Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2409283

Unveiling the Momentum in Tennis

Summary

People have discovered that athletes exhibit momentum during matches, shaping their performance in a certain pattern throughout the game. To delve deeper into the patterns of momentum and the flow of the match, we utilized related data and established mathematical models to explore the following questions.

In Task 1: To capture the flow of play and dynamically quantify player performance during matches, we developed a Tennis Player Performance Evaluation Model. Using the **TOPSIS Method with a hybrid weight of entropy and CRITIC weight**, we scored player performance during matches. **Grey Relational Analysis** was introduced to evaluate the hybrid model. The results of both models were coordinated. Processing the match problem mentioned, we **visualized the dynamic changes** in their performances during this epic showdown. Through match result validation, our model effectively captured the changes in match flow, achieving an accuracy of 82.31%.

In Task 2: We considered momentum as a manifestation of **long-term memory** in match time series data. Using the **R/S Analysis Method**, we calculated player's Hurst exponent in multiple matches. We found that when using points as the time unit, the Hurst exponent was always greater than 0.5, indicating that momentum extends and amplifies player performance trends. Conversely, in matches consisting of multiple game sets, the Hurst exponent was always less than 0.5, indicating weakened momentum effects and mean reversion characteristics.

In Task 3, we used the **XGBoost model** to predict points where swings occur during matches. Exploring the most relevant factors using **Shapley Value** from game theory, we identified "Continuous Success" as the key factor related to swing occurrence. Testing the model with data from the Wimbledon Men's Singles, we achieved a prediction accuracy of 92.64%, indicating excellent model performance.

In Task 4, we tested the robustness and generalization of our Swing Predictor Model using match datasets from the US Open and the French Open tournaments, yielding an accuracy of only around 71%. We attribute the decreased accuracy to variations in **match surfaces**. Therefore, in future model optimization, we will consider the impact of surfaces on match scoring indicators.

Finally, we summarized our conclusions and wrote a memo providing coaches with advice on the role of "momentum" and how to prepare players to respond to events affecting the tennis match process.

Keywords: CRICTIC; Entropy-weighted TOPSIS; Grey Relational Analysis; R/S Analysis Method; XGBoost; Shapley Value

Contents

1 Introduction

1.1 Problem Background

Wimbledon stands as one of the four tennis Grand Slam tournaments, revered for its antiquity, tradition, and prestige. Among its events, the Wimbledon Men's Singles competition commands particular attention, drawing elite male tennis players from around the globe.

Remarkable swings in momentum occur during matches, favoring players who gain point advantages. This phenomena is common in tennis. Such patterns fuels athletes and enhances their performance.

1.2 Restatement of the Problem

In the 2023 Wimbledon Men's Singles final, 20-year-old Spanish prodigy Carlos Alcaraz triumphed over 36-year-old Novak Djokovic. MCM compiled the data from this match into a dataset, tasking us with addressing several key inquiries:

- Develop a model that quantifies players' performances during specific intervals of the match, capturing the flow of play as points unfold.
- Establish a model to investigate the role of momentum in matches, assessing its correlation with players' success or failure, thus scrutinizing a tennis coach's perspective that swings in play and runs of success by one player are random.
- Based on the above analysis, the following needs should be further addressed:
  - Utilize provided data to predict when the momentum will shift from favoring one to another, identifying factors most closely associated with these shifts.
  - Offer recommendations for players facing different opponents based on the variance in momentum shifts observed in past matches.
- Test the model on other matches, evaluating its accuracy. Should the model exhibit occasional inadequacies, analyze potential factors for future model enhancements. Validate the model's generalizability to other matches and sports.
- Summarize the preceding conclusions and compose a one to two-page memo, offering advice to coaches on the role of "momentum" and strategies for preparing players to respond effectively to events that influence the flow of mathces.

1.3 Our work

2 Assumptions and Justifications

1.Assuming that the player's scores in the match can well reflect their competitive performance, the data we use are reliable.

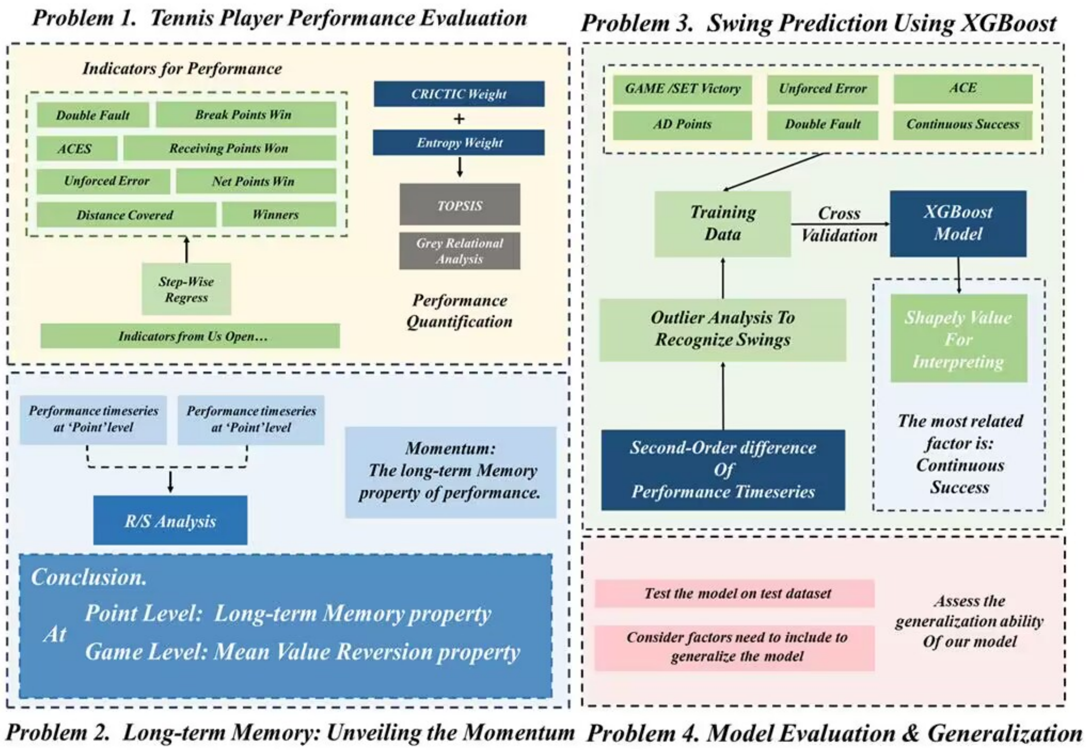


Figure 1: The Process of Our Work

2.**Deciphering Momentum**.

