the men's singles (Klaassen & Magnus, [8]). However, to precisely measure the momentum of a player, we need to alleviate the fact that serving will significantly increase the chances of winning a point or a game.

(3) *Untouchable shot*. The ability to hit shots that the opponent cannot reach demonstrates superior skill and strategy. Higher levels indicate a player's effectiveness in creating and exploiting openings, leading to easy points and disrupting the opponent's rhythm (Cui et al., [3]).

(4) *Backhand won*. In tennis, backhand swing is considered as a weaker and more defensive swing type compared to a forehand swing. Therefore, in a point, if the player hit a backhand shot and results in the winning of the point, it is considered as confidence boost because the player scored in a difficult pose, making this a positive influencer for momentum(Croucher, [2]).

(5) *Ace*. Serving aces directly wins points without rally exchange, showcasing serve dominance. More aces indicate a powerful serve, contributing to winning games more easily by reducing the opponent's chances to initiate play (Hughes et al., [5]).

Potential Negative Influencers

(6) *Consecutive point lost*. Losing consecutive points can negatively impact a tennis player's momentum by eroding their confidence and increasing mental pressure, making it harder to execute strategies effectively and maintain composure (Sim et al., 2020; Jeon & Juyong, [7])

(7) *Unforced error*. Unforced errors can similarly disrupt momentum, as they not only gift points to the opponent but also lead to frustration and self-doubt, potentially affecting a player's focus and decision-making in subsequent points (Filipčič, [4]).

(8) *Double faults*. Double faults means directly conceding points to the opponent without any exchange, highlighting vulnerabilities in the serve and increasing psychological stress. This can lead to a loss of confidence in one of the game's most crucial aspects, the serve, making it difficult for the player to recover mentally and affecting their overall performance in the match (Taylor & Demick, [10]).

Variable	**Influencer**	**Weight**
x_{CPw}	Consecutive Point Won	0.1680
x_S	Serving	0.6270
x_{US}	Untouchable Shot	0.0009
x_{BP}	Backhand Point	0.0164
x_{CPl}	Consecutive Point Lost	-0.0317
x_{UE}	Unforced Error	-0.0171

Table 1: Influencer Relevance

With these potential influencers, we use a Machine Learning Regression method to study their relevance to the current game outcome. Through machine learning regression analysis, we investigated the influence of various scoring patterns on immediate game outcomes, identifying those most indicative of a player's psychological momentum. This analysis revealed a significant correlation between certain patterns, such as consecutive points won, and the outcome of the game, underscoring their value as predictors of psychological momentum. Conversely, other patterns previously thought to be influential, like executing an untouchable shot, were found to have negligible impact on game outcomes. This suggests that while some actions are pivotal in shifting momentum, others, contrary to initial assumptions, do not significantly affect the flow of play. Our findings challenge conventional beliefs about momentum triggers, indicating the need for a nuanced understanding of what truly influences psychological momentum and its consequent effect on match dynamics.

2.2 Psychological Momentum

Taylor & Demick (1994) showed that previous performance results are capable of instantiating a momentum chain reaction. Therefore, we assume that positive and negative scoring patterns will influence a player's psychological momentum. In particular, we postulate that the scoring patterns till time $t-1$, P_{t-1}, influences the player's momentum at time t, M_t, where ideally we want to find an evaluation function that gives M_t based on P_{t-1},

$$f(P_{t-1}) = M_t$$

Meanwhile, the player's momentum at time t will influence the game outcome O_{t+1} (represented as the probability of winning the current game) at the next moment, as reflected by the score at time $t+1$,

$$g(M_t) = O_{t+1}$$

Given that the scoring pattern is cumulative results of the scores O_t at each moment, our model agrees with Taylor & Demick's momentum chain mechanism that momentum and game outcome are interactive. However, given the limitation that psychological momentum is not directly observable, it is impossible to create a regression model that simulates f or g. Instead, we use a regression model that simulate $g \circ f$ using the scoring pattern at time $t-1$ and the resulting game output at time $t+1$ to see how much these potential influencers are contributing positively or negatively on the game outcome. When determining the outcome of a game, the target variable is defined as a dummy variable: it is assigned a value of "1" for a win and "0" for any other outcome. Similarly, in scenarios where the game is lost, the target variable also adopts a value of "1" for a loss, and "0" for otherwise.

