Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2420216

The Secret of “Momentum”

Summary

People often attribute remarkable swings in tennis matches to “momentum”. In order to explore the secrets behind “momentum”, we analyse the factors influencing “momentum”and match swings. Subsequently, we establish models for evaluation, prediction, and interpretation.

For task 1: Firstly, we clarify the concept of “momentum”in tennis matches. Next, through literature review and experiential analysis, applying the **sliding window** approach, we process the data and obtain 11 factors, and use PCA to distill them into 5 principal components. Lastly, we employ AHP and EWM to derive their weights, and multiply the serving coefficient to establish the **Multi-angle Momentum Evaluation Model (Model I)**. Utilizing the model, we calculate RM(Relative Momentum) to quantitatively assess a player's performance during a specific time period, and describe the flow of play through visual methods.

For task 2: To demonstrate the role of “momentum”in matches, We only need to demonstrate that RM could indicate swings in the play. We use a combination of qualitative and quantitative analyses. In qualitative analysis, we visualize match trends, and use “momentum”to explain 2023-wimbledon1701 match which has incredible swings. In quantitative analysis, we treat RM as the score expectation, and use actual scores as true labels, resulting in $RMSE = 0.303$, $R^2 = 0.87$. Additionally, we construct a binary classifier using RM as the indicator, plot the ROC curve, and obtain an AUC of 0.79, indicating excellent classifier performance and role of “momentum.”

For task 3: We innovatively create the RAS(Relative Advantage Swing) as an indicator to describe changes in the flow of play. Upon this, we develop the **XGBoost-SHAP Interpretable Swings Prediction Model (Model II)**: firstly, we balance the dataset using undersampling, and then incorporate the factors from task 1 along with point_victory as input features. Using XGBoost, we predict swings with an accuracy of 84.2%. Employing SHAP analysis for feature importance, we identify the top four relevant factors as overall point difference, recent scores, server, and successful serves, with SHAP values of 0.32, 0.30, 0.17, and 0.16. Based on these results and considering the historical differences in momentum swings, we provide recommendations for players focusing on training emphasis, tactical arrangements, and mental adjustments.

For task 4: We collect additional match data from relevant websites, as detailed in the table 7. Applying the model to five matches in the first two rounds of the 2023 Wimbledon Men's Singles, the results are presented in the table 8. We observe instances of suboptimal performance. Combining model analysis with literature review, we believe that future models should also consider individual factors such as player technical characteristics, professional rankings, match experience, tactics, and mental states.We extended the application of the model to the US Open, women's tennis matches, and table tennis events, with results documented in the table 9. We found that the model demonstrated good generalization capabilities across tennis matches with varying venues and genders and relatively moderate in table tennis events.

Finally, We analyze the model sensitivity. **For Model I**, we calculate the CV(Coefficient of Variation) under different serve coefficients (S) to explore the sensitivity of the model to serve coefficients. **For Model II**, we integrated SHAP and PDP (Partial Dependence Plot) to further investigate the impact of individual features and combined features on the model output.

Keywords: Tennis; Momentum; AHP-BWM; XGBoost; SHAP

Contents

1 Introduction

1.1 Problem Background

Tennis, the world's second-largest sport, is widely played on a global scale, with renowned events such as the four Grand Slam tournaments capturing widespread attention each year. In particular, the men's singles final of the 2023 Wimbledon Tennis Championships witnessed the 20-year-old Spanish sensation, Carlos Alcaraz, triumph over the formidable Novak Djokovic. The intense match went the full five sets, lasting nearly five hours, providing tennis enthusiasts worldwide with a spectacular visual feast.

In addition to the collision of outstanding skills, what captures attention is the roller-coaster-like trajectory of the matches. People are astonished to discover that the seemingly advantaged side exhibits greater fluctuations in the subsequent games, prompting thoughts about the crucial factor influencing the course of the match - "momentum". Momentum is an abstract concept, accumulated through a series of actions and events in the game, akin to psychological factors. It is challenging to quantify but holds significant sway over athletes

In fact, countries around the world have gradually made tennis research at the technical and tactical levels more transparent. In the current intensifying landscape of tennis competitions, the key factors determining athletic performance have shifted from technical and tactical elements to psychological factors experienced by athletes during the matches. The study referenced as [1] explores the influencing factors of "momentum"and captures phenomena related to pivotal moments in momentum shifts. This research not only helps us better understand this abstract concept but also provides proactive strategic recommendations for athletes in their preparation and adaptation to changes on the competitive stage.

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

- Establish a mathematical model to assess the performance of athletes at a given time. Apply this model to one or more games and provide a visual description of the competition process.
- Examine the fluctuations in the "momentum"during the game and determine whether the success of athletes is random.
- Develop a model to predict the fluctuations in "momentum"during the game and explore factors more closely related to these fluctuations.
- Provide recommendations to athletes based on the differences in "momentum"fluctuations in past games against different opponents.
- Test the established model to explore its applicability to fluctuations in "momentum"during games, identify areas for improvement in case of poor performance, and assess its generalizability to other sports, playing surfaces, and competitions.
- Based on the results of the research, prepare a memorandum for the coach, providing some recommendations.

1.3 Our Work

Our main workflow is shown in the following figure 1.

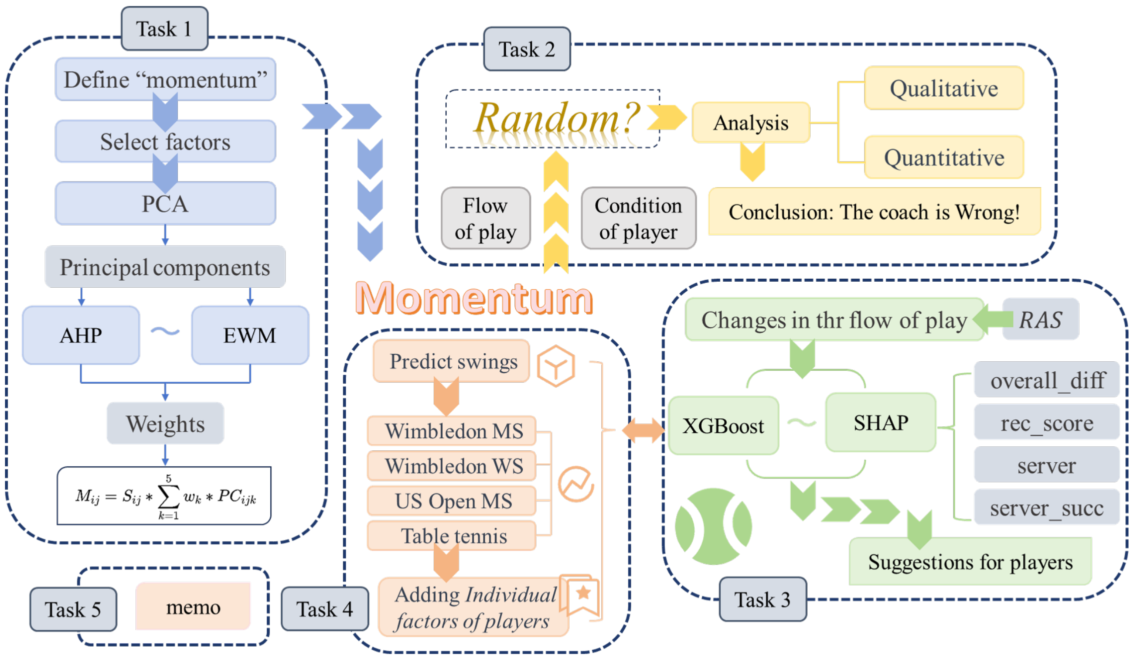


Figure 1: Our Work

2 Assumptions and Justifications

Based on the adequate analysis of problems, we make the following properly justified assumptions to simplify our models.

- **Assumption 1: Player's physiological functions, health conditions, and emotional states do not show significant differences during the game.**

  This helps capture the essence of the match more accurately, free from interference by individual variations.

- **Assumption 2: The competitive condition and playing style of professional tennis players remain stable throughout a tournament.**

  This enables the prediction of momentum fluctuations in the current round based on the performance trends from previous rounds.

- **Assumption 3: The data provided by the MCM officials is authentic and the recorded game data is unbiased.**

  This ensures that the data used by the model is both objective and effective.

- **Assumption 4: Disregarding external factors such as weather conditions and court facilities.**

  This enhances the simplicity and interpretability of the model.