In statistics, a stationary process is a stochastic process whose unconditional joint probability distribution remains unchanged when shifted in time.[1] Conversely, a non-stationary process exhibits long memory in time series, indicating it is non-random.[2]

Drawing from the literal meaning of "momentum" as "strength or force gained by motion or by a series of events"[3] , **we conjecture that the momentum exhibited by players in their match performances is essentially a form of long memory: the current (or past) performance values in a match's time series significantly influence future performance values beyond the extent of random fluctuations**.

3 Tennis Player Performance Evaluation Model

The timely observable data in the game is the player's score in each game, which captures the performance of the player in each stage. When points occur, we can consider the player's performance during the time they attempt to earn that point, regardless of success. A remarkable score indicates a commendable performance during that period, while a lackluster score suggests otherwise. Therefore, we segment time into points as the smallest interval unit, quantify the player's performance using their score in each game.

3.1 Selection of Evaluation Indicators

Step 1: Initial Indicators from Relevant Literature

Table 1: Initial Metrics for Athlete Performance

Indicator	Description
ACES	Direct service points
DOUBLE FAULTS	Double faults
FIRST SERVE % IN	First serve percentage in
WIN % ON 1ST SERVE	Win percentage on first serve
WIN % ON 2ND SERVE	Win percentage on second serve
NET POINTS WON	Net points won
BREAK POINTS WON	Break points won
RECEIVING POINTS WON	Receiving points won
WINNERS	Winning shots
UNFORCED ERRORS	Unforced errors
TOTAL POINTS WON	Total points won
DISTANCE COVERED	Distance covered
DISTANCE COVERED/PT	Distance covered per point

We first consult relevant literature to select indicators for evaluating players' performance, with reference to the data on the official website of the US Open(see Table (1)).

Step 2: Filtering Under the Span of 'Game'

Utilizing the dataset provided, we excluded some not well-defined indicators. In elite-level athletic competitions, such specific scoring situations lack the general applicability required for evaluating competitive performance.

Step 3: Metric Screening Using Step-wise Regression

Statistical screening was introduced. We employed a regression analysis method, with total points won as the dependent variable Y and other indicators as independent variables X. Assuming there are n samples, p independent variables, and one dependent variable, the multiple linear regression model can be represented as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i.$$

where

- $y_i$is the dependent variable for the i^{th}observation;
- $x_{ij}$is the value of the j^{th} independent variable for the i^{th} observation;
- $\beta_0, \beta_1, \beta_2 \ldots \ldots$ are the parameters (regression coefficients) of the model;
- ϵ_i is the error term of the model, representing the unexplained random error.

The significance level for variable elimination was set at 0.8. The multiple correlation coefficient R ranges from 0 to 1, measuring the degree of linear relationship among

Table 2: Model Fit Assessment

Model	R	R^2	AdjustedR^2	residuals
1	0.761	0.58	0.563	12.17

variables. We get $R = 0.761$, indicates a close relationship between the dependent and independent variables, which is suitable for regression analysis(See table (2)).

The variance analysis test on the equation shows a p-value of 0.008, indicating high significance ($p < 0.01$). This means that regression analysis can be confidently conducted, as the relationship between the variables is statistically significant.

Below are the final evaluation indicators, all of which demonstrating significant influence on the athlete's overall score, i.e., performance in the match.

Table 3: Final Metrics for Athlete Performance

Indicator	Description	Symbol
ACES %	Direct service points as a percentage of total serves	ACE
DOUBLE FAULTS	Double faults divided by total points in a game	DF
WIN % ON 1ST SERVE	Win percentage on first serve	ACE1
NET POINTS WON %	Net point winning percentage	NPW
BREAK POINTS WON %	Break points won divided by break points opportunities	BPW
RECEIVING POINTS WON %	The proportion of points won by the serving opponent in the sum of both players' points	RPW
WINNERS %	Percentage of winners in total points	W
UNFORCED ERRORS	Unforced errors divided by total points in a game	UE
DISTANCE COVERED/GM	Distance covered per game	DIS

3.2 Combined Weight TOPSIS Evaluation Model

3.2.1 CRICTIC - Entropy Weighting Model for Composite Weights

The CRITIC method, proposed by Diakoulakietal, in 1995, is an objective weighting approach that evaluates the objective weights of indicators based on their comparative strength and conflict. Comparative strength is represented by the standard deviation, while conflict is indicated by the correlation coefficient.

Drawing on research [4], improvements to this method are proposed as follows:

1. Replacing the standard deviation with the coefficient of variation to eliminate dimensional influences.
2. Taking the absolute value of the correlation coefficient to eliminate the impact of positive and negative signs.

The CRITIC method assesses the comparative strength and conflict between indicators but skip the dispersion between indicators. In contrast, the entropy weighting method determines indicator weights based on the dispersion between indicators.

Therefore, the combined use of the CRITIC method and entropy weighting method provides a more objective reflection of indicator weights. The process for calculating the composite weights of indicators is outlined as follows:

Given m evaluation objects and n evaluation indicators, with original data represented as X_{ij}, where $i = 1, ..., m$ and $j = 1, ..., n$, the first step involves standardizing the data

$$x_{ij} = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}}(\text{ Positive Indicators}), x_{ij} = \frac{X_{max} - X_{ij}}{X_{max} - X_{min}}(\text{ Negative Indicators}).$$

Where X_{max} and X_{min} represent the maximum and minimum values, respectively, for the j^{th} indicator. x_{ij} denotes the processed data. Utilizing the CRITIC method to compute weights, we calculate the information content for the j^{th} indicator.

$$c_j = \frac{\sigma_j}{\bar{x}_j} \sum_{i=1}^{m} (1 - |r_{ij}|)$$

Where σ and x represent the standard deviation and mean value, respectively, for the j^{th} indicator, and r_{ij} denotes the correlation coefficient between the i^{th} and j^{th} indicators.

Calculate the weight of the j^{th} indicator:

$$w_1 = \frac{c_j}{\sum\limits_{j=1}^{n} c_j}$$

According to the entropy weighting method, calculate the weight by computing the probability of occurrence for the j^{th} indicator of the i^{th} evaluation object:

$$P_{ij} = \frac{x_{ij}}{\sum\limits_{i=1}^{m} x_{ij}}$$

Calculate the information entropy of the j^{th} indicator:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} p_{ij} \ln p_{ij}$$

Calculate the weight of the j^{th} indicator:

$$w_2 = \frac{1 - e_j}{\sum\limits_{j=1}^{n} (1 - e_j)}$$

The combined weight of the j^{th} indicator is mean value of the two weights.