Subsequently, we obtain the desired mechanism f by composing $g \circ f$ with g^{-1}, where $g^{-1}$is the explanatory function of how the score at time $t+1$ is influenced by the player's psychological momentum at time t. Factors that positively contribute to the player's momentum will thereby contribute to winning the next point, and factors that negatively influence the player's momentum will thereby be held culpable for losing the next point. A high probability of winning the next point indicates a positive momentum, while a high probability of losing the next point indicates a negative momentum,

$$g^{-1}(O_{t+1}) = M_t$$

We then construct the Psychological Momentum Model (PMM) by scoring pattern as

$$PM = g^{-1} \circ g \circ f(P_{t-1}) = \frac{1 + g \circ f(P^+_{t-1})}{1 + g \circ f(P^-_{t-1})} = \frac{1 + \sum \omega_i X_i^+}{1 + \sum \nu_i X_i^-}$$

where X_i^+ are positive influencers and X_i^- are negative influencers. The model will produce a positive value, and when performances do not impact the momentum, the outcome of the model will approach to 1. If there are numerous instances of performances that have a positive effect on flow, the outcome will exceed 1. Conversely, if there are many instances of performances that adversely affect flow, the outcome approach to 0.

Finally, we use the filtered influencers relevant to the game outcome listed above in Table 1 as the scoring pattern variables. For example, the serving variable is removed because momentum should be calculated irrelevant to who's serving. We replace the weight ω_i and ν_i in the formula for PM with the respective weight of each influencer, where we take absolute value for negative weight in the denominator, and we have the formula to quantify a player's psychological momentum at any given time t based on the historical scoring pattern:

$$PM = \frac{1 + (0.1680x_{CPw} + 0.0009x_{US} + 0.0164x_{BP})}{1 + (0.0317x_{CPl} + 0.0171x_{UE})}$$

2.3 Flow of Play Prediction Model

The Random Forest model, tailored for psychological momentum analysis in tennis, leverages engineered features reflecting point differentials, psychological states, and match contexts, distributed across 9 categories. Stratified sampling guarantees balanced training data representation. Hyper-parameter optimization—number of trees, tree depth, and node split criteria—was conducted via grid search with cross-validation, aiming to enhance prediction accuracy across match stages. This approach underscores a meticulous balance between model complexity and predictive capability, targeting precise momentum state identification.

3 Result

3.1 Visualization of Flow of Play

The graphical analysis distinctly demonstrates a robust correlation between quantified psychological momentum and match outcomes, affirming our hypothesis on the pivotal role of psychological momentum in determining match results. Notably, the visualization reveals a pattern where surges in the quantified momentum (illustrated by peaks in the blue line) consistently precede sequences of successful outcomes (indicated by green markers). This observation suggests that a rise in psychological momentum contributes to a player's winning streak, reflecting an enhancement in confidence and performance. Such momentum gains not only boost the player's morale but also appear to impact the mental state of the opponent, further influencing the match dynamics.

The graph shows multiple instances of momentum shifts where the blue line crosses the zero mark. These crossings are pivotal, possibly denoting critical moments such as successful serve breaks or key rallies won, emphasizing the importance of recognizing and quantifying such shifts.

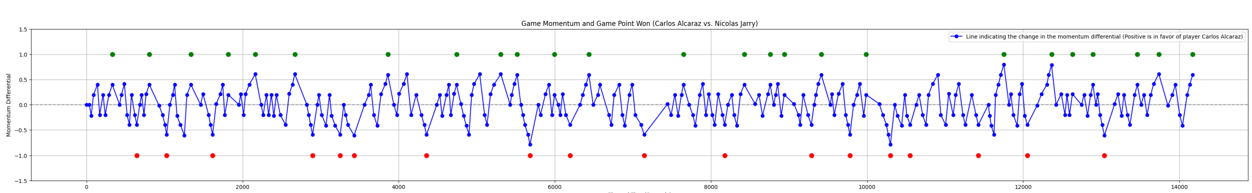

Figure 2: Visualization of the Flow of Play, Carloz Alaraz vs Nicolas Jarry.

Understanding these moments is key to analyzing the ebb and flow of psychological advantage throughout the competition, providing insights into how players navigate and influence the match dynamics through gains and losses in momentum.

