Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2425454

Changing Tides: Momentum Swings in Tennis

Summary

Situations change rapidly in tennis, and the incredible swings are often attributed to **momentum**. We analyze momentum through Wimbledon 2023 men's matches data.

For Task 1, we develop a **Leverage Quantitative Momentum Capture Model**. Firstly, we consider the effect of short-term motivation and confidence boosting on serve percentage, and utilize the **Monte Carlo** algorithm to compute the winning probability(match/set/game) at each point. Then we define **Leverage** as how much winning or losing a ball changes the winning probability and the momentum as **exponential moving average** of leverage. Finally, the Wimbledon final match flow is visualized and the results are shown.

For Task 2, we use a statistical test to verify that "momentum" does play role in the match. Through the **run test**, we found that at a 90% confidence level, the proportion of non-random game scores in the given dataset accounts for 45.11% of the total, which indicates the swings in the game are not random. Furthermore, through the **contingency table independence test**, we determine, at a 90% confidence level, that swings in play and runs of success by one player are correlated to the momentum.

For Task 3, we establish the **Grid-Search Random Forest Swing Predictor Model** to forecast short-term swings in momentum.We select 43 indicators as inputs for the model to predict **momentum transition scenarios** (no transition of advantage, the advantage tilting towards oneself, the advantage transitioning to the opponent) in the next s points. The model achieved a **94**$\%$ successful classification rate on the training set, demonstrating good classification performance. Utilizing GSRF, we can also identify that the factor most related to match fluctuations is the **opponent's total score**.

It is necessary to consider he top 15 factors in the GSRF model in terms of feature importance, so the four factors of dimensionality reduction into **Expenditure, Expenditure, Running Distance, and Game Situation Factor** are utilized by **factor analysis**. We suggest players focus on improving physical fitness, enhancing mental resilience, and paying special attention to break and net point scoring techniques.

For Task 4 , by modifying the Monte Carlo simulation winning rule to **three sets to two wins**, we apply the model in the 2023 **Wimbledon women's singles final**, which verifies that the model is generalizable. It is worth noting that the most related factor of swing in women's tennis is different from that in men's singles, it is precise **the total running distance**, which may be be related to the physical strength difference between men and women.

We conduct **sensitivity analysis** on the parameters of the short-term motivation effect in the Leverage Quantitative Momentum Capture Model and the parameters in the GSRF Swing Predictor Model, which are low in sensitivity and have high robustness.

Finally, a one-page memo with suggestions to tennis coaches is also produced to help coaches instruct players to control the pace of play.

Keywords: Monte Carlo Simulation, Runs Test, GSRF, Factor Analysis

Contents

1 Introduction

1.1 Background

Tennis matches are exciting and ever-changing. Players who seem to be dominating experience incredible swings from time to time. For example, a player may be in control of the match with a series of points, but at one rally, the opponent suddenly breaks and the situation is reversed. We often attribute this dramatic turn of events to "momentum".

"Momentum" heightens emotional intensity in the match, creating unpredictability and anticipation, making each moment charged and intense.

Figure 1: The fast-changing, heart-stopping swings of a tennis match.

1.2 Restatement and Analysis of the Problem

We need to complete the following tasks based on the data given in the question and additional data from ourselves:

- **Task 1 : Build a model to capture the flow of play as points occur.**

  The model is required to be able to qualitatively and quantitatively describe performance of players at a given time and provide a visualization. In particular, the model should take the higher probability of the server winning into account.

- **Task 2 : Assess whether swings in play and runs of success by one player during a match are random.**

- **Task 3 : Look for indicators that identify the point in time when the game situation is reversed.**
  - Develop a model to predict swings in a match. Find the indicator with the highest correlation.
  - Based on momentum swings in past games, we need to give players advice on how to deal with other opponents.

- **Task 4 : Evaluate the generalisability of the model we have developed to other competitions.**

- **Task 5 : Write a memo summarising the results and giving targeted advice to coaches and players.**

After some analysis, we found that there is a close relationship between the tasks: In Task 1 and Task 2, we need to build a model describing the situation of the game (i.e. the relative performance of two players) and explore the relationship between "momentum" and changes in the game situation, respectively. The results of these two tasks are the basis of task 3. In task 3, we need to further build a model that can predict the situation fluctuations in the competition and find out the key factors that affect the competition. We also need to start to consider the applicability of our model. Can it provide competition suggestions for players to achieve better results? The fourth task is to consider the extensibility of our model. Are we still valuable in other competitions? Finally, in the memo, we need to focus on our model and its value.