Table 4: Indicator Weight Calculation Results

Model	ACE	ACE1	BPW	NPW	RPW	W	DF	UE	DIS
Entropy Weight	0.266	0.247	0.074	0.064	0.062	0.221	0.009	0.019	0.037
CRICTIC	0.091	0.059	0.112	0.145	0.072	0.122	0.074	0.203	0.122
composite Weight	0.179	0.153	0.093	0.105	0.067	0.171	0.041	0.111	0.080

The weight calculation results for each indicator are presented in Table (4).

The serving side holds the advantage of initiating the game, controlling the rhythm, and taking the initiative, which may lead to higher performance scores. Our evaluation criteria take into account factors where the serving side is more likely to score points or win, and therefore, serving-related indicators have been incorporated into the evaluation system with significant weighting.

3.2.2 The TOPSIS quantification scoring model.

In 1981, Hwang and Yoon first proposed the TOPSIS method, which ranks evaluation objects based on their proximity to idealized targets. This method measures the distance between evaluation objects and the best and worst solutions.

By combining the entropy weighting method and the CRITIC method to calculate weights, it effectively overcomes the limitations of the traditional TOPSIS method, which fails to reflect the correlation and importance of variables.

1. Calculate the weighted matrix.

$$V = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix}$$

   Where $v_{ij} = x_{ij}w_j$,w_j represents the weight of the j^{th} indicator.

2. Calculate the positive ideal solution and the negative ideal solution :

$$V^+ = \left(v_1^+, v_2^+, \cdots, v_n^+\right) = \{\max v_{ij} \,|j \in J_1, \min v_{ij}|\, j \in J_2\}$$
$$V^- = \left(v_1^-, v_2^-, \cdots, v_n^-\right) = \{\min v_{ij} \,|j \in J_1, \max v_{ij}|\, j \in J_2\}$$

   Where J_1represents the set of positive indicators, and J_2 represents the set of negative indicators.

3. Calculate the distance from the evaluation object to the positive and negative ideal solutions:

$$S_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2}, S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}$$

4. Calculate the relative closeness of the i-th evaluation object to the ideal solution:

$$\delta_i = \frac{S_i^-}{S_i^+ + S_i^-}$$

Where $0 \leq \delta \leq 1$, the values of δ_i are sorted based on their magnitude, with larger values indicating closer proximity to the optimal level.

3.3 Grey Relational Analysis Evaluation Model

Given the complex and interdependent relationship between each indicator and the overall assessment, we also established a Grey Relational Evaluation Model to comprehensively analyze and evaluate athletes' performance in competitions. Grey Relational Evaluation, a branch of Grey Theory, is commonly employed to assess interacting factors.

Step 1:Calculate the correlation coefficient.

Using the total points scored by players in each match as the reference sequence and the numerical values of various indicators as the comparative sequence, after normalizing the data, grey relational coefficients are calculated:

$$\zeta_i(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|}$$

In which,

- $|x_0(k) - x_i(k)|$ represents the absolute difference between the comparative sequence of each evaluation object and the reference data sequence.
- Resolution ratio ρ (0,1), the difference of correlation coefficient increase with ρ decreasing. Herein, we choose $\rho = 0.5$.

Step 2: Calculate the correlation of indicators.

$$r_j = \frac{1}{n}\sum_{j=1}^{n} \xi_{ij}$$

where,r_j represents the grey correlation degree for the j^{th} indicator.

Step 3: Calculate the weight of the indicator.

$$w_j = \frac{r_j}{\sum_{j=1}^{m} r_k} \quad (j = 1, 2, 3, ..., m)$$

Step 4: Calculate the scores and normalize them.

$$S_i = \sum_{j=1}^{m} Zij \cdot r_j$$

$$\tilde{S}_i = \frac{S_i}{\sum_{k=1}^{m} S_k}(i = 1, 2, \cdots, n)$$

Based on above steps, we can get results by SPSS. The results are shown in Table (5).

Table 5: Results of Grey Relational Analysis

Indicator	ACE1	ACE	BPW	NPW	RPW	W	DF	UE	DIS
Correlation Coefficient	0.480	0.455	0.587	0.798	0.536	0.525	0.616	0.848	0.683

The quantitative scores obtained from the Grey Relational Analysis Evaluation Model are close to those given by the Combined Weight TOPSIS Evaluation Model. After comprehensive consideration, we can obtain the final quantitative evaluation scores for the athletes' performance.

3.4 Performance Evaluation and Visualization

In evaluating the match-id=2023-wimbledon-1701, the final match between Carlos Alcaraz and Novak Djokovic, we quantified their real-time performance scores, capturing the flow of play as points occurred.

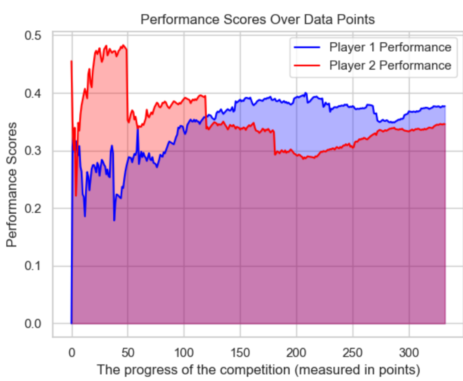


Figure 2: The Dynamic Performance of Players

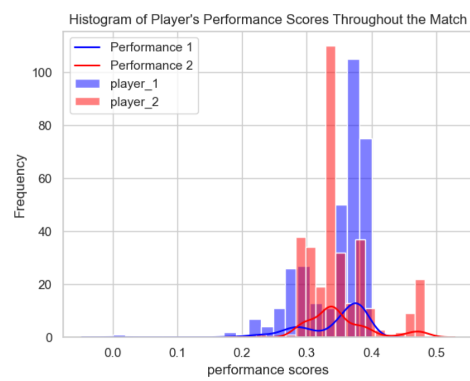


Figure 3: The Overall Performance of Players

We use "Player 1" to refer to Novak Djokovic and "Player 2" to refer to Carlos Alcaraz.Figure 2 illustrates the dynamic changes in the performance of both players:

Initially, Novak Djokovic's performance was significantly better than Carlos Alcaraz, with ratings above 0.4; but after 50 points, his performance dropped sharply, aligning with Carlos Alcaraz's. Between 60 and 130 points, Novak Djokovic regained the lead.

Carlos Alcaraz, despite initially trailing, steadily improved from 50 to 200 points, peaking at 0.4. After 135 points, he surpassed Novak Djokovic and maintained the lead thereafter.Despite Novak Djokovic's gradual improvement post-200 points, narrowing the gap with Carlos Alcaraz, his performance still slightly trailed until the end of the match.