3.2 Correlation Analysis

To investigate into this problem, we calculated the Pearson Correlation between ΔM, the difference of momentum of two players in the game with RoS_t, runs of success, and SoP_t, swings in play. We first derived a mathematical definition of the "swings in play" and the "runs of success".

$$RoS_t = \max(\text{Times of Consecutive Scoring})_{t:t+3}$$

RoS_t is the runs of success of player 1 at given moment t. It is the maximum times of consecutive scoring in the following next three time. Therefore, the higher ΔM, the higher the max number of consecutive scoring in the future.

$$SiP_t = \sum CP_t - \sum CP_{t-1}$$

SiP_t is the swings in play in the game with respect to player 1 at given moment t. $\sum CP_t$ is the net consecutive of player 1 at given moment t. $\sum CP_{t-1}$ is the net consecutive of player 1 in the previous point. Swings in play is the shifts or changes in the dynamics of a match. For example, one player is dominating the match and suddenly the momentum shifts, allowing the opponent to start winning points. Therefore, we decided to use the difference between net consecutive points in different points in time to denote swings in play.

We used the Pearson Correlation to validate whether there swings in play or runs of success are random. For runs of success, the null hypothesis is:

H_o : There is no correlation between runs of success for a player and their momentum difference with the opponent in the match.

H_A : There is a correlation between runs of success for a player and their momentum difference with the opponent in the match.

Parameter	Value
Pearson Coefficient	0.689
p-value	6.64×10^{-34}

Table 2: Correlation between Momentum Differential and Runs of Success.

Parameter	Value
Pearson Coefficient	0.566
p-value	3.66×10^{-254}

Table 3: Correlation between Momentum Differential and Swings in Play.

In table 2, we see a Pearson Coefficient of 0.689, and a p-value of 6.64×10^{-34}. Therefore, we reject the null hypothesis of there is no correlation between the runs of success in a match and the momentum differential.

For swings in play, the null hypothesis is:

H_o : There is no correlation between the swings of play in a match and the swings in the momentum differential.

H_A : There is a correlation between the swings of play in a match and the swings in the momentum differential.

If p-value < 0.05, we reject the null hypothesis at a confidence level of 95%. In table 3, we see a Pearson Coefficient of 0.566, and a p-value of 3.66×10^{-254}. Therefore, we reject the null hypothesis of there is no between the swings of play in a match and the swings in the momentum differential. Therefore, we have proved that momentum plays a role in the match, resulting to changes in runs of success and swings in play.

3.3 Flow of Play Prediction

The predictive analysis using a Random Forest model, applied to the 2023 Wimbledon Men's Singles Tennis dataset, provides profound insights into the dynamics of momentum swings. The effectiveness of the model is showcased through graphical representations, illuminating key aspects of match play and strategic considerations essential for both coaches and players. This analysis underscores the model's capability to decipher complex patterns of momentum shifts, offering a valuable tool for enhancing understanding and strategic planning in competitive tennis.

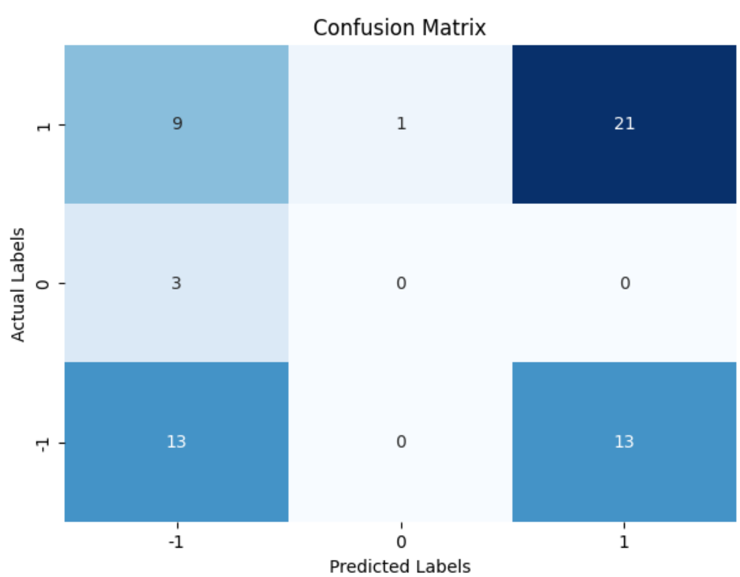

Figure 3: The Confusion Matrix for Random Forest Prediction Model.