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Definition
w	Weight of each principal component in AHP-EWM
S	Serve coefficient
M	Momentum
RM	Relative momentum
RMR	Relative momentum ratio
pw	Point won
pwn	Point won num
RA	Relative advantage
RAS	Relative advantage swing
rec_score	Recent scores
$overall\_py\_diff$	Overall score difference
$game\_key\_pt$	Key point in a game
set_key_pt	Key point in a set

4 Multi-angle Momentum Evaluation Model

4.1 Description of Momentum

"Momentum"is an elusive concept, capturing a player's short-term performance. Unlike the overall win rate, a player can trail significantly in the global game but, inspired by a few points, may experience a surge in "momentum", resulting in a burst of positive play. In tennis, we define "momentum"as a player's recent performance over a few points, influencing the likelihood of winning the next point.

According to the above the definition, "momentum"is a short-term concept, akin to derivatives in functions :$f' = \frac{dy}{dx}$.

Based on the above analysis and considering the following two factors:

- Typically, the closer the scores are, the greater the impact on the athlete.
- The problem is inherently discrete, while the concept of derivatives is continuous.

We utilize a sliding window, evaluating the "momentum"of both sides in the current score contention based on the most recent 3 rounds. Here, the recent 3 rounds are analogous to dx, and the relative change between an athlete's momentum and the opponent's, determined by evaluation criteria, is analogous to dy. The number 3 represents the window size in the sliding window concept.

4.2 Momentum Factor Indicator Selection

Various factors impact tennis "momentum". To quantify a player's "momentum", we conducted a literature review and made informed speculations. We preliminarily identified factors influencing "momentum", as depicted in Figure 2, which will be elaborated in the following sections.

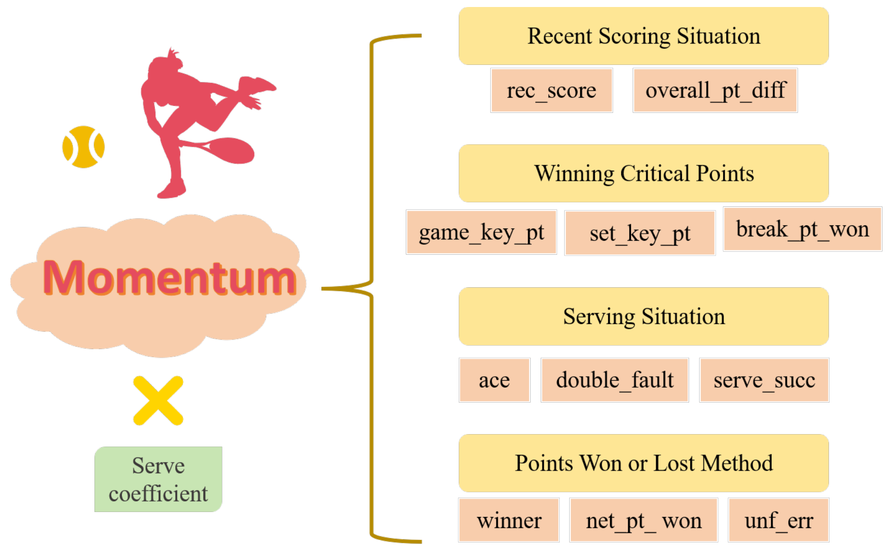


Figure 2: Overview of selected factors

4.2.1 Analysis of Recent Scoring Situation

Scores are the sole pathway to victory in tennis matches. Recent scoring accurately assesses a player's short-term performance. The smaller the score difference between opponents, the more intense the match, and the greater the impact on "momentum". Due to the dynamic and real-time nature of point differentials, better reflecting a player's current state, we use coefficients of 0.1, 0.3, and 0.6 to weight describe the differences in point, game, and set scores.

4.2.2 Analysis of Winning Critical Points

Research indicates a strong correlation between winning crucial points and overall match outcomes. Winning key points in recent matches, as discussed in [2], undoubtedly significantly enhances a player's "momentum". Considering the relative nature of the "momentum". values we calculate for two players, which represent a zero-sum game, we do not explicitly address situations where key points are lost. We have selected three specific indicators to assess the scenario of winning crucial points.

1. **Game Key Points:** We define game key points as scoring points that occur at a critical juncture within a game. Specifically, for a standard game, a point is considered a game key point if at least one player has a score (p_score) of 40 before this point occurs. The advantage of this selection and definition is that, under this definition of game key points, it encompasses all game points, including situations where the score is 40:40, which we consider equally significant. Additionally, we account for the uniqueness of "tiebreak"games, where, in a "tiebreak", if at least one player has a score of 6 or more, we consider it a game key point.

2. **Set Key Points:** Similar to game key points, if at least one player has a game score of 5 or more before this point occurs in a set, we consider it a set key point.
3. **Break Point:** Break points are typically crucial moments in a match. Not only are they inherently game key points, but also considering the difficulty players face in winning the receiving game, they naturally have a significant impact on a player's "momentum".

Considering that match points serve as both game break points and set break points, we haven't separately defined parameters related to match points as indicators. In practical processing, we sum the scores of game break points and set break points to assess the criticality of match points.

4.2.3 Analysis of Serving Situation

In tennis, serving is a powerful weapon controlled solely by the player. [3] Considering the relativity of "momentum", we specifically evaluate the serving player's performance, categorizing it into three situations:

1. **Ace:** When the serving player delivers an ace, it significantly enhances self-confidence.
2. **Double Fault:** Occurrence of a double fault often indicates a fluctuation in the serving player's mentality, exerting a certain inhibitory effect on the development of "momentum."
3. **Successful Serves:** The number of successful serves in a short period reflects the serve success rate, which research shows significantly impacts match outcomes . [4]

4.2.4 Analysis of Points Won or Lost Method

In tennis matches, winning or losing points through different means has diverse impacts on "momentum". We prioritize three situations, with cumulative scoring in case of overlap:

1. **Winner:** When a player's high-quality shot prevents the opponent from making a return, it undoubtedly enhances the player's "momentum."
2. **Net Point won:** Winning a point at the net not only secures points quickly but also acts as a deterrent to the opponent, providing a competitive advantage.
3. **Unforced Error:** It is easy to understand that when a player makes an entirely unnecessary error, it can have varying degrees of impact on their confidence.

4.3 Data Processing and Analysis

4.3.1 Harmonization of Evaluation Metric Types

Standardize the two extremely small indicators, double faults, and unforced errors, by converting them into extremely large indicators for consistency. For minimizing indicators $x_i (i = 1, 2)$ a translation transformation is applied: $x' = M_i - x_i$, where M_i represents the maximum possible value that the indicator m_i can take.

4.3.2 Dimensionless Processing of Evaluation Metrics

Applying Min-Max normalization to 11 indicators of n samples $x_{ij}(i = 1, 2, ..., 11; j = 1, 2, ...n)$. Letting $M_i = max\left\{x_{i1}, x_{i2}, ..., x_{in}\right\}, m_i = min\left\{x_{i1}, x_{i2}, ..., x_{in}\right\}$

the dimensionless values after normalization are

$$x_{ij}^* = \frac{x_{ij} - m_i}{M_i - m_i} \in [0, 1] \tag{1}$$

4.3.3 Spearman Rank Correlation Coefficient Test

Rank transformation on the sample data yielded Spearman rank correlation coefficients, along with significance p-values. The following heatmap illustrates the results.

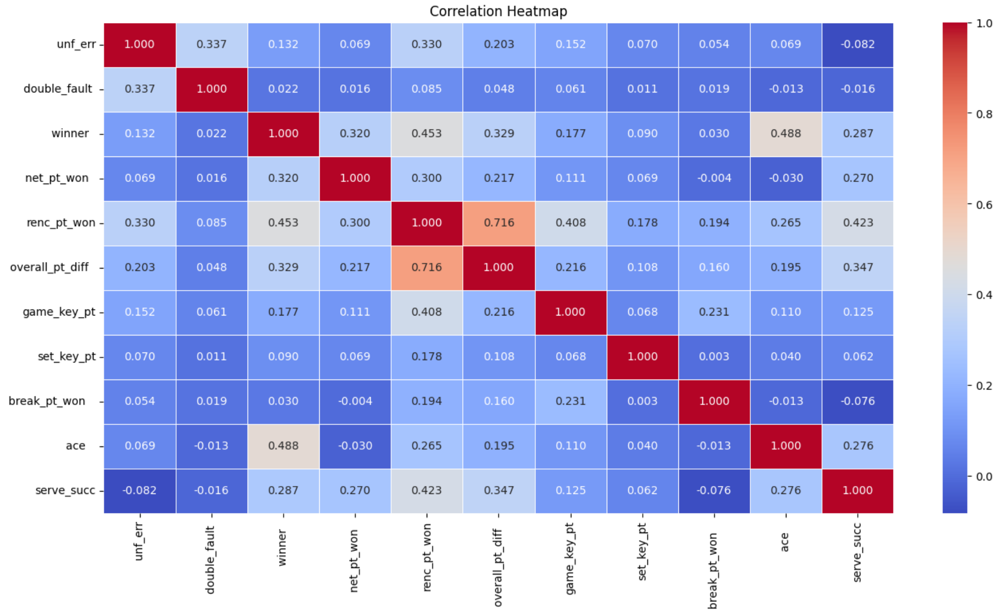


Figure 3: Spearman Rank Correlation Coefficient Test

The p-values show no correlation between key games, net points won, and break points. There is no correlation between ACE and double faults. However, significant correlations exist among the remaining indicators, indicating the need for dimensionality reduction.