Figure 3 displays their performance distributions: Carlos Alcaraz's concentrated around 0.4, Novak Djokovic's around 0.3, indicating Carlos Alcaraz's superiority.

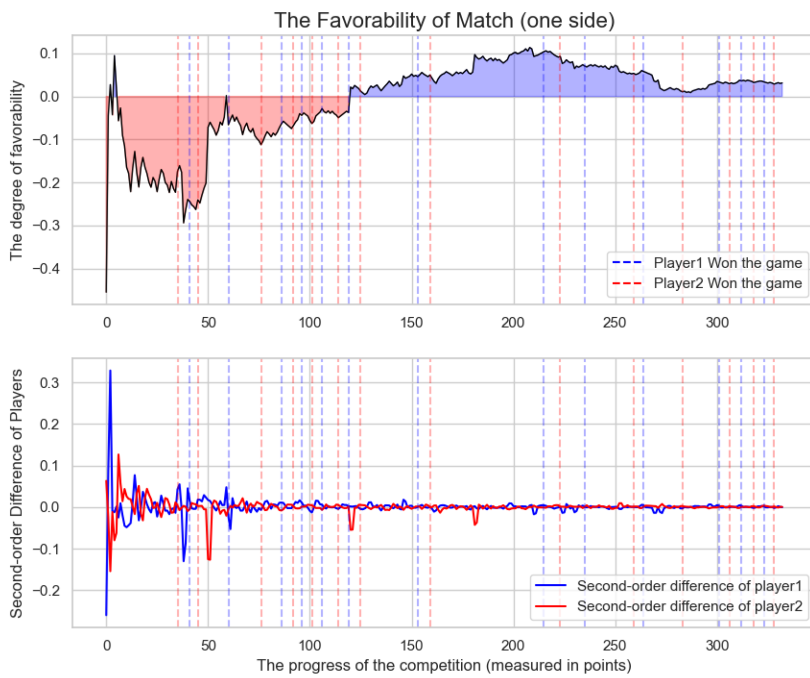


Figure 4: The Flow of the Game

Analyzing the dynamic performance of both players in conjunction with Figure 4 reveals the shifts in the match's momentum. Initially, Novak Djokovic had better performance, controlling the match's momentum, as evidenced by winning the first set 6:1. However, as Carlos Alcaraz's performance improved, the match's momentum gradually shifted away from Novak Djokovic, completely favoring Carlos Alcaraz after 135 points and maintaining this trend until the end of the match.

Validating the match's results and the outcome of each game against the final match result, we achieved a model accuracy of **94.31**%.

3.5 Sensitivity and Robustness Analysis

The model consists of two independently evaluated models: the TOPSIS model and the Grey Relational Analysis model. Sensitivity analyses were conducted separately for each model. If the model is highly sensitive to changes in weight parameters, it may lack robustness and require further adjustments or improvements. Conversely, if the model is less sensitive to changes in weight parameters, its results may be more reliable and possess better predictive capabilities.

(1) By fluctuating the resolution coefficient ρ, which involves varying the weight parameters within the model, we assessed the impact of weight parameter changes on the

stability and reliability of the Grey Relational Analysis results.

The results are shown in Table 6.Here, $Value_i$ and $Rank_i(i = 1, 2, 3)$ respectively represent the correlation values and rankings obtained from Grey Relational Analysis when ρ **=0.4,0.5, and 0.6**.

Table 6: The Correlation Coefficient Results

[5pt]

Indicator	0.4	Rank	0.5	0.6	Rank	
ACE1	0.533	8		8	0.577	8
ACE	0.51	9	0.554	9		
BPW	0.636	5	0.675	5		
NPW	0.826	2	0.846	2		
RPW	0.589	6	0.632	6		
W	0.577	7	0.618	7		
DF	0.664	4	0.702	4		
UE	0.866	1	0.88	1		
DIS	0.723	3	0.754	3		

As shown in the table 7, it is evident that the fluctuation in ρ does not affect the weight parameters, indicating the stability of the model's results.

(2) When exploring the variation of weights for each indicator within the range of $-50\%, -40\%, ..., 40\%, 50\%$, we investigate whether the evaluation scores from the TOPSIS model change. It is observed that the alternative evaluation results remain largely unchanged, indicating that the impact of indicator weights does not significantly alter the final evaluation scores of TOPSIS. Thus, the results of the TOPSIS evaluation model are deemed reliable.

4 Long-term Memory Property : Unveiling the Momentum

we conduct a long memory test on the time series of a match to assess whether momentum plays a role in the match.

4.1 R/S Analysis Method

R/S Analysis, first introduced by Hurst in 1951, assesses the long-memory characteristics of a time series by calculating the Hurst exponent (H):[2]

- If $0 < H < 0.5$, the time series exhibits mean-reverting properties.
- If $H = 0.5$, the time series is considered a random process.
- If $0.5 < H < 1$, the time series demonstrates long memory.

The specific calculation process is as follows[5]:

Step 1 : Begin with a time series of length M. Convert this into a time series of length N= M- 1 of logarithmic ratios:

$$\mathbf{N_i} = \log\left(\mathbf{M}_{(i+1)}/\mathbf{M_i}\right), \mathrm{i} = 1, 2, 3, \ldots, (\mathbf{M} - 1)$$

Step 2 : Divide this time period into A contiguous subperiods of length n, such that $A \ldots n = N$. Label each subperiod I_a, with $a = 1, 2, 3, ..., A$. Each element in I_a is labeled $N_{k,a}$ such that $k = 1, 2, 3, ..., n$. For each I_a of length n, the average value is defined as:

$$e_a = \frac{1}{n}\sum_{k=1}^{n} N_{k,a}$$

where e_a = average value of the N_i contained in subperiod I_a of length n.

Step 3 : The time series of accumulated departures ($X_{k,a}$) from the mean value for each subperiod I_a is defined as:

$$X_{k,a} = \sum_{i=1}^{k}(N_{i,a} - e_a) \quad \mathbf{k = 1, 2, 3, \ldots, n}$$

Step 4 : The range is defined as the maximum minus the minimum value of $X_{k,a}$ within each subperiod I_a:

$$R_{l_a} = max(X_{k,a}) - min(X_{k,a})$$

where $1 \le k \le n$.