The confusion matrix provides a granular view of the model's classification prowess. The accurate predictions for extreme classes (-1 and 1) indicate the model's efficacy in discerning clear swings in match momentum. The complete absence of predictions for the neutral class (0) may indicate either a class imbalance within the dataset or a model insensitivity to nuanced, non-dominant patterns of play that do not result in immediate momentum shifts. It reflects the difficulty of the model to predict future change in the flow of play when there is no obvious momentum advantage for either of the two players, i.e. the psychological effect of momentum is neutral. This also suggests that future model iterations might benefit from a re-balancing of the dataset or an exploration into more sophisticated models that can capture subtler shifts in momentum.

The delineation of feature importances provides an invaluable insight into the match elements that drive momentum. The preeminence of the number of consecutive win and the number of consecutive loss underpins the intuitive notion that momentum is intrinsically linked to sequences of successful point acquisition or loss. This affirms the tactical emphasis on sustaining performance across consecutive plays and minimizes the occurrence of errors. The relative importance of direct point-winning features (e.g., ace, serving) compared to double fault accentuates a strategic preference for assertive play over conservative strategies aimed merely at avoiding errors.

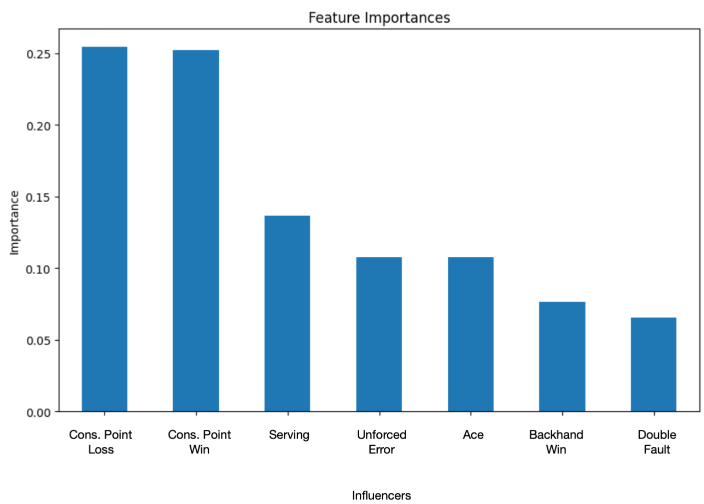

Figure 4: The relative feature importance in the Random Forest Prediction Model.

3.4 Robustness Test

In the formation of the momentum model, we used the first 13 games of whole data. Therefore, to carry out the robustness test, we decided to apply the model to a match not in the training set for analysis. We chose the match of Alexander Bublik and Maximilian Marterer. Figure 5 is the visualization of the momentum differential.

From the Figure 5, we see that the flow of play still follows the game points in the match, meaning that when the momentum differential is favoring one player, the player will win the game. Also, the momentum returns to zero after a game because a new game is starting. Therefore, the momentum model is applicable to other matches.

For testing the correlation between runs of success and momentum differential. We initially defined runs of success as the maximum times of consecutive wins in the next three points. We

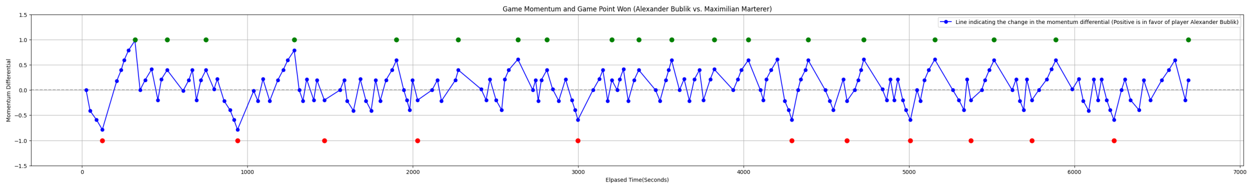

Figure 5: Robust analysis for Psychological Momentum Quantification.

can also increase the standard by changing next three points to next four points, making it to make a more long-term prediction. From Table 4, the new Pearson Correlation Coefficient is 0.558, lower than the original, but the p-value suggests that the correlation is still statistically significant. Therefore, the momentum differential is still valid in predicting runs of success in a long-term scale.

Parameter	Value
Pearson Coefficient	0.558
p-value	2.69×10^{-24}

Table 4: Correlation between Momentum Differential and Runs of Success.

4 Discussion

4.1 Implications for Player Strategy and Coaching

The analytical results from the Random Forest model implicate a strategic focus on the accumulation of consecutive points as a primary driver for securing match momentum. Coaches and players may infer that strategies and training that enhance consistency and reduce unforced errors can be crucial in altering the match dynamics in favor of the player.