4.3.4 Data Partitioning and Selection

For a more reasonable evaluation, we use 80% of the data provided by the MCM official (first 25 matches) to build a comprehensive evaluation model based on AHP and EWM. This amount of data is already more than sufficient. The remaining 20% of the data (6 matches, from 1502 to 1701) will be used for subsequent verification of the correlation between momentum and volatility.

4.4 Principal Component Analysis

For momentum assessment, while each indicator has distinct underlying causes, there exist correlations among different indicators. To uncover these latent factors and their respective dominant effects, this study employs PCA to solve these issues.

4.4.1 KMO and Bartlett's Tests

We conduct the KMO and Bartlett's tests to determine the feasibility of conducting PCA. The result of the KMO test is 0.669 and Bartlett's sphericity test yields a significance p-value of 0.000, showing statistical significance. Hence, principal component analysis is effective.

The eigenvalues and cumulative contribution rates for the principal components are presented in the table 2 below:

Table 2: Factor Weight Analysis

Component	Variance Explained	Cumulative Variance Explained(%)	Weight(%)
Principal Component 1	0.29818	29.818	38.160
Principal Component 2	0.16123	45.941	20.634
Principal Component 3	0.11571	57.512	14.808
Principal Component 4	0.10978	68.490	14.049
Principal Component 5	0.09649	78.139	12.348

With a cumulative variance of 78.139%, we identified 5 principal components. Notably, Principal Component 1 and Principal Component 2, with high cumulative contributions, may be key indicators for potential energy assessment.

The heatmap in Figure 4 displays the factor loading matrix from principal component analysis, enabling the assessment of the significance of latent variables in each principal component.

	unf_err	double _fault	winner	net_pt _won	rec_score	overall _pt_diff	game _key_pt	set _key_pt	break_ pt_won	ace	serve _succ
pc1	0.373	0.149	0.685	0.444	0.881	0.747	0.471	0.326	0.187	0.472	0.552
pc2	0.645	0.589	-0.279	-0.136	0.109	0.046	0.269	0.072	0.374	-0.362	-0.464
pc3	0.379	0.515	0.238	-0.058	-0.131	-0.152	-0.377	-0.028	-0.614	0.357	0.011
pc4	-0.001	-0.084	0.21	-0.615	-0.045	-0.063	0.149	-0.262	0.366	0.609	-0.203
pc5	0.001	-0.175	-0.121	-0.379	0.04	0.049	-0.058	0.848	-0.139	0.074	-0.094

Figure 4: Factor Loading Matrix Heatmap

In the Momentum Evaluation indicators, PC1 combines serve_succ, game_key_pt, overall_pt_diff, rec_score, and winner. PC2 combines double_fault and unf_err. PC3 mainly involves break_pt_won. PC4 combines ace and net_pt_won. PC5 is related to set_key_pt. This provides a comprehensive assessment of momentum across the 11 indicators.

4.5 Establishment of the Momentum Evaluation Model

4.5.1 Analytic Hierarchy Process

In this section, we employ AHP (Analytic Hierarchy Process) to establish a set of scoring rules. AHP is a decision analysis method that relies on decision-makers' experiential judgments to assess the relative importance of criteria in determining the feasibility of various measurement objectives. [5]

- **Construction of Hierarchical Structure Model**

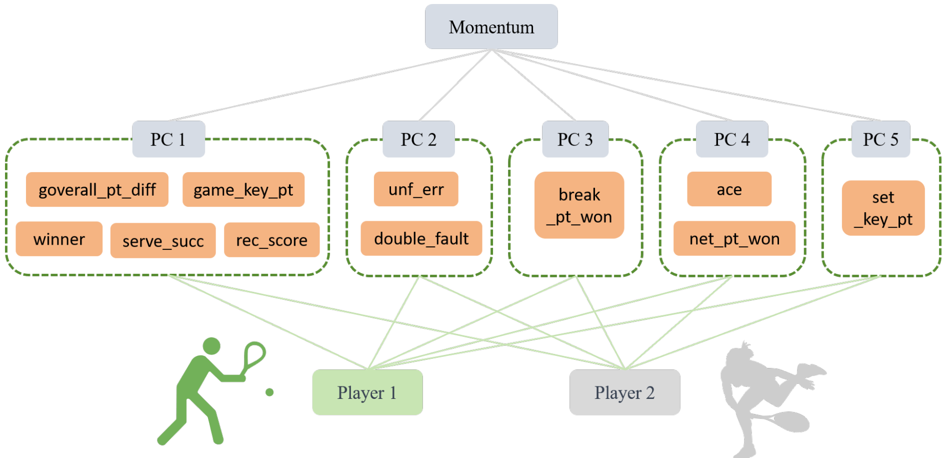


Figure 5: Hierarchical structure model

- **Construction of Judgment matrix** First, we create a judgment matrix using literature and experience, and then normalize it.

$$J_{ij} = J_{ij} \div \sum J_{ij} \tag{2}$$

  Then, based on the matrix, we calculated the weight vector and obtained the eigenvalues.

- **Inspection and Evolution**

  Next, a consistency check is conducted, and the results are shown in the following table 3.

Table 3: AHP Consistency Check Results

Consistency Check Results				
Maximum Eigenvalue	CI Value	RI Value	CR Value	Consistency Check Result
5.06	0.015	1.11	0.013	passed

  The calculation results indicate that the maximum eigenvalue is 5.06. According to the RI table, the corresponding RI value is 1.11. Therefore, it passes the consistency check.

$$CR = \frac{CI}{RI} = 0.013 < 0.1 \tag{3}$$

4.5.2 Entropy Weight Method

Entropy weight method is a multi-attribute decision analysis approach that assigns weights to each attribute by considering the differences in information entropy among attributes, facilitating comprehensive evaluation and decision-making. [6]

- **The calculation of the information entropy of each variable**
  As the steps of"Construction of matrix"and "Normalization of Indicators"have been previously completed, we can now directly use the following formula to calculate the entropy of the $j^{th}, (j = 1, 2, 3, 4, 5)$ variable.

$$H_j = -\frac{1}{\ln n} \sum_{i=1}^{5} p_{ij} * \ln p_{ij}, \quad p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{n} y_{ij}} \tag{4}$$

- **The calculation of the weight of each indicator**

  The weights of each indicator can be calculated according to the following formula:

$$W_j = \frac{1 - H_j}{P - \sum_{j=1}^{p} H_j} \tag{5}$$

4.5.3 Establishing the AHP-EWM Integrated Evaluation Model

We obtained two different sets of weighting factors, one being subjective and the other objective. Since they are respectively subjective and objective, we combine them and calculate the weighted average of the weights obtained from both methods. In the cases of AHP, EWM, and AHP-EWM, the weight distribution of each principal component is illustrated in Figure 6 as follows:

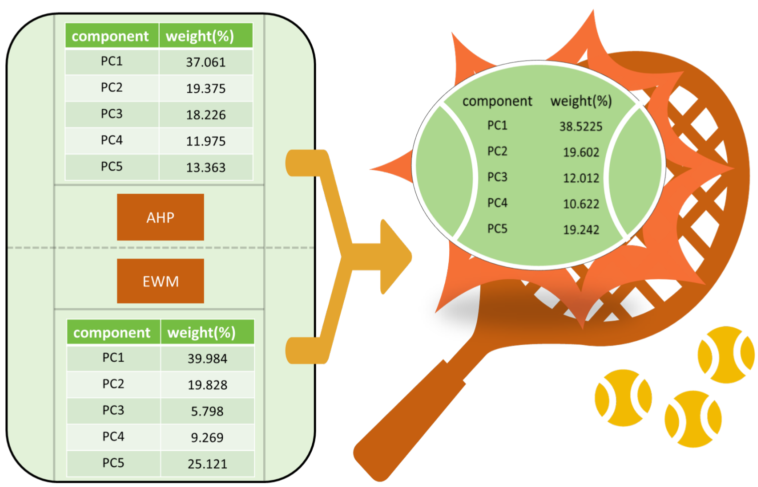


Figure 6: Three methods of scoring weight table

The final weights obtained by the AHP-EWM method are referred to as: $w_k, (k = 1, 2, 3, 4, 5)$.