Step 5 : The sample standard deviation calculated for each subperiod I_a:

$$S_{l_a} = (\frac{1}{n}\sum_{k=1}^{n}(N_{k,a} - e_a^2))^{0.50}$$

Step 6 : Each range,R_{l_a}, is now normalized by dividing by the S_{l_a} corresponding to it. Therefore,the rescaled range for each I_a subperiod is equal to $\frac{R_{l_a}}{S_{l_a}}$. From Step 2 above, we had A contiguous subperiods of length n. Therefore, the average R/S value for length n is defined as:

$$(R/S)_n = \frac{1}{A}\sum_{a=1}^{A}(R_{1_a}/S_{l_a})$$

Step 7 : The length n is increased to the next higher value, and $\frac{M-1}{n}$ is an integer value. We use values of n that include the beginning and ending points of the time series,and Steps 1 through 6 are repeated until $n = \frac{M-1}{2}$.

$$\mathrm{(R/S)_n = cn^H} \tag{1}$$

$$\log(\mathbf{R}/\mathbf{S}_n) = \log(\mathbf{c}) + \mathbf{H}\log(\mathbf{n}) \tag{2}$$

We can now apply equations(1) and (2) by performing an ordinary least squares regression on log(n) as the independentvariable and log(R/S)n as the dependent variable.The intercept is the estimate for log(c), the constant. The slope of the equation is the estimate of the Hurst exponent,H.

4.2 Results and Analysis of Hurst Exponent

We initially employed the match score dataset provided by MCM, focusing on the match data for match-id=2023-wimbledon-1301. Within the entire match time series, we investigated:

- Whether the performance of players exhibits long memory at the level of points, with **points serving** as the smallest time interval within a match.
- Whether the performance of players exhibits long memory at the level of games, with **games** being the smallest time interval within a match.

Based on above steps, we can get results by MATLAB programming. The results are shown in Table (7).

Table 7: The calculation Results of the Hurst Exponent

Name	Player ID	H-Points	H-Games
Carlos Alcaraz	player1	0.6221	0.4491
Novak Djokovic	player2	0.5459	0.4116

We utilized a dataset provided by MCM, consisting of randomly selected matches, to compute the Hurst exponent reflecting players' performances during the matches. The illustration of these findings is presented in the following figure.

4.2.1 The Impact of Momentum on Performance at the Point Level

The results indicate that, **at the point level, both athletes exhibit long-term memory in a match, suggesting the presence of momentum**. This momentum influences the players' performance in future points based on their current and previous point scores. If a player's performance scores rise due to points earned in the preceding period, there is a continued upward trend in the next period, indicating strong momentum. Conversely, the trend reverses accordingly.

In summary, **momentum in a match extends and amplifies the trend of player performance changes over consecutive points, affecting the player's scoring or losing streak in subsequent points**. The closer the Hurst exponent is to 1, the stronger the impact of this momentum.

In this match, Carlos Alcaraz's Hurst exponent is 0.6221, higher than Novak Djokovic's 0.5459. This suggests that Carlos Alcaraz's performance is more sensitive to previous

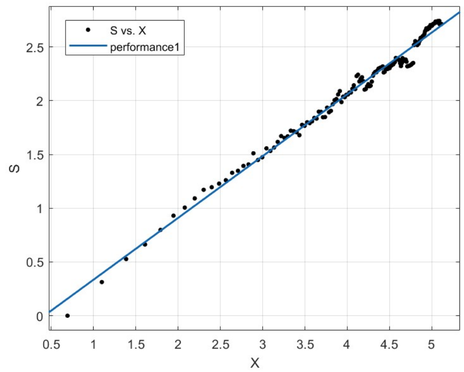


Figure 5: Carlos Alcara's H at Point Level

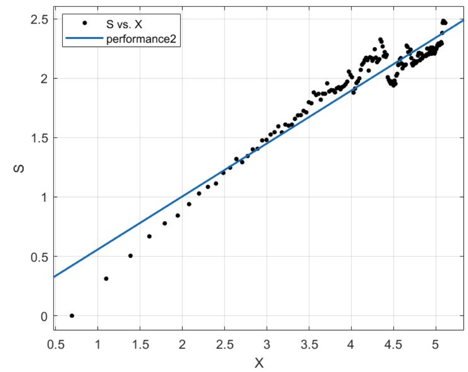


Figure 6: Nicolas Jarry's H at Point Level

scores. When influenced by current or past scores, his performance fluctuates more in subsequent points. This susceptibility and greater variability on court align with Carlos Alcaraz's youthful identity as a tennis prodigy.

As one of the greatest Grand Slam players, Novak Djokovic maintains a stable competitive mindset. His lower Hurst exponent indicates less sensitivity to match scores, resulting in more consistent performance. However, he lacks the youthful energy and momentum of Carlos Alcaraz, which may explain his loss in this match.

4.2.2 The Impact of Momentum on Performance at the Game Level

The results indicate that, **when using the game as the minimum time interval, the performances of both players in a match exhibit mean reversion characteristics, showing no long memory**. In other words, across different game sets, a player's performance shows discontinuity, with significant fluctuations within each game set, **yet overall converging towards the player's average competitive level, devoid of momentum on the game scale**.

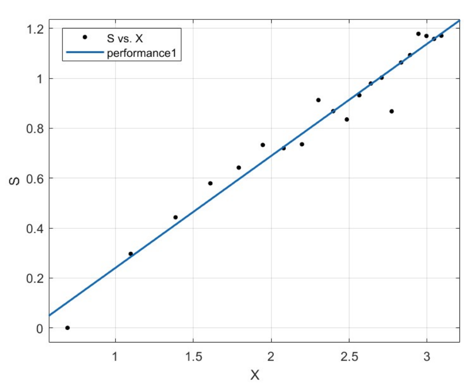


Figure 7: Carlos Alcaraz's H at Game Level

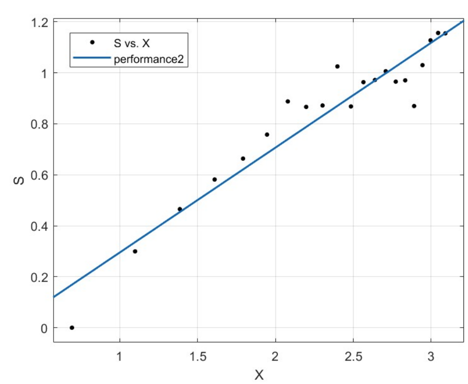


Figure 8: Novak Djokovic's H at Game Level

Although players may experience psychological burdens or encouragement due to known match scores, such psychological effects persist inconsistently across different game sets.