Furthermore, the apparent prediction difficulties surrounding neutral momentum states suggest an opportunity for players to develop specialized strategies to capitalize on these pivotal moments. The ability to shift the momentum in one's favor during these critical junctures may prove to be a decisive factor in match outcomes. In light of the differential impact of various match-play elements, players may be advised to engage in aggressive, point-winning plays, particularly against opponents with differing play styles. Such a strategy should be balanced with a focus on minimizing streaks of point losses to prevent adverse momentum swings.

In summary, this model offers a quantified testament to the tactical aspects that significantly influence match outcomes. The implications drawn from this analysis provide a data-driven foundation for strategic decision-making in high-stakes match play.

4.2 Generalization

Our model is designed to accurately predict momentum swings in a variety of competitive sports with numerical scoring systems, such as tennis, badminton, squash, boxing, and fencing. By utilizing a regression framework, it integrates a wide range of observable variables from environmental conditions to pre-game statistics, enhancing its predictive accuracy across different matches and conditions.

The model's adaptability was tested across various sports, demonstrating a high level of accuracy in predicting match outcomes. It is particularly effective in sports with rapid score changes, such as table tennis, due to the rich data these sports provide. However, the model's performance can vary based on the sport and available data, suggesting the need for continuous refinement and the inclusion of additional variables like psychological factors and fatigue levels.

For non-confrontational sports like swimming and shooting, where psychological momentum is a direct performance influencer, our model shows even greater potential. The straightforward impact of psychological elements in these sports aligns well with our regression model, promising enhanced predictive capability.

4.3 Area of Future Research

The quantification of psychological momentum is a complex task, and while the current method aligns closely with game outcomes, it may not capture all nuances of a player's psychological state. Further research could involve more granular data collection, including biometric data, to enhance the model's accuracy. Our research did not encompass variables such as the game/set period, game status, competitive load, or match congestion. These factors could potentially influence the psychological momentum experienced by players and, consequently, the outcomes of matches. The game/set period might affect a player's psychological state differently at various stages of the match, while the game status (e.g., leading or trailing) could alter the intensity of psychological momentum.

Furthermore, the competitive load, referring to the frequency and intensity of matches played over a period, and match congestion, indicating the scheduling density of competitions, may also impact a player's psychological resilience and momentum. The exclusion of these variables from our analysis means that the study's insights into the nuances of psychological momentum in tennis are somewhat constrained. Future research should consider incorporating these factors to provide a more comprehensive understanding of the dynamics at play.

5 Conclusion

Our research offers significant insights into the role of psychological momentum in Men's Elite Tennis, quantitatively capturing its influence on match outcomes. Based on past research in relevant areas, we used Machine Learning Regression method to find scoring patterns that will influence the player's psychological momentum, and used the historical scoring patterns to numerically generate the momentum. Statistical analysis we performed showed that certain scoring patterns in Tennis matches, commonly referred to as s̈wings" and r̈un of success" are indeed correlated with the player's psychological momentum, not simply random phenomenon or illusion. By employing a novel predictive model that integrates scoring patterns with psychological momentum, we have demonstrated the capacity to anticipate shifts in the flow of play. Our findings underscore the profound impact of psychological factors on athletic performance, challenging conventional understandings of momentum in sports. Furthermore, the predictive accuracy of our model across various match stages and conditions suggests its broader applicability in sports psychology and coaching strategies.

References

[1] Peter Adler and Patricia A. Adler. The role of momentum in sport. *Urban Life*, 7(2):153–175, 1978.

[2] J.S. Croucher. The conditional probability of winning games of tennis. *Research Quarterly for Exercise and Sport*, 57:23–26, 1986.

[3] Yixiong Cui et al. Data-driven analysis of point-by-point performance for male tennis player in grand slams. *Motricidade*, 15(1):49–61, 2019.

[4] Tjaša Filipčič, Aleš Filipčič, and Tomaž Berendijaš. Comparison of game characteristics of male and female tennis players at roland garros 2005. *Acta Universitatis Palackianae Olomucensis. Gymnica*, 38(3), 2008.

[5] Mike Hughes et al. *Essentials of Performance Analysis in Sport*. Routledge, 2015.

[6] Akihiro Ito and Michiko Miyamoto. Proposing a model to catch the momentum of games: Visualization of momentum in japanese professional baseball. page 9, 7 2019. July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios.