In summary, we should effectively build a model that can predict the fluctuations of the competition situation, provide practical advice to players and coaches, and be extensible across different competitions.

1.3 Overview of Our Work

On the basis of the above analysis, we have carried out our work. Our working framework is shown in Figure 2.

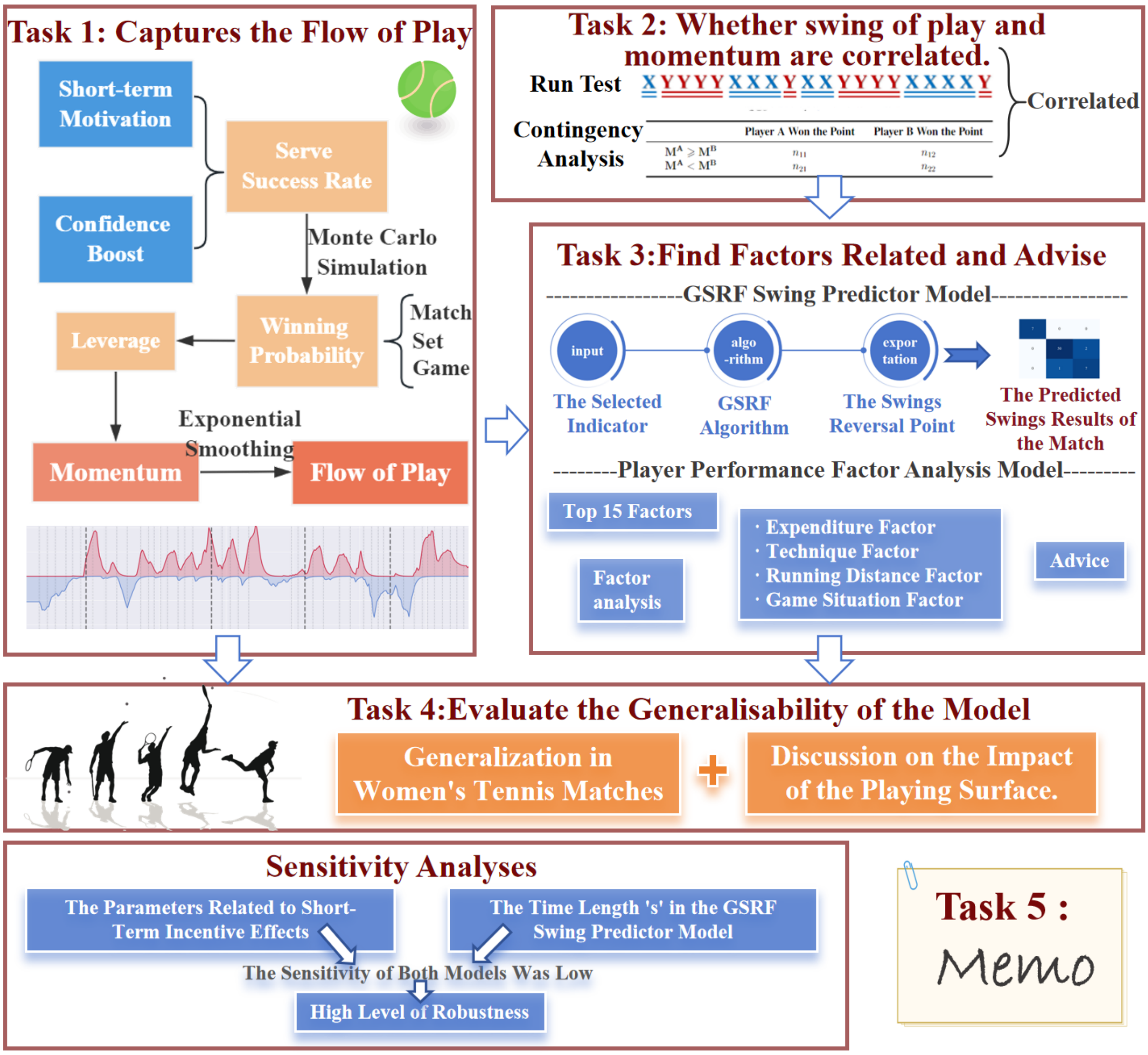


Figure 2: Our Work Overview Schematic Diagram

The detailed modeling process and model details are presented in the following article.

2 Assumptions and Justification

To simplify the problem and make it convenient for us to simulate real-life conditions, we make the following basic assumptions, each of which is properly justified.

- **When a player wins a point, he experiences a short-term "motivational" effect that increases the probability of winning the next point.** The motivation effect can lead to increased self-confidence on a psychological level as well as increased physical conditioning on a physiological level.
- **Lifting effects are cumulative**. A string of successes may have a positive psychological and emotional effect on a player, so that a player winning more points in a row will result in a greater probability of winning the next point on serve.