We analyzed the reasons behind these findings. Considering that each game set involves alternating serves, granting each player serve advantages alternatively, and within a match, there are intervals such as changing sides and breaks between games, which interrupt player momentum. This reduces the impact of previous game results on a player's current performance, allowing players to reset to a competitive level at the beginning of each game set. This contributes to the fairness of competition between players, ensuring that the match outcome better reflects the players' inherent abilities.

4.3 Conclusion Recap

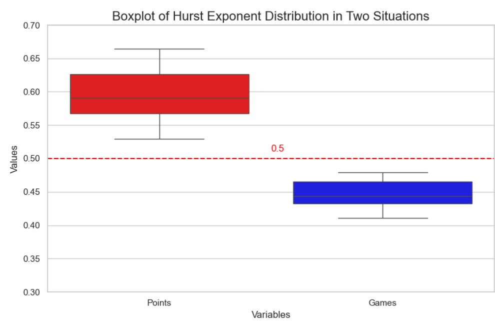


Figure 9: Boxplot of Hurst Exponent Distribution in Two Situations

A tennis coach questioned the role of momentum in matches, suggesting that swings in play and runs of success by one player are random.

However, **our model validation contradicts this assertion**.We utilized a dataset provided by MCM, consisting of multiple randomly selected matches, to compute the Hurst exponent reflecting players' performances during the matches. The results are depicted in the figure**??**, **indicating the universality of momentum**:

In each game, when using the point as the time unit, momentum extends and amplifies the player's performance trends, allowing them to surpass the range of random fluctuations around the mean level of competition. In a match consisting of multiple game sets, the effect of momentum is weakened, and players' performances tend to regress towards their average competitive level, demonstrating mean reversion characteristics.

5 Swing Predictor Model Based on XGBoost and SHAP

5.1 Features Used to Measure the Swings in the Match

Player performance is evaluated using a series of scoring-related indicators, thus the overall match tends to favor the side with the currently higher performance score. As revealed in the second question, players exhibit momentum in their performance over consecutive points during a match, thereby indicating a continuity in the match's dynamics over several points. When a swing in momentum occurs, it signifies a shift in a player's performance.

Hence, combining the evaluation model from the first question, which assigns significant weight to indicators affecting player scores, considering the influence of momentum, and referring to the rules of the game, we have selected features to measure whether the match dynamics are likely to undergo a swing:

Table 8: Features Used to Measure the Swings in the Match

Aspect	Feature
Decisive Points	AD Point Game Victory Set Victory
Serious Mistake	Unforced Error Double Fault
Outstanding Performance	ACE Cont-success

As we aim to forecast changes in match dynamics, we treat consecutive wins or losses of points as continuous scoring for one side, using only "Cont-success" as a feature. While breaks and changeovers offer chances for player adjustments, we excluded these rules as they do not directly cause swings in match dynamics, which hinge primarily on in-game performance.

5.2 XGBoost Model

5.2.1 Model Description

XGBoost stands out as the premier model for handling standard tabular data, typically stored in Pandas DataFrames. It implements the **Gradient Boosted Decision Trees algorithm**, offering distinct technical advantages over other implementations.

Gradient boosting is a method that iteratively augments models into an ensemble through cycles. Initially, the ensemble starts with a single model, which may produce somewhat naive predictions. However, subsequent iterations address any inaccuracies by incorporating new models into the ensemble.

Here's how the cycle unfolds:

- Initially, the current ensemble generates predictions for each observation in the dataset by aggregating predictions from all models.
- These predictions are used to compute a loss function, such as mean squared error.
- Next, a new model is fitted to the data based on the loss function, with parameters optimized to reduce the loss. The "gradient" in "gradient boosting" refers to the use of gradient descent on the loss function to determine these parameters.

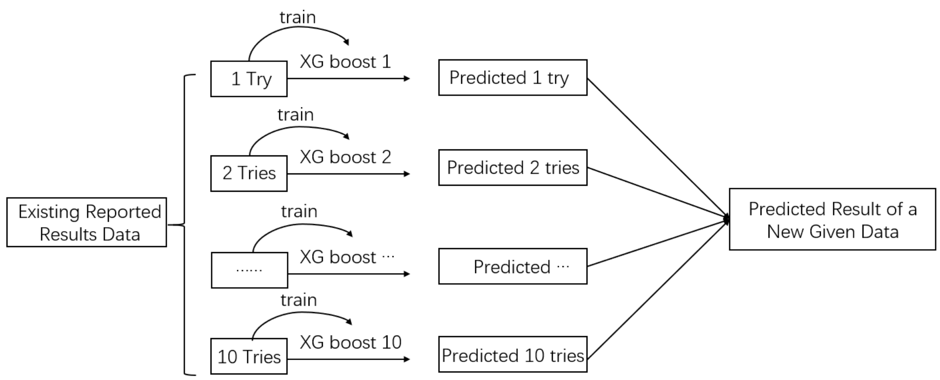


Figure 10: The Process of XGBoost

- Finally, the new model is added to the ensemble, and the process repeats itself.

5.2.2 Recognize Swings: Outlier Analysis of Performance Difference

In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis that seeks to build a hierarchy of clusters.[6]

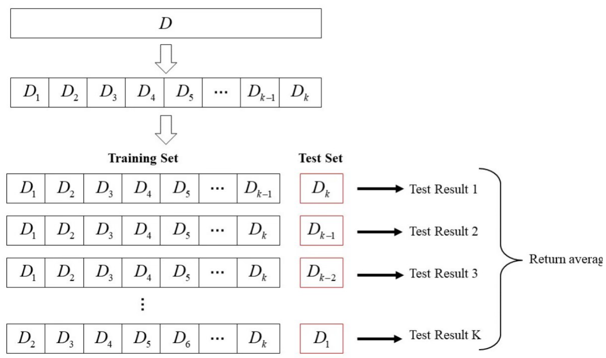
we conduct a differential analysis on the scores of both sides in a game, obtaining the changes in performance over the past five points. Calculate the difference between these two changes. When there is a significant disparity in the changes in performance of the two players, it can be inferred that a swing in the situation is likely to occur. Conversely, it can be inferred that their states remain similar to the previous moment, with no abrupt changes, and a swing in the game is unlikely to occur. This approach is rigorous yet elegant, and aligns with the style of mathematical modeling papers.

5.2.3 The Process of XGBoost Training

As shown in Figure10,We employ 10 models and utilize the aforementioned 7 features to predict whether the match's momentum will swing in the next point based on the performance data from the past 5 points.(See Table (8)).

During the training process, **cross-validation** was used to determine the corresponding hyperparameters of each XGBoost model. Using the determined hyperparameters, the model's coefficient of determination (R^2) is **0.71**, showing that the model is indeed capable of predicting.