[7] Gyuhyeon Jeon and Juyong Park. Characterizing patterns of scoring and ties in competitive sports. *Physica A: Statistical Mechanics and its Applications*, 565:125544, 2021.

[8] F.J.G.M. Klaassen and J.R. Magnus. Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96:500–509, 2001.

[9] G. Knight and P. O'Donoghue. The probability of winning break points in grand slam men's singles tennis. *European Journal of Sport Science*, 12:462–468, 2012.

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

[11] Robert J. Vallerand, Pasquale G. Colavecchio, and Luc G. Pelletier. Psychological momentum and performance inferences: A preliminary test of the antecedents-consequences psychological momentum model. *Journal of Sport and Exercise Psychology*, 10(1):92–108, 1988.

Appendices

Appendix A Memo

To the Coaching Staff,

I hope this letter finds you well. Following our comprehensive study on psychological momentum in Men's Elite Tennis, I wish to share with you some pivotal insights and recommendations that could significantly enhance our coaching strategies and player performance.

Our research has illuminated the profound impact that psychological momentum has on the outcomes of tennis matches. We have developed a predictive model that not only captures this dynamic but also offers a framework for anticipating shifts in the flow of play. This model underscores the importance of mental resilience and strategic adaptation in securing a competitive edge.

In light of our findings, I advocate for a coaching approach that emphasizes mental fortitude and strategic awareness among our players. It is crucial that we equip them with the skills to recognize and capitalize on scoring patterns indicative of momentum shifts. Moreover, integrating simulation drills that mimic high-pressure situations could profoundly benefit our players, enhancing their ability to remain composed and strategically agile under duress.

Furthermore, the effective use of timeouts and match breaks can serve as critical tools in managing the game's psychological aspects. These moments should be optimized for mental recovery and strategic planning, potentially disrupting an opponent's momentum.

Encouraging a culture of flexibility and adaptability will be key to navigating the unpredictable nature of match play successfully. Our players need to be prepared to adjust their game plans in response to their current psychological state and that of their opponents.

In conclusion, the insights gleaned from our study offer a valuable blueprint for fostering a resilient and strategically savvy team. By incorporating these recommendations into our coaching paradigm, I am confident we can achieve a marked improvement in our players' performance and overall match outcomes.

Thank you for considering these insights. I look forward to discussing them further and integrating these strategies into our coaching regimen.

Warm regards,

Your Friends

Appendix B Program

Here are the modeling program codes we used in our model.

Machine Learning Regression Analysis for the momentum model:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
from sklearn.linear_model import LinearRegression

data = pd.read_csv('modified_file.csv')
selected_columns = ['max_con_point','serve','cum_untouchable','cum_ace','cum_b
    ↪ ','max_con_point_loss','cum_unf_err']
X = data[selected_columns]
y = data['game_won']
print(type(X))

model = LinearRegression()

# Fit the model on the data
model.fit(X, y)

# Display the coefficients and intercept
print("Coefficients:", model.coef_)
print("Intercept:", model.intercept_)
# Predict the target values
y_pred = model.predict(X)

# Convert the predictions to integers (classification)
y_pred_class = [round(pred) for pred in y_pred]
accuracy = accuracy_score(y, y_pred_class)

print("Accuracy:", accuracy)
```

Code for Visualization in match momentum differential:

```
import pandas as pd
from datetime import datetime
import matplotlib.pyplot as plt
# Load the CSV file into a DataFrame
df = pd.read_csv('Wimbledon_featured_matches_copy.csv')

def cal_momentum(con_win, untouch, back, con_loss, unf_err ):
  b_0 =  0.009394573521668753;
  b_1 = 0.16845085;
  b_2 = 0.00091111;
  b_3 = 0.0163816;
  b_4 = 0.03170258;
  b_5 = 0.017100571;
  result = (1 + b_1 * con_win + b_2 * untouch + b_3 * back) / (1+ b_4 *
      ↪ con_loss + b_5 * unf_err);
  return float(result)
```