Research indicates that in tennis matches, the serving side has a notably higher probability of winning points. Consequently, a player's psychological state is more influenced by their performance in serving rounds, impacting their "momentum"values. [7] Different coefficients are set based on whether the player is serving or receiving. The serving player's serving coefficient S is set to $S_t = 1.1$, while the receiving player is set to $S_f = 1.0$.

Combining the above analysis, for a match, assuming the total number of points in the match is N, the momentum values of two players $i(i = 1, 2)$ at the $j^{th}(j = 1, 2, \ldots, N)$ point are:

$$M_{ij} = S_{ij} * \sum_{k=1}^{5} w_k * PC_{ijk} \tag{6}$$

where $S_{ij} = \{S_t, S_f\}$ is the serving coefficient of player i in the j^{th} match, $w_k(k = 0, 1, ...5)$is the weight of the k^{th} principal component, and PC_{ijk} is the value corresponding to the k^{th} principal component for player i at the j^{th} point.

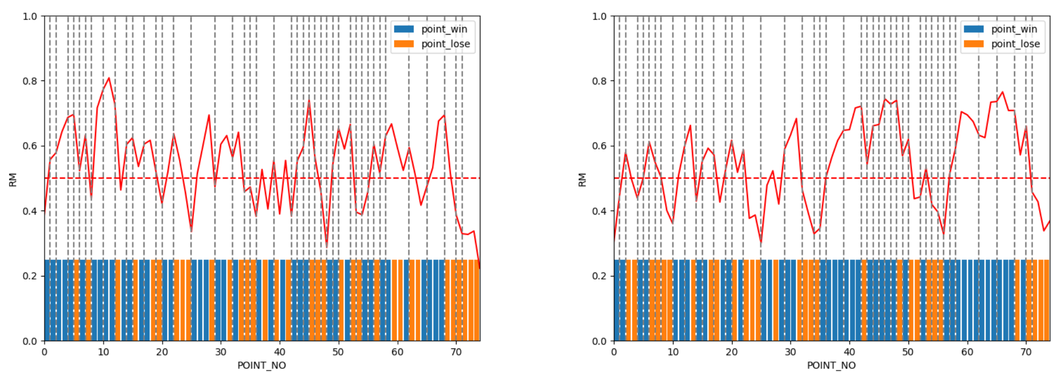
Due to the relativity of momentum, to more accurately and conveniently describe the likelihood of a player scoring at the j^{th} point, the following formula is used to calculate the relative momentum $(RM_i(i = 1, 2))$ between Player 1 and Player 2:

$$\begin{cases} RM_{1j} = \frac{M_{1j}}{M_{1j}+M_{2j}} \\ RM_{2j} = \frac{M_{2j}}{M_{1j}+M_{2j}} \end{cases} \tag{7}$$

Intuitively, the larger the relative momentum of a player at a particular score, the better their condition at that moment, indicating a higher likelihood of winning that point.

4.6 Application of the Model

We utilized the first two unanalyzed matches from the dataset (match_id = 2023-wimbledon-1502 and match_id = 2023-wimbledon-1503) for the model. Subsequently, we generated a trend chart illustrating Player 1's relative momentum from the 1st to the 75th point of each match. The outcomes are presented in Figure 7:



(a) 2023-wimbledon-1502

(b) 2023-wimbledon-1503

Figure 7: The relative momentum change chart for Player 1 in the two matches

The red fluctuation line in the graph represents the relative momentum changes of Player 1. The RM value at a given point_no reflects the performance of Player 1 in the recent period. A higher RN value indicates better performance compared to a lower RM value. If the RM value is greater than 0.5, it means that Player 1 has a stronger momentum relative to Player 2 at that time. Additionally, the performance can be measured by the specific value of RM or the indicator $RMR(Relative\ Momentum\ Ratio) = \frac{RM}{1-RM}$. The larger the values of RM or RMR, the better the performance.

If a qualitative assessment is desired, after statistical analysis, the degree of good performance can be determined based on the corresponding values in the table 4 below.

Table 4: Performance Evaluation Form

Correspondence between RM Value and Performance Situation						
Outstanding	Very Good	Good	Average	Fair	Poor	Very Poor
$>$0.7	0.65–0.70	0.55–0.65	0.45–0.55	0.35–0.45	0.3–0.35	$<$0.3

For example, in Figure 7a, at point_no=11, Player 1's RM value exceeds 0.8, and the calculated RMR value surpasses 4, indicating an outstanding performance by Player 1 at that moment. The bar chart below visually displays Player 1's win-loss situation at the specified point_no. Blue indicates Player 1 winning the point, and orange signifies losing the point. In Figure 7b, during intervals from point_no=38 to point_no=48 and from point_no=57 to point_no=67, Player 1's RM values mostly ranged between 0.6 and 0.75, indicating exceptional performance and dominance. The bars below are predominantly blue, confirming Player 1's strong performance during intervals with higher RM values.

5 Analysis of the Impact of Momentum on Matches and Players

This section's overview is illustrated in the following figure, as shown in Figure 8.

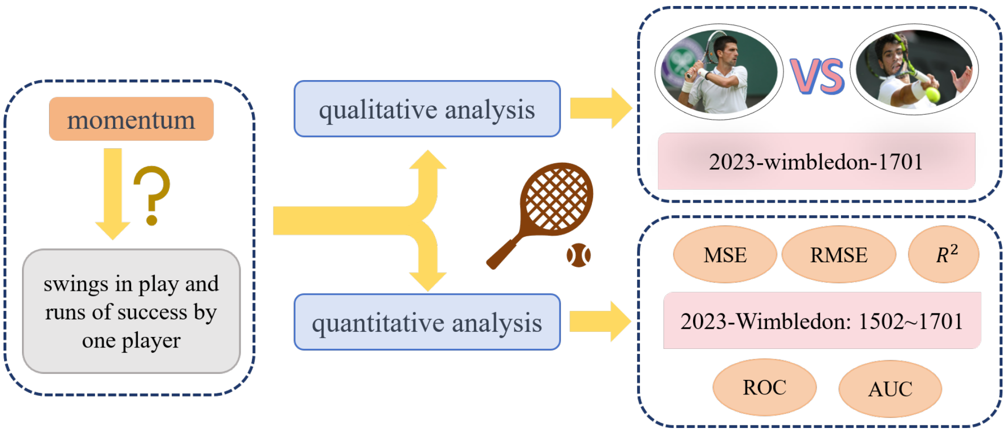


Figure 8: Task 2 overview

5.1 Visualization and Qualitative Analysis

In the qualitative analysis, we selected a match (match_id = 2023-wimbledon-1701) that has not been used to construct the comprehensive evaluation model. This match corresponds to the men's singles final of the 2023 Wimbledon Tennis Championship, with Player 1 being Carlos Alcaraz and Player 2 being Novak Djokovic. In the following figure 9, the x-axis represents the momentum and score comparison of the players during various time intervals in this match:

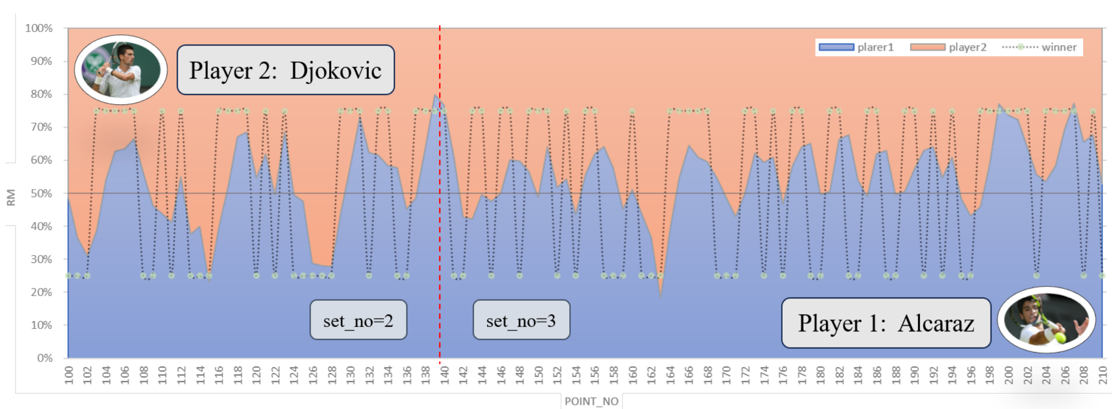


Figure 9: Momentum and score comparison chart

Scattered points above indicate Player 1 won, while points below indicate Player 2's victory. The blue area below and the pink area above represent the momentum magnitudes for Player 1 and Player 2. Observing momentum fluctuations in relation to actual scores reflects momentum's impact on the match. It can be observed that the small peaks where the player's momentum is prominent are often enclosed by small trapezoids formed by connected scattered points. In other words, when having the advantage in momentum, there is a higher likelihood of winning the point.