We tested the model using data from the 2022 Wimbledon Gentlemen's tournament, achieving an accuracy rate of 82.31$\%$ in predictions. The result is shown in the figure12,suggesting that the model exhibits good predictive performance.



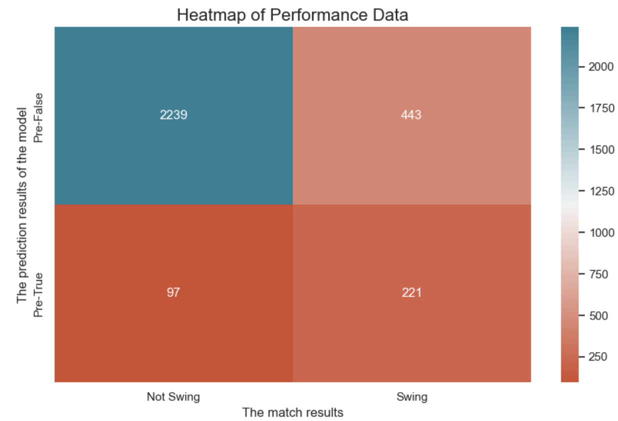


Figure 11: The Process of cross-validation Figure 12: The prediction results of the model

5.3 The Most Related Factor Recognition With Shapley Value

SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values.[7]

In the perspective of game theory, a prediction can be explained by assuming that each feature value of the instance is a “player” in a game where the prediction is the payout. Shapley values – a method from coalitional game theory – tells us how to fairly distribute the “payout” among the features, and thus make the “black box” of machine learning model explainable.

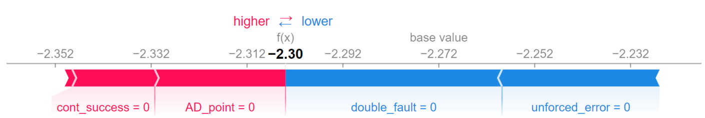


Figure 13: Explanation for one sample using SHAP

Taking the first sample as an example, We analyze its sensitivity to various features. The model output is -2.30, with a baseline defined as the model output minus the average value of the training data. The numbers below the arrows represent the feature values of this instance. For example, if Cont-success is 0 and AD Point is 0, the feature is represented as 0. Features that push the prediction higher are shown in red, while those that push it lower are in blue. The longer the arrow, the greater the impact of the feature on the output. Changes in the x-axis tick marks indicate the amount of increase or decrease in the influence.

From Figure 13, it is evident that Cont-success and AD-Point have a positive effect on the reversal of the situation, while Double-Fault and Uforced-Error have the opposite effect.

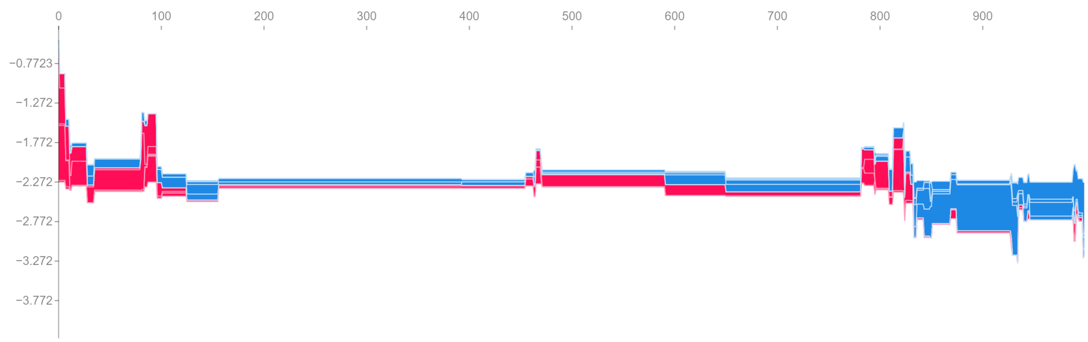


Figure 14: Explanation for 1000 samples using SHAP

For the first 1000 samples, the influence of each feature is analyzed similarly, and the results are shown in the figure14.

We **plot the Shapley value for each feature of every sample**, where the position on the y-axis is determined by the feature and the position on the x-axis is determined by each Shapley value. The color represents the feature value (red for high, blue for low), allowing us to match how changes in feature values affect changes in risk. Overlapping points are jittered along the y-axis so that we can understand the distribution of Shapley values for each feature, and **these features are sorted according to their importance**.

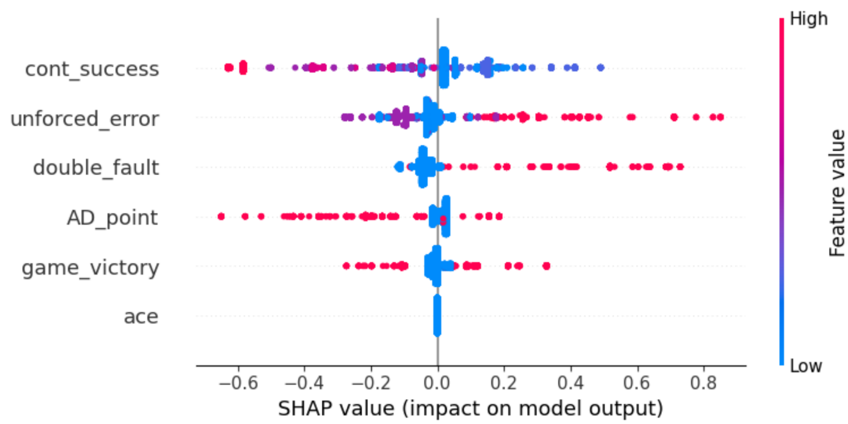


Figure 15: Explanation for 1000 samples using SHAP

Using Shapley Value, we could conclude the most related factor is **Continuous Success**.

5.4 The Advice For Tennis Players

Given the differential in past match "momentum" swings, we offer the following recommendations for a player entering a new match against a different opponent:

Opponent Analysis:

- Analyze recent matches of the opponent pre-match, using statistical methods to identify specific momentum fluctuations, particularly during critical points.

- Compare the player's momentum fluctuations with the opponent's, identifying similar and significantly different momentum trends.

Strategic Planning:

- Targeted training should address score ranges where the opponent's momentum fluctuates significantly, especially during points with notable increases. Training should also focus on countering the opponent's preferred scoring methods.
- During the match, promptly call timeouts when the opponent's momentum spikes significantly. Implement strategic responses to counter the opponent's primary scoring methods.