```
# Calculate
df['Momentum_1'] = 0;
df['Momentum_2'] = 0;
for index, row in df.iterrows():
  if index > 1:
    df.at[index, 'Momentum_1'] = cal_momentum(row['p1_consecutive_points_won'
        ↪ ], row['p1_winner'], row['p1_backhand_won'], row['
        ↪ p1_consecutive_points_loss'], row['p1_unf_err'])
    df.at[index, 'Momentum_2'] = cal_momentum(row['p2_consecutive_points_won'
        ↪ ], row['p2_winner'], row['p2_backhand_won'], row['
        ↪ p2_consecutive_points_loss'], row['p2_unf_err'])
df['Momentum_delta'] = df['Momentum_1'] - df['Momentum_2']

# Plot
line_label = 'Momentum Differential Positive: Carlos Negative: Nicolas'
plt.figure(figsize=(40, 5))
plt.plot(df['elapsed_time'].head(300), df['Momentum_delta'].head(300), marker=
    ↪ 'o', linestyle='-', color='b', label = line_label)
plt.title('Game Momentum and Game Point Won (Carlos Alcaraz vs. Nicolas Jarry)
    ↪ ')
plt.xlabel('Elpased Time(Seconds)')
plt.ylabel('Momentum Differential')
plt.grid(True)

for index, row in df.head(300).iterrows():
  if row['game_victor'] == 1:
    plt.plot(row['elapsed_time'],1, 'o', color='green', markersize=8, zorder
        ↪ =3, label = 'Game won by Carlos')
    if g == 0:
      plt.legend();
      g = 1
  elif row['game_victor'] == 2:
    plt.plot(row['elapsed_time'],-1, 'o', color='red', markersize=8, zorder=3,
        ↪  label = 'Game won by Nicolas')
    if r == 0:
      plt.legend();
      r = 1

plt.axhline(y=0, color='grey', linestyle='--', zorder=1)
plt.ylim(-1.5, 1.5)
plt.show()
```

Code for correlation analysis:

```
import pandas as pd
from datetime import datetime
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
df = pd.read_csv('momentum_file.csv')

# runs of success
df['runs_of_success'] = 0
for index, row in df.iterrows():
  next_three = df['consecutive_points_won'][index:index+4]
```

```
  df.at[index, 'runs_of_success'] = next_three.max()

# Calculate Pearson correlation coefficient and p-value
col_A = df['Momentum_delta']
col_B = df['runs_of_success']
correlation_coefficient, p_value = pearsonr(col_A, col_B)

# Print the results
print(f"Pearson Correlation Coefficient: {correlation_coefficient}")
print(f"P-Value: {p_value}")

# Check if the correlation is statistically significant
alpha = 0.05
if p_value < alpha:
    print('There is a statistically significant relationship between the
       ↪ variables.')
else:
    print('There is no statistically significant relationship between the
       ↪ variables.')

# swings in play
df['net_p1_runs_of_success'] = df['p1_consecutive_points_won'] - df['
   ↪ p2_consecutive_points_won']
df['net_p1_runs_of_success_t-1'] = df['net_p1_runs_of_success'].shift(1)
df['swings_respected_to_p1'] = df['net_p1_runs_of_success'] - df['
   ↪ net_p1_runs_of_success_t-1']
df['Momentum_delta_t-1'] = df['Momentum_delta'].shift(1)
df['swings_in_Momentum_delta'] = df['Momentum_delta'] - df['Momentum_delta_t-1
   ↪ ']

# Calculate Pearson correlation coefficient and p-value
col_A = df['swings_respected_to_p1'].fillna(0)
col_B = df['Momentum_delta'].fillna(0)

correlation_coefficient, p_value = pearsonr(col_A, col_B)