3 List of Notation

Symbol	Meaning
p, q	Probability of Winning
m_i	Increase in the Rate of Serve points
T_{AB}	The Cumulative Total Score
L_i	Leverage of the ith Point
G	The leverage Gained
M	Momentum
λ	Tuning Parameter

Noted: Symbols not specified in the table should be referred to their first occurrence.

4 Data Pre-processing

- **Outlier Handling**

  No clear outliers were observed in the data. However, in a match between Frances Tiafoe and Grigor Dimitrov during the third round, a notable occurrence occurred with a 24-hour time gap between two points at a set score of 2:0, as shown in Table 1.

Table 1: Suspected Outliers

match_id	playerA	playerB	elapsed_time
2023-wimbledon-1303	Frances Tiafoe	Grigor Dimitrov	0:53:55
2023-wimbledon-1303	Frances Tiafoe	Grigor Dimitrov	**0:54:22**
2023-wimbledon-1303	Frances Tiafoe	Grigor Dimitrov	**24:56:34**
2023-wimbledon-1303	Frances Tiafoe	Grigor Dimitrov	24:57:00

Upon consulting relevant information, we found the following explanation:

"...... Saturday's game was suspended due to inclement weather, so it was postponed to a day later to continue on Sunday.. ...... "

Therefore, the data is correct.

- **Missing Value Handling**

  In the dataset given in the question, the quantities $serve_width$, $serve_depth$, and $return_depth$ have missing values, for which we use the **nearest-neighbour** filling method to fill them.

- **Tag Encoding of String Type Data**

  There are many categorical variables in the data given in the title, presented as strings. In order to make it easier for the model to receive this type of data, we preprocess this type of data by means of label coding, using virtual coding for the last three columns of $serve_{width}$, $serve_{depth}$, and $return_{depth}$ in the data of the tennis match.

5 Task 1 : Captures the Flow of Play

5.1 Specifying Conceptual Significance

Since a relatively large number of concepts are introduced in our model, we first identify and clarify the meaning of the introduced concepts here.

- **Short-term Motivation** refers to a brief period of above-average performance by a player following a motivating event.
- **Confidence Boost** is when a player, drawing upon their self-awareness developed through multiple games in their professional career and an assessment of the current situation, generates positive expectations regarding the outcome of the match.
- **Serve Success Rate** is probability of scoring in a service game.
- **Winning Probability** is the likelihood of the player winning the game/set/match.
- **Leverage** measures how a ball's outcome impacts the margin of victory.
- **Momentum** aims to describe which player is in control at any given point in the match—based on who is currently winning more points and who is prevailing in crucial (high-leverage) moments. It is defined as a player's exponentially weighted leverage average.

  At each point in the match, both competing players will have their own momentum values.
- We define the fluctuations in momentum as the **Flow of Play**.

The relationship between the above concepts, i.e., the framework of the model, is shown schematically below, and in the following we show the model building process in full.

5.2 Leverage Quantitative Momentum Capture Model

5.2.1 Impact of Short-term Motivation and Confidence Boost on Serve Success Rate

Short-term Motivation refers to a brief period of above-average performance by a player following a motivating event (e.g., a breaking serve) during a match, a phenomenon known in sports as "hot hands"[1]. We developed a probabilistic model that quantifies Short-term Motivation as an increase in serve success rate.

The following two assumptions are necessary:

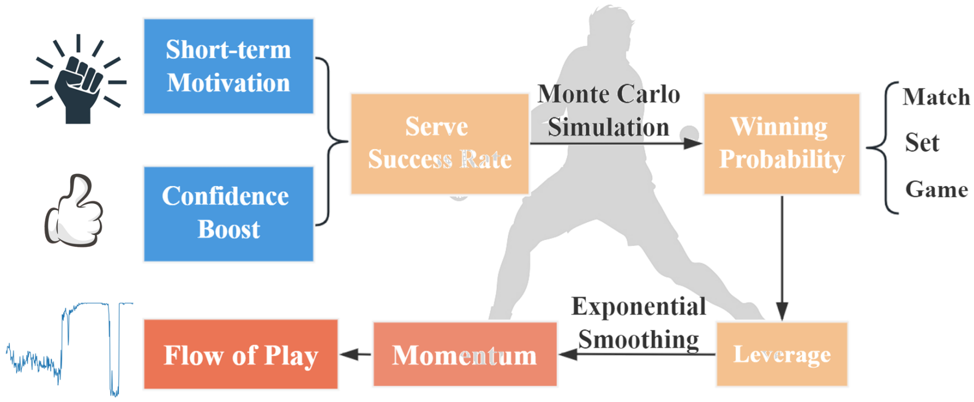


Figure 3: Relationships Between Concepts, i.e., Schematic of the Framework of the Model

- When a player wins a point, he experiences a short-term "motivational" effect that increases the probability of winning the next point.
- Lifting effects are cumulative. A string of successes may have a positive psychological and emotional effect on a player, so that a player winning more points in a row will result in a greater probability of winning the next point on serve.