Now let's consider dividing this figure into two segments:

- From the 100^{th} to the 140^{th} point, in the middle to later stages of the second set, the chart indicates even performance between the two players, with momentum fluctuations showing a relatively stable and periodic pattern. Towards the end of this segment, Player 1 gained overwhelming momentum. Combining this with the raw data, we observed intense competition in the second set, with the players taking turns winning until reaching a tiebreaker. Ultimately, Alcaraz won the tiebreaker, and at that point, his momentum significantly surpassed Djokovic, aligning well with the actual circumstances.
- From the 141^{th} point to the 210^{th} point covers the entire third set. As depicted in the graph, during this period, Alcaraz had stronger momentum, consistently dominating Djokovic. By considering the raw data, it's evident that Alcaraz secured a significant victory in the third set with a score of 6:1, showcasing a clear advantage in the overall outcome.

The observations and analyses above suggest that momentum influences the match situation, and changes in the match situation reflect momentum fluctuations. Therefore, qualitatively, the relationship between momentum and shifts in play and runs of success by one player is not random; instead, it exhibits a higher level of correlation.

5.2 Evaluation Metrics and Quantitative Analysis

Visual results suggest momentum plays a significant role in determining the outcome of the competition. To objectively measure its impact, we will use quantitative analysis. Using our momentum definition and calculation results from Question One, momentum values and relative momentum can assess a player's likelihood of winning a specific point. Higher momentum increases the probability of winning, enabling the creation of a binary classifier with the momentum indicator. Selecting the confidence threshold α determines the point outcome, resulting in the prediction $\hat{y}$ for score achievement:

$$\hat{y} = \begin{cases} 0 & \text{if } RM \leqslant \alpha \\ 1 & \text{if } RM > \alpha \end{cases} \tag{8}$$

5.2.1 MSE, RMSE and R^2

Considering the use of momentum to assess the expected score $E(0 \leq E \leq 1)$, with RM as the predicted value sequence and the scoring situations as the actual value sequence, the calculated values of MSE, $RMSE$ and R^2 are shown in the table 5 below.

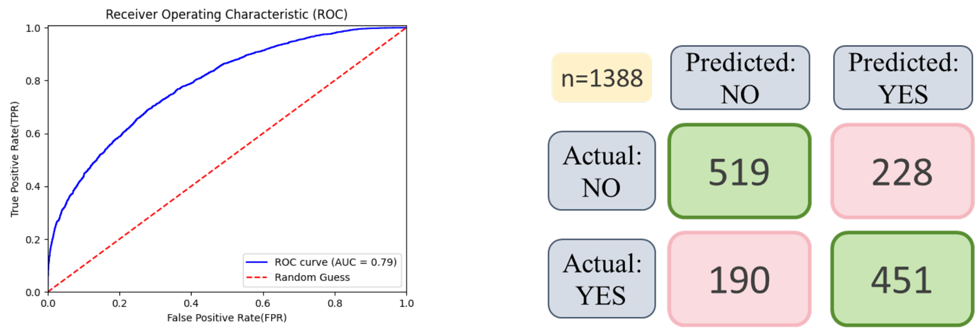
Table 5: The value of MSE, RMSE and R^2

MSE	RMSE	R^2
0.092	0.303	0.78

Under random prediction conditions, $RMSE$ should be 0.5. With $RMSE$ = 0.303, it indicates that the numerical fluctuations of the predicted sequence relative to the actual sequence are only 60% compared to random predictions, demonstrating better predictive accuracy. $R^2 = 0.78$ indicates that using RM for predictions shows strong explanatory power for scoring situations.

5.2.2 ROC and AUC

Using RM as the classification metric for a binary classifier, with the actual values representing the scoring situations, the ROC curve is shown in the figure 10 (a) below at different classification thresholds. The Area Under the Curve (AUC) for this ROC curve is 0.79.



(a) ROC curve (b) Confusion Matrix

Figure 10: ROC and Confusion Matrix

The binary classifier utilizing momentum as an indicator exhibits excellent discrimination performance. This implies that momentum plays a crucial role in determining match outcomes, suggesting a significant impact on match results. It indicates that match fluctuations and outcomes are not entirely random.

6 XGBoost-SHAP Interpretable Swings Prediction Model

This section's overview is illustrated in the following figure, as shown in Figure 11.

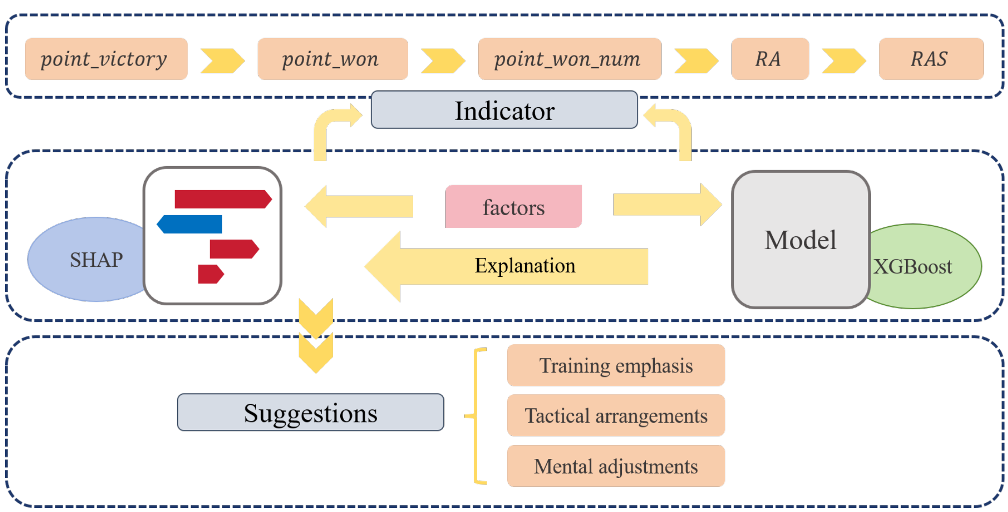


Figure 11: Task 3 overview

6.1 Construction of Indicator for Changes in the Flow of Play

The most intuitive factor reflecting the flow of play is the scoring situation. Therefore, we base our subsequent considerations on point_victory. In order to better describe the changes in the flow of play, it is necessary to construct relevant indicators based on the following steps (considering the relativity of both players, let's continue to focus on Player 1):

1. **point_won(pw):** $pw = \{1, 0\}$ is defined as whether Player 1 wins the point. For Player 1:
$$pw_t = \begin{cases} 1, & \text{if } point\_victory = 1 \\ 0, & \text{if } point\_victory = 2 \end{cases} \tag{9}$$
where $t = point_{no}$.

2. **point_won_num(pwn):** $pwn = \{0, 1, 2, 3, 4, 5\}$ is defined as the number of points scored from two points before this point to two points after this point, totaling 5 points. This definition is made because the actual flow of play in this point needs to be considered in conjunction with the actual score situation before and after this point. Additionally, it aligns with the concept that momentum is a relatively short-term phenomenon mentioned earlier. The calculation is as follows:
$$pwn_t = \sum_{i=t-2}^{t+2} pw_i \tag{10}$$

3. **relative_advantage(RA):** $RA = \{1, 0, -1\}$ is defined as the situation of relative advantage, where 1 indicates a relative advantage, 0 indicates an even match, and -1 indicates a relative disadvantage. The calculation is as follows:
$$RA_t = \begin{cases} 1, & \text{if } pwn_t = 5 \text{ or } 4 \\ 0, & \text{if } pwn_t = 3 \text{ or } 2 \\ -1, & \text{if } pwn_t = 1 \text{ or } 0 \end{cases} \tag{11}$$

4. **relative_advantage_swing(RAS):** $RAS = \{1, 0, -1\}$ is defined as the situation of relative advantage swing, where 1 indicates a swing towards advantage, 0 indicates no significant swing, and -1 indicates a swing towards disadvantage. The calculation is as follows:
$$RAS_t = RA_t - RA_{t-1} \tag{12}$$

Thus, the changes in the flow of play can be measured by the value of RAS_t.