# Print the results
print(f"Pearson Correlation Coefficient: {correlation_coefficient}")
print(f"P-Value: {p_value}")

# Check if the correlation is statistically significant
alpha = 0.05
if p_value < alpha:
    print('There is a statistically significant relationship between the
       ↪ variables.')
else:
    print('There is no statistically significant relationship between the
       ↪ variables.')
```

Code for Random Forest Prediction Model analysis:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
from datetime import timedelta
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report,
    ↪ mean_squared_error, confusion_matrix
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
import lightgbm as lgb
import xgboost as xgb
import seaborn as sns
from statsmodels.tsa.arima.model import ARIMA
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.utils import to_categorical
from sklearn.preprocessing import LabelEncoder

match_data_path = 'momentum_file_final.csv'
df1 = pd.read_csv(match_data_path)
df = df1[df1['match_id'] == '2023-wimbledon-1311']
df['Momentum_delta'] = df['Momentum_delta'].shift(-1)
df = df[:-1]
df['Momentum_delta_categorical'] = df['Momentum_delta'].apply(lambda x: 1 if x
    ↪  > 0 else (-1 if x < 0 else 0))

conditions = [
    df['p1_winner'] == 1,
    df['p2_winner'] == 1
]

choices = [1,2]

default = 0

df['winner'] = np.select(conditions, choices, default=default)

X = df[['ace', 'winner', 'double_fault', 'consecutive_points_won', '
    ↪ backhand_won', 'unf_err', 'consecutive_points_loss']]
y = df['Momentum_delta_categorical']
y = y.astype(str)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
    ↪ random_state=42)

# Training the RandomForest model
model = RandomForestClassifier(n_estimators=100, random_state=42)  # Or
    ↪ RandomForestClassifier for classification
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.4f}")
```

```
feature_importances = model.feature_importances_
features = X.columns
importances = pd.Series(feature_importances, index=features)
sorted_importances = importances.sort_values(ascending=False)

# Display the sorted importances
print(sorted_importances)

# Generate the confusion matrix
cm = confusion_matrix(y_test, y_pred, labels=['-1', '0', '1'])

# Convert the confusion matrix to a DataFrame for better labeling
cm_df = pd.DataFrame(cm,
                     index=['-1', '0', '1'],  # Actual labels
                     columns=['-1', '0', '1'])  # Predicted labels

# Plotting the confusion matrix
............
plt.show()
```

Report on Use of AI

1. ChatGPT (GPT-4 April 2023)
   Query1: What is "the flow of the play"

   Output: "The flow of the play" refers to the dynamic progression of events and the rhythm of a game or a dramatic piece. It can pertain to different contexts:

   1. In Sports: It describes how the game progresses from moment to moment. In team sports like soccer or basketball, it refers to the movement of players and the ball, the transitions between offense and defense, and the strategies that unfold during the game. In tennis or similar individual sports, it might refer to the exchange of shots, the players' positioning, and the changing advantage between players.
   2. In Theater: It refers to the sequence and pacing of scenes and actions in a play. Good flow maintains the audience's interest and ensures that the narrative progresses smoothly without confusing or jarring transitions.

   In both uses, the flow is about the seamless, logical, and engaging progression of actions or events.

2. ChatGPT (GPT-4 April 2023)
   Query1: what is logistic regression analysis

   Output: Logistic Regression Analysis is a statistical method used for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). It is used extensively in various fields, including the medical and social sciences, for predicting the likelihood of occurrence of an event by fitting data to a logistic curve.

3. ChatGPT (GPT-4 April 2023)
   Query1: What is double faults in tennis

   Output: In tennis, a double fault occurs during a service play when the player serving fails to make a correct serve twice in a row. Each serve attempt must land in the correct service box (the diagonally opposite box from the server, without touching the net or going outside the specified area). If the first serve attempt fails (called a "fault"), the server gets a second chance. However, if the server also fails on the second attempt, it results in a double fault, and the server loses the point.

   Double faults are considered significant errors because they directly award a point to the opponent without them having to play the ball. Reducing double faults is crucial for players as they can significantly impact the outcome of a match, especially in crucial moments.

4. ChatGPT (GPT-4 April 2023)
   Query1: Can you tell me how to improve the visualization so make it more suitable for academic paper?

   Output: Certainly! To make the visualization more suitable for an academic paper, we can make several improvements:

1. Increase the Figure Size: Ensure the figure is large enough to be clearly legible when printed.
2. Improve Labeling: Use clear and descriptive axis labels, and if necessary, a legend.
3. Title and Caption: Add a descriptive title and a caption explaining the figure.
4. Font Size: Increase the font size for better readability.
5. Bar Colors: Use a more subtle color palette that is print-friendly.
6. Gridlines: Add horizontal gridlines to make it easier to read the importance values.
7. Save as a Vector Graphic: Save the figure in a vector format like SVG or PDF to prevent pixelation when zoomed or printed.

5. ChatGPT (GPT-4 April 2023)
   Query1: Explain in two sentences how each of these factors might negatively influence a tennis player's momentum: Consecutive Point Lost, Unforced Error

   Output: Losing consecutive points can negatively impact a tennis player's momentum by eroding their confidence and increasing mental pressure, making it harder to execute strategies effectively and maintain composure. Unforced errors can similarly disrupt momentum, as they not only gift points to the opponent but also lead to frustration and self-doubt, potentially affecting a player's focus and decision-making in subsequent points.