Here i denotes the number of consecutive points scored by a player A or his opponent B; and m_i denotes the increase in the rate of serving points on a player's next serve after winning $i$points consecutively due to the short-term motivation gained. Then, the probability of winning the next serve under short-term motivation p_s versus q_s is calculated by equations (1) .

$$\begin{aligned} p_s &= p_f(1 + m_i^A); \\ q_s &= q_f(1 + m_i^B), \end{aligned} \tag{1}$$

where p_f and q_f are the scoring probabilities before player A and player B are motivated, respectively.

The short-term motivation that a player receives for scoring a game n times in a row is referred to as n-order short-term motivation. Using the dataset given in the title, we statistically obtained a plot of the total number of short-term motivations occurring in a match versus order, and the results are shown in Figure 4. It is clear to see that short-term motivations of order four and above rarely occur in the dataset used in this study. Therefore, only short-term motivations of order three and below are considered in this paper.

Short-term motivation focuses on short-term status enhancement during matches and ignores sources of motivation on a longer time scale in the player's lifetime, so we continue to introduce a **confidence-boosting factor**.

A new tuning parameter λ is introduced, which represents the effect of the confidence boosting factor on the success rate of the next serve [2], the probabilities after the effect are expressed as p_s and q_s – see equation (2)

$$\begin{aligned} p_c &= \frac{T_{AB}}{\lambda} p_t + (1 - \frac{T_{AB}}{\lambda}) p_0; \\ q_c &= \frac{T_{AB}}{\lambda} q_t + (1 - \frac{T_{AB}}{\lambda}) q_0, \end{aligned} \tag{2}$$

where T_{AB} denotes the cumulative total score of the opposing players since the start of the match, and p_t and q_t denote the probability of scoring for player A and player B, respectively,

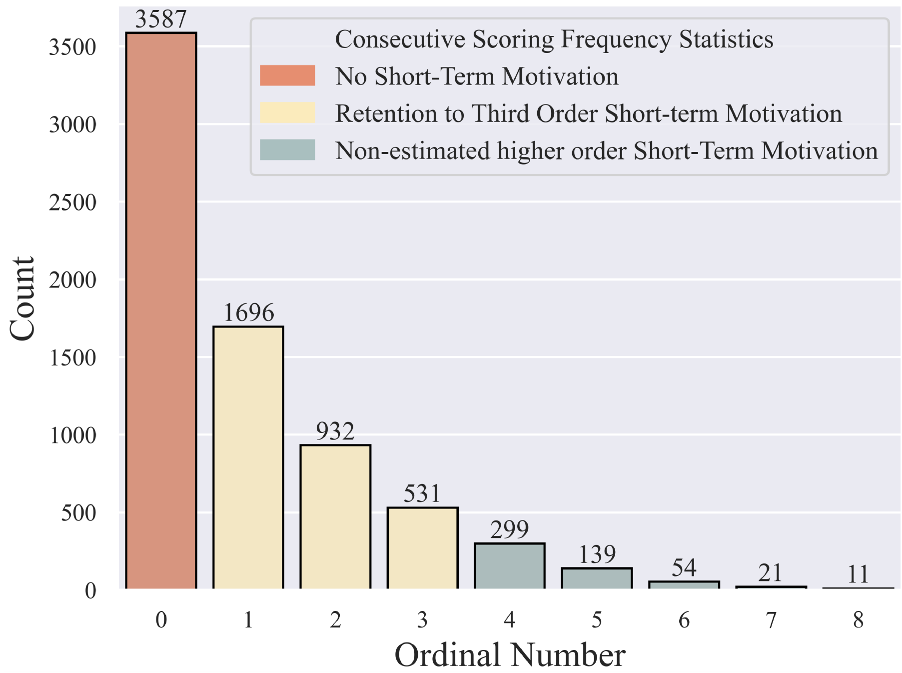


Figure 4: Histogram of the Total Count of Short-Term Incentives Occurring in Competitions as a Function of Order

since the start of the match up to the present. p_0 and q_0 denote the historical data on the success rate of the opposing teams' serves on the grass court surface prior to the current match (source: https://www.infosys.com/)

Eq. (2) describes the phenomenon that a player's historical results always have the greatest impact on player status at the start of a match, and then diminish as the match progresses - the performance of the current match begins to influence player play more.