6.2 XGBoost Prediction Model Establishment

6.2.1 Analysis of Model Selection Reasons

XGBoost is a gradient boosting algorithm that trains and predicts using an ensemble of multiple decision tree models. In situations with limited dataset size, XGBoost excels due to its adaptability to small datasets. Built upon gradient boosting, XGBoost sequentially adds decision tree models, with each tree aiming to correct the residuals of the previous one, gradually improving the overall model performance. The core objective function and principle diagram are depicted below:

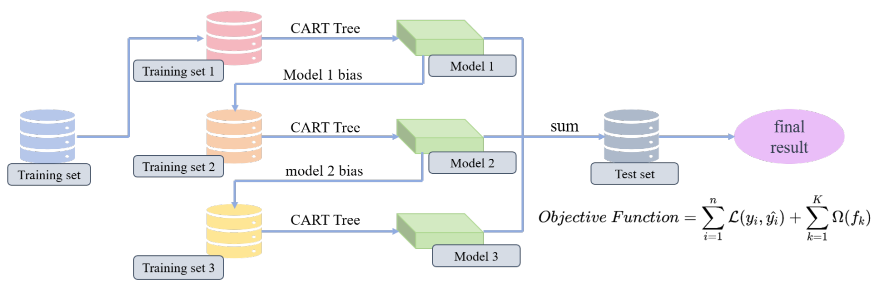


Figure 12: XGBoost principles diagram

Here, $\mathcal{L}$ represents the loss function, $y_i$is the true value, $\hat{y}$ is the model's predicted value, and $\Omega(f_k)$ is the regularization term concerning each tree f_k, employed to control the complexity of the model.

Considering that our training dataset comprises only 4284 samples, coupled with XGBoost's superior performance on small-scale datasets and its gradient boosting mechanism, we chose XGBoost as the prediction model.

6.2.2 Analysis of Reasons for Feature Selection

Through the argumentation in tasks one and two, we can conclude that momentum significantly influences the swings in the match. Consequently, we have chosen to use the 13 features (12 initially selected features plus point_victor used as the label) employed in the comprehensive evaluation model as the input variables for our model. We combine favorable and unfavorable swings into a single category, representing swings with label 1, while instances with no significant swings are labeled as 0. These together form the outcome vector.

6.2.3 Sample Balance

Due to the relatively infrequent occurrence of stable conditions in actual competitions compared to volatile situations, the ratio of stable fluctuation samples to volatile fluctuation samples in our training set is 5057:2025. The imbalanced distribution of data samples can affect the effectiveness of model training. Therefore, we removed 3000 samples with a label of 0 to achieve a balanced dataset.

6.2.4 Model Result Evaluation

After parameter tuning, the final selection includes a learning rate of 0.05, a maximum tree depth of 10, and the construction of the model using five-fold cross-validation. The training process took 18.3 seconds. The specific details are presented in the following table 6:

Table 6: Model evaluation results

	Accuracy	Recall	Precision	F1 score
Training set	0.863	0.831	0.842	0.856
Test set	0.842	0.833	0.8728	0.766

The model demonstrates good performance in predicting fluctuations in tennis matches, with an accuracy of 84.2% and precision of 73.3% on the test set. The recall rate and F1 score are relatively high, indicating the robustness and accuracy of the model in predicting fluctuation categories.

6.3 SHAP Analysis of Feature Importance

6.3.1 Analysis of Model Selection Reasons

An overarching issue in machine learning is the relatively poor interpretability. While the XGBoost algorithm provides feature_importance to showcase the most important N features of the model, the situation for individual samples may not align with the overall model. Therefore, the SHAP algorithm is employed to reveal the contribution of different features in a sample with numerical values. The key advantage of SHAP lies in its ability to reflect the impact of each feature within each sample, displaying both the positivity and negativity of the influence. [8]

6.3.2 Principles and Calculation of SHAP Values

SHAP is an additive interpretability model inspired by Shapley values. It interprets the model's prediction as the sum of estimated values for each input feature. For each predicted sample, the model generates a prediction, and SHAP values allocate values to each feature in that sample. Assuming x_{ij} represents the j^{th} feature of the i^{th} sample, the model's prediction for the i^{th} sample is y_i and y_{base} is the mean of the target variable for all model samples, then the SHAP values follow the equation:

$$y_i = y_{base} + f(x_{i,1}) + f(x_{i,2}) + ... + f(x_{i,13}) \tag{13}$$

Where $f(x_{i,j})$ is the SHAP value for $x_{i,j}$, representing the contribution of the j^{th} feature to the final prediction. When $f(x_i, 1) > 0$, it implies that the feature increases the predicted value; conversely, it means the feature decreases the predicted value.

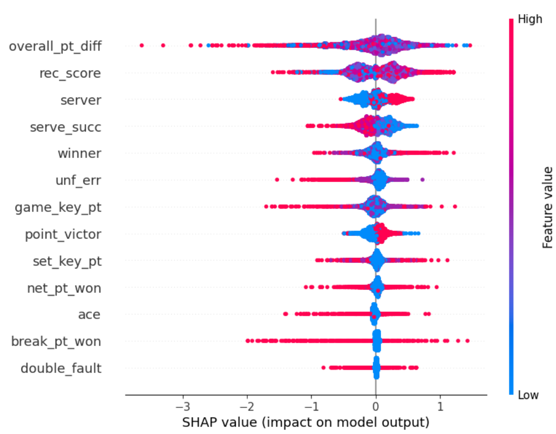
Specifically, when the model is nonlinear or input features are not independent, SHAP values calculate the weighted average of all possible feature rankings. According to the following equation:

$$\phi_j = \sum_{S \subseteq \{x_1, \cdots, x_{13}\} \setminus \{x_j\}} \frac{|S|!(13 - |s| - 1)!}{13!} (f_x(S \cup \{x_j\}) - f_x(S)) \tag{14}$$

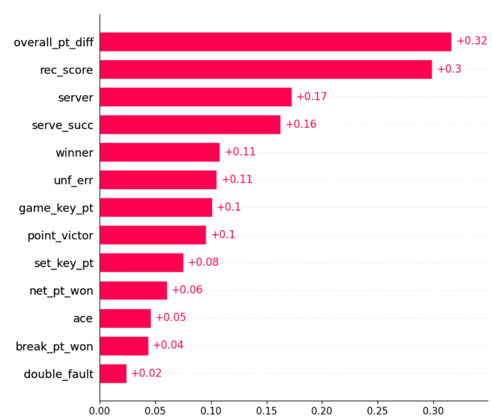
Where $\{x_1, \cdots, x_{13}\}$ is the set of all input features, $\{x_1, \cdots, x_{13}\} \setminus \{x_j\}$ is the set of all input features excluding $\{x_j\}$, and $f_x(S)$ represents the prediction for the feature subset.

6.3.3 Results Analysis

Through SHAP analysis, we obtained feature importance. The visualization results and analysis are as follows. Figure 13(a): Bee swarm plot depicting the impact of features on samples. Each row represents a feature, with SHAP values on the x-axis. Each point represents a sample, where a redder color indicates a higher feature value, and a bluer color indicates a lower feature value. We can visually observe that the overall point difference is a crucial feature, and a larger point difference tends to decrease the probability of match swings. The red-blue concentration is significant in whether it is the serving side and the number of successful serves, which noticeably affects fluctuations. Specifically, being the serving side is negatively correlated with fluctuations, while the number of successful serves is positively correlated with fluctuations.



(a) SHAP value



(b) mean |SHAP value|

Figure 13: SHAP calculation results

Figure 13(b): The absolute mean-sorted impact of features on swings. It can be observed that the overall point difference and the recent scores for the last 33 balls have the greatest impact on fluctuations. Following these, whether it is the serving side and the recent success count of serves also exhibit significant influences on the feature.