6 Swing Predictor: Evaluating Model Performance and Generalizability

6.1 The Results of the Model's Generalization Test

We have collected some scoring datasets from the US Open and the French Open tournaments to test the robustness and generalization of our Swing Predictor Model. The predicted results regarding swings are shown in the table below. Compared to the predictive performance on the Wimbledon Men's Singles dataset, there is a significant decrease in accuracy, indicating insufficient generalization of our model.

Table 9: The Model's Prediction Accuracy Varies Across Different atches

Name	2022 US Open	2022 French Open	2022 Wimbledon Men's Singles
Accuracy (%)	71.93	69.22	81.96

6.2 Considering the Court Surfaces Factors

We found significant differences in the playing surfaces among the three tournaments: The US Open features **hard courts**, The French Open is played on **clay courts**, and Wimbledon's men's singles matches are held on **grass courts**. Considering that our model selects predictive features based on their high contribution rates in scoring, we analyzed some of the effects of the court surfaces on players' performances:

1. **Ball Speed and Bounce**: Hard courts offer higher and faster ball bounce; grass courts provide lower and quicker bounce, favoring volleying; clay courts have slower and higher bounce, demanding endurance.
2. **Player Movement**: Hard and grass courts require similar movement, risking more ankle and knee injuries; clay courts allow sliding, demanding less strain but more skill.

3. **Ball Stability**: Clay and hard courts offer more consistent bounce, aiding trajectory anticipation; grass courts' bounce can vary due to wear and humidity.

Due to our Wimbledon-based data, our model's feature selection has limitations. Grass court instability leads to higher service and net points, significant in match scoring. Other surfaces may lack such data impact, limiting predictive momentum swings. On clay, monitor player movement for stamina and injury risks; on hard courts, prioritize overall player ability, focusing on serving points.

6.3 Extending to Other Sports

To extend the predictive model to forecast swings in other sports, it's essential to consider the rules and conditions specific to those sports, including the playing surface. Then, select more appropriate features to train the XGBoost model. Simply applying our model directly may lead to inaccuracies.

7 Strengths and weaknesses

7.1 Strengths

Outstanding Explanation for Momentum: Long-term Memory at Point/Game Level

Our theoretical explanation for momentum is the long-term memory property of the performance timeseries. Momentum means that fluctuation of performance is influenced by previous game situation, that's to say, the performance tends to maintain the original trend. This is depicted by long-term memory property of timeseris vividly.

We look into the performance in both point and game levels. Our result suggests that performance trends tend to be more extreme at point level, for instance the 6-1 game. But at game level, the performance tends to regress towards the mean value. This is consistent with our experience, in which a game of 6-1 is common but a match of 5-1 is rare.

Excellent Model Interpretability for XGBoost Using SHAP

Machine learning model explanation is such a difficult problem that machine learning is called 'Black Box'. We using the shapley value, from the perspective of game theory to interpret the result and get the most related feature.

7.2 Weaknesses

Absence of Quantified Momentum Definition

The R/S Analysis Method can only test if a given timeseries has the property of long-term memory, hence we could only conclude the existence and intensity of momentum. For the incredible swings in the situation, our model is powerless.

Poor generalization of the model

Some key features our model use is related to the tennis, which means our model is difficulty to apply to other sports. But the idea of using 'Decisive Points', 'Serious

Mistake' and 'Outstanding Performance' to identify the incredible swings is feasible.

8 The Memorandum to Tennis Coaches

To: Tennis Coaches
From: MCM Team #2409283
Subject: Leveraging Momentum and Preparing Players for Dynamic Match Scenarios
Date: February 5, 2024

As we explore tennis match dynamics, it's clear that grasping and utilizing momentum are pivotal in determining match results. Our analysis, rooted in player performance data and statistical methods, underscores the importance of momentum. This memo summarizes our findings and offers practical guidance to you on leveraging momentum and preparing players to adapt to match-changing events.

Momentum in tennis refers to the psychological and performance-driven advantage that a player gains from winning consecutive points or dominating specific phases of a match. From the perspective of statics, the **long-term property** of a timeseries is the manifestation of momentum.

Key Findings and Recommendations:

Pre-match Analysis of Opponents.

Analysis of opponent's recent matches will help us identify patterns of momentum swings. Utilize statistical methods to capture critical moments, such as breakpoints and pivotal game stages, where significant momentum shifts occurred. Compare the momentum curves of our players with those of the opponents to identify areas of convergence and divergence.

Strategic Formulation.

Customize training sessions to address specific match points where opponents experience momentum swings. Prepare players to conquer opponents' preferred scoring methods and disrupt their momentum during crucial match moments. Encourage players to maintain a stable competitive mindset and capitalize on opportunities to seize momentum shifts in their favor.

Utilizing Statistical Methodologies.

Test the long-term memory aspect of player performance using R/S Analysis Method, identifying trends and fluctuations influenced by previous match data. That's to say, quantify how much a player is affected by the momentum during a match.

Swing Prediction and Interpretation.

Define swing as performance turning points and employ our model including Hierarchy Clustering and XGBoost for accurate swing prediction. Using such model to help players to prepare for the swing of momentum. Utilize the Shapley Value to interpret model predictions and identify key factors influencing momentum shifts, thus assisting players to make futher improvement.

Conclusion.

In conclusion, momentum serves as a pivotal factor in matches, influencing performance and match outcomes. By understanding the nuances of momentum and employing statistical methods, coaches can empower players to respond effectively to dynamic match scenarios, capitalize on momentum shifts, and emerge victorious on the court.

This memorandum aims to equip coaches with actionable insights and strategies to optimize player performance and navigate the complex landscape of tennis match dynamics. Thank you for your attention and commitment to player development.

Sincerely,

MCM Team #2409283

References

[1] P. A. Gagniuc, *Markov Chains: From Theory to Implementation and Experimentation*. USA, NJ: John Wiley & Sons, 2017, pp. 1–256, ISBN: 978-1-119-38755-8.

[2] H. E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, no. 1, pp. 770–799, 1951.

[3] *The official website of the us open*, https://www.merriam-webster.com/dictionary/momentum, 2024.

[4] F. Weizhong and C. Liuping, "Research on the evaluation of high quality development of manufacturing industry from the perspective of integration of the yangtze river delta———topsis evaluation model based on improved critical-entropy weight method," *Journal of Industrial Technological Economics*, vol. 9, 2020.

[5] E. E. Peters, *FRACTAL MARKET ANALYSIS Applying Chaos Theory to lnvestment and Economics*. John Wiley & Sons, 1994.

[6] *Hierarchical clustering*, https://en.wikipedia.org/wiki/Hierarchical_clustering, 2024.

[7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, Eds., Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.