6.4 Exploratory Analysis and Advice

6.4.1 Exploratory Analysis

We conducted an analysis of the fluctuation differences in past matches and generated heatmap plots illustrating the impact of factors on historical swings:

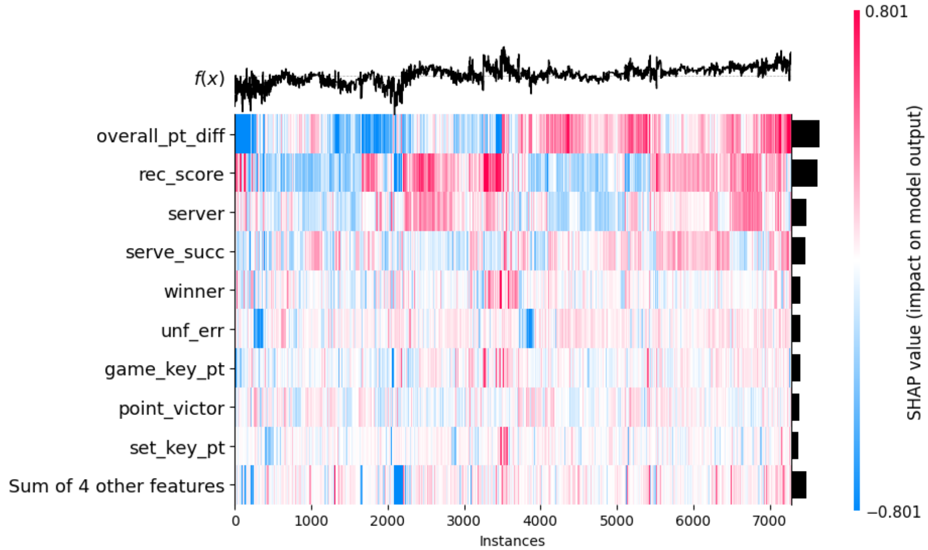


Figure 14: Heatmap for factors and swings

The feature heatmap visualizes how model input features relate to instances using SHAP values. Instances are on the x-axis, features on the y-axis. Default hierarchical clustering highlights similar model output instances. Model outputs are shown above the heatmap, and global importance is indicated by the right-side bar chart. Darker colors in the heatmap signify higher feature correlation.

Positive SHAP trends for overall point difference indicate increased importance in momentum fluctuations when leading or trailing by a specific score. Significant SHAP values for recent 3-point scores suggest their substantial impact, potentially being decisive. SHAP values for serving status indicate an advantage in controlling the game's pace or being influenced by opponent counterattacks. SHAP values for successful serves may reveal serving's role in momentum shifts. This information helps teams devise smarter game tactics.

6.4.2 Suggestions for preparation

1. **Training Emphasis:** Focus on specific match volatility indicators during training and develop the ability to recognize them. From the analysis above, it is evident that changes in point differentials, the rotation of serving sides, and the success or failure of serves have the most significant impact on match volatility. Paying adequate attention to these indicators is crucial for players to grasp match fluctuations effectively.
2. **Tactical Arrangement:** Make corresponding tactical adjustments when volatility occurs to enhance the ability to respond to and control match fluctuations. In situations favorable to oneself, employ a stable tactical strategy based on the impact of various indicators to stabilize

one's advantage. In situations where momentum is balanced or the opponent has the upper hand, adopt a tactical strategy that promotes match volatility to regain control of the game.

3. **Mental Adjustment:** Facing unfavorable situations such as the opponent scoring or personal mistakes, maintain psychological stability, focus on the present, and avoid letting emotions impact performance. In favorable situations, stay calm and avoid becoming overly confident.

7 Model Application and Generalization Testing

7.1 Data Source and Description

We obtained all data sources and competition information used for Task 4 model testing by scraping multiple websites. The details are as follows:

Table 7: Data source collation

Match Type	Match Session	Data Source Website
2023 Wimbledon MS & WMS	1101|1102|1103|1231|1232 2503|2504|2601|2602|2701	https://www.ultimatetennisstatistics.com/ https://www.tennisvisuals.com/
2023 US Open MS	1102|1105|1107	https://www.usopen.org/ https://www.livesport.com/en/
2020 Olympic Table Tennis MS	Zhendong Fan VS Ma Long	https://olympics.com/olympic-games

7.2 Model Application Testing

Testing was conducted on the data of the five 2023 Wimbledon men's singles matches as shown as follows:

Table 8: 2023-wimbledo-MS test

2023-wimbledo-MS test						
Maximum Eigenvalue	1101	1102	1102	1231	1232	all
Accuracy	83.8%	83.2%	82.2%	76.6%	84.2%	83.6%

Analyzing the results in Table 8, our predictions for fluctuations in the matches were quite good, with an overall accuracy of 83.6%. The highest accuracy was achieved on 2023-wimbledo-1232, reaching 84.2%. However, the accuracy for predicting fluctuations in the match 2023-wimbledo-1232 was relatively low at 76.6%, significantly differing from other match instances. A check on Wikipedia revealed a substantial difference in the world rankings between the players Stan Wawrinka and Tomás Martín Etcheverry in this particular event. Further analysis will focus on the first set swing in this match:

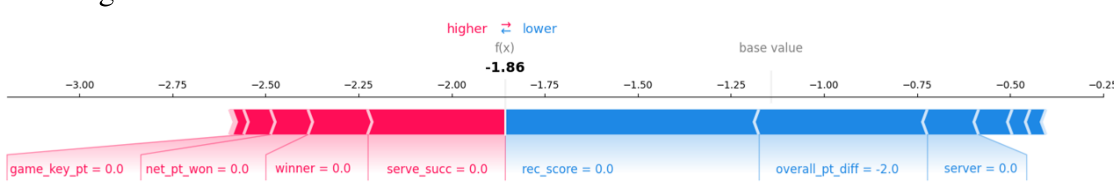


Figure 15: Analyse of 2023-wimbledo-1232

In cases of significant skill gaps, point differentials play a dominant role, diluting the impact of the sliding window metrics we employed. Clearly, match fluctuations are influenced by multiple factors. Based on the above analysis and experience, we believe it's essential to consider factors related to the players themselves, such as their technical characteristics, professional rankings, match experience, tactics, and psychological state.

7.3 Evaluation of Model Generalization Capability

To assess the model's generalization, we chose test sets based on gender, court surface, and sports type. Considering gender, we used five women's singles tennis matches from the 2023 Wimbledon events (2503, 2504, 2601, 2602, 2701). For court surface, we selected five men's singles tennis matches from the 2023 US Open (1102, 1105, 1107), where Wimbledon had grass surface, and the US Open had a hard court surface. Regarding sports type, we picked the match between Fan Zhendong and Ma Long from the table tennis event at the 2020 Tokyo Olympics. [9]

After processing these three datasets, the sample sizes were 894, 1354, and 226, respectively. The confusion matrices post-testing are shown in the following figures 16(a):

1354 samples	Condition positive	Condition negative
Test outconme positive	364	88
Test outconme negative	64	378

(a) Wimbledon WMS

1354 samples	Condition positive	Condition negative
Test outconme positive	511	202
Test outconme negative	140	501

(b) USopen MS

226 samples	Condition positive	Condition negative
Test outconme positive	83	30
Test outconme negative	32	81

(c) Olympic Table Tennis MS

Figure 16: The confusion matrix for different test sets results

Calculating various metrics based on the three confusion matrices in the above figures, the results are presented in the table below:

Table 9: Results of different matches

Match sessions	Accuracy	Precision	Recall	F1 Score
Wimbledon WMS	83.4%	83.4%	80.2%	82.7%
USopen MS	79.8%	78.1%	82.3%	80.1%
Olympic Table Tennis MS	74.2%	72.4%	70.1%	71.2%

According to table 9, the model performs well in the Wimbledon women's singles and the US Open men's singles, with high accuracy, balanced precision and recall, and a higher F1 score. This indicates strong generalization across different gender and competition categories. However, on the Olympic table tennis men's singles, the model shows relatively poorer performance, with lower accuracy, precision, and recall compared to the first two datasets, suggesting weaker generalization in the domain of table tennis.

8 Sensitivity Analysis

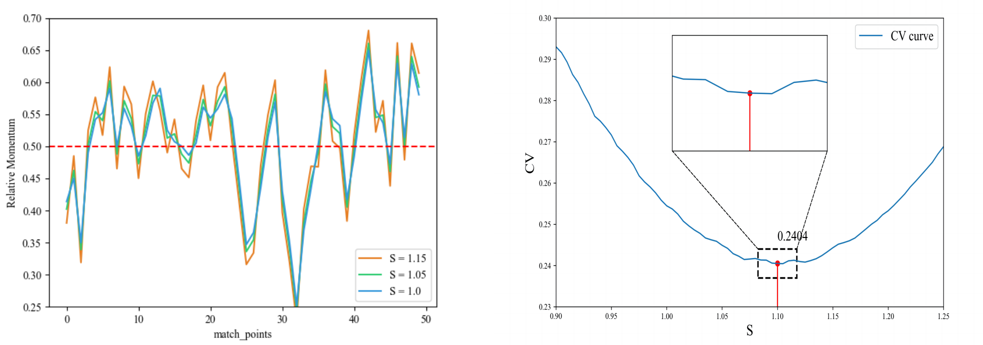
8.1 Sensitivity Analysis of Multi-angle Momentum Evaluation Model

In the model, the serve coefficient S_t significantly influences the model results, necessitating a sensitivity analysis to determine the optimal choice.

Due to the significant fluctuations of momentum, making it challenging for intuitive visualization (as shown in Figure 17(a)), we employ the coefficient of variation (CV) to characterize the variability of momentum distribution under different S values.

$$CV = \frac{\sigma}{\overline{y}} \times 100\% \tag{15}$$

Here, σ represents the standard deviation of momentum, and $\overline{y}$ is the mean of momentum. Calculating the CV values under different S values yields the following results:



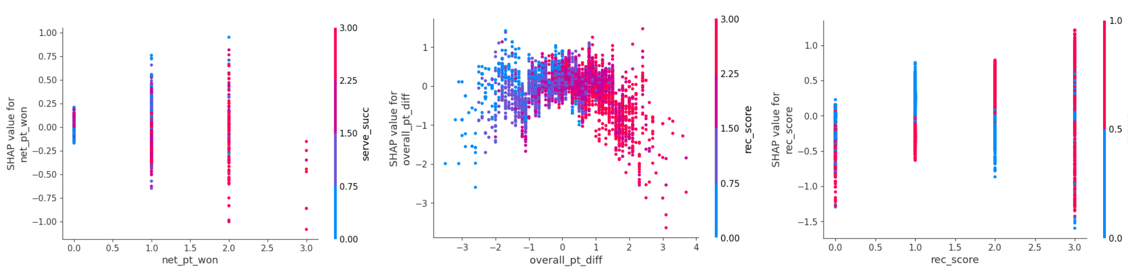
(a) Swings visualization (b) Coefficient of variation curve

Figure 17: Visualizations of *Model-I* sensitivity analysis

From the figure 17(b), it can be observed that the CV reaches its minimum value of 0.2404 at $S = 1.100$. As the S value increases or decreases, the overall trend of the CV shows a significant upward trend, indicating an increasing fluctuation and higher instability in the momentum. Therefore, we choose $S = 1.1$ as our optimal serve coefficient.

8.2 XGBoost sensitivity analysis.

We have investigated the impact of individual features on the model outputIn Task 3. Now, we delve deeper into the influence of individual features and feature combinations on the model output through Partial Dependence Plots (PDP). We continue to use the magnitude of SHAP values as an indicator to assess feature importance. For our model, the y-axis of PDP is in units of the logarithmic odds of match fluctuations. The color corresponds to the value of a second feature that may interact with the current feature. If there is interaction, it will be displayed as a distinctive vertically shaded pattern. Below are the PDP plots for the top four features in terms of importance:



(a) Wimbledon WMS (b) USopen MS (c) Olympic Table Tennis MS

Figure 18: Partial Dependence Plots

Analysis of the Pdp plots leads to the following conclusions:

In figure 18(a), when the net score is 0, an increase in successful serves has a significant positive impact on the output. However, when the net score is 1, there is a negative correlation between the number of successful serves and the output.

Figure 18(b) indicates that the overall point difference has almost no interaction with recent 3-point scores, but there is a strong positive correlation between the features. To optimize the model, consideration can be given to dimensionality reduction for the features.

Figure 18(c) highlights that when the recent 3-point score is 1, the serving side has a greater impact on the output than the non-serving side. Conversely, when the recent 3-point score is 2, the non-serving side has a greater impact on the output than the serving side.

9 Strengths and Weaknesses

9.1 Strengths

- Adopting a combined approach of AHP and EWM, we integrate subjective and objective measures, enhancing the accuracy and credibility of the evaluation.
- SHAP demonstrates prominent advantages in reflecting the impact of features within each sample, addressing the challenge of poor interpretability in machine learning.
- Using 11 indicators to measure momentum from four different aspects, the composition of momentum is analyzed comprehensively from multiple perspectives.

9.2 Weaknesses

- Model II exhibits relatively poor generalization capabilities for non-tennis matches.

References

[1] Jiawei Xu and Zhiyuan Zhang. Exploring factors affecting athlete psychology in tennis matches and countermeasures. *Innovation in Ice and Snow Sports Research*, 1(164-166), 2023.

[2] Yin Xue. A study on the key points grasping and the relationship between success and failure of outstanding male tennis players worldwide. Master's thesis, Chengdu Sports College, 2018.

[3] Zhencong Liang. A brief discussion on tennis serving tactics. *Journal of Chifeng University (Natural Science Edition)*, 25(06):99–100, 2009.

[4] Ting Jiang and Qing Li. A study on the factors influencing the outcome of men's singles matches in professional tennis tournaments: Based on the four grand slam events from 2014 to 2018. *China Sport Science and Technology*, 57(07):62–68, 2021.

[5] K. M. A. S. Al-Harbi. Application of the ahp in project management. *International Journal of Project Management*, 19(1):19–27, 2001.

[6] Y. Zhu, D. Tian, and F. Yan. Effectiveness of entropy weight method in decision-making. *Mathematical Problems in Engineering*, 2020:1–5, 2020.

[7] Zhenmin Wang and Lianchun Zhou. The role of psychological control in the tennis serving process. *Liaoning Sports Science and Technology*, 1(04):31–37, 1998.

[8] Ali Baghban Parsa, Alireza Movahedi, Hamed Taghipour, et al. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136:105405, 2020.

[9] Markus Junker, Roland Hoch, and Andreas Dengel. On the evaluation of document analysis components by recall, precision, and accuracy. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*, pages 713–716. IEEE, 1999.

The Secret of Momentum

Memorandum

To: Coaches

From: Team # 2420216

Subject: Research on Momentum and Recommendations for Player Preparation

Date: February 5, 2024

Dear Sir or Madam:

People often attribute incredible swings during a match to "momentum." We reckon that everyone involved, players and coaches alike, would be keen on unraveling the mysteries behind this "momentum."

I. What is "momentum" ?

"Momentum" is a relatively abstract concept, specifically in tennis matches, where we define it as the player's performance over the last few points and the additional impact of this performance on the likelihood of winning the next point.

II. The Secret of Momentum — Our Research Results

1. We establish a Multi-angle Momentum Evaluation Model to assess players' momentum based on various factors. By calculating the Relative Momentum (RM) between a player and their opponent during a specific time period, we quantitatively determine who is performing better on the court and to what extent.

2. Through a qualitative combined with quantitative analysis, we determine that "momentum" plays a significant role in the game and can serve as a crucial factor in assessing the course of a match. The magnitude of "momentum" shows a high correlation with swings in the game and the likelihood of a player winning.

3. We construct the Relative Advantage Swing (RAS) as an indicator describing changes in the course of a match. Additionally, we develop the XGBoost-SHAP Interpretable Swings Prediction Model. Our findings reveal that overall point differentials, recent scores, the server, and the number of successful serves are the four most correlated factors influencing the shift in the match dynamics from one player to favoring another. The corresponding SHAP values are 0.32, 0.30, 0.17, and 0.16, respectively.

4. Our model shows strong generalization across tennis matches of different genders and surfaces, but might not be as effective for players with distinctive traits or in matches where one side significantly outweighs the other in strength.

III. Recommendations for coaching your players

Next, please allow us to offer recommendations regarding the role of "momentum," how to prepare players, and how to address events that can impact the course of tennis matches.

1. Emphasize daily training in serving and receiving serves. Serving-related metrics are crucial factors influencing momentum and the course of the game. Strengthening serving and receiving skills can help your players gain a greater advantage.

2. Develop tactical strategies based on the situation of your players and opponents. For instance, if the opponent is stronger and has consistently performed well in the earlier rounds, encourage your players to adopt a relatively aggressive approach during the match, initiating proactive attacks to gain an advantage. If your players have superior skills but show noticeable momentum fluctuations in the early rounds, advise them to employ a relatively steady strategy during the game, seeking stability for eventual victory.

3. Maintain a good status. The "momentum" exhibits a certain cyclicality, and sustaining a good condition is beneficial in handling various match situations. When players experience consecutive point losses in a game, they should believe they still have a chance; and when in an advantageous position, be prepared for the possibility of the opponent launching a counterattack. This encompasses physical health, outstanding performance, superior fitness, and a stable mental state.

We sincerely hope that our model and research findings prove helpful to you. Wishing you and your players great success in your matches!

Yours sincerely,
Team # 2420216