We then performed a linear interpolation of the obtained probabilities. The interpolation coefficient is the cumulative total score of the current pair divided by the coefficient λ (i.e. $\frac{T_{AB}}{\lambda}$).

The short-term motivation is fluctuating as the match progresses, while the confidence boost takes into account the player's self-perception established over the course of his career as well as the current situation in the match, which is the "undertone" of the motivated state, therefore, we use p_c and q_c in Eq. (2) instead of p_f and q_f in Eq. (1), we can get the success rate of the next serve after considering the short-term motivation and confidence improvement:

$$p = [\frac{T_{AB}}{\lambda} \cdot p_c + (1 - \frac{T_{AB}}{\lambda})p_t](1 + m_i^A);$$
$$q = [\frac{T_{AB}}{\lambda} \cdot q_c + (1 - \frac{T_{AB}}{\lambda})q_t](1 + m_i^B).$$

The above equation shows the effect of short-term motivation and confidence enhancement on serve percentage.

In tennis matches, players typically have an advantage when serving[3]. According to the data we have collected, the serving success rate of professional players exceeds 50%. Based on this understanding, **players gain confidence in serving games**. Due to this "confidence boost," we have incorporated the player's performance in serving games into consideration in our model.

5.2.2 Monte Carlo Simulation for Winning Probability

As a sports competition, there is a great deal of uncertainty in tennis, for which Monte Carlo algorithms can provide probabilistic estimates of the outcome.

Based on the serve success rate obtained in the previous section, we use the Monte Carlo algorithm to simulate each ball, and ultimately obtain the winners and losers of each game, set and even the whole match. Then in this process, the randomness and diversity of the simulation are ensured by generating a large number of random number seeds.

Here is the pseudo-code for the algorithm we used:

Algorithm 1: Monte Carlo Simulation for Winning Probability

Input: $phist$,$data$
Output: win_p

begin
    for i ***in*** *scores* **do**
        for i *1* ***to*** n **do**
            repeat
                Calculate p_{cur} based on previous data
                Calculate scoring probability based on the formula.3
                Update players' score/game/set based on tennis rule
                Record n_{win} of game/set/match
            until *one side wins the match*;
        end
        $win\_p[i] \leftarrow$ n_{win}/n
    end
end

We use the above algorithm to perform $N = 1000$ simulations of the remaining fixtures after each ball and determine the final outcome of that game. If the number of wins of player $K(K = A, B)$ in the result is n, then we say that the winning rate of player K after that ball is $\frac{n}{N} \times 100\%$.

The graph below illustrates the real-time fluctuations in match-winning probabilities, simulated through the Monte Carlo algorithm, for Djokovic (highlighted in red) and Alcarez (highlighted in green) during the 2023 Wimbledon final.

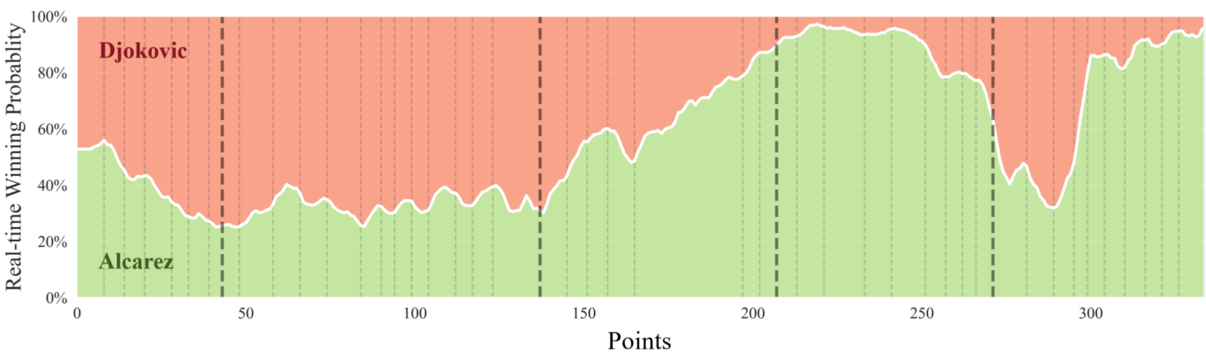


Figure 5: Wimbledon Final Real-Time Predicted Win Probability Swings Chart. The thick gray dashed lines represent set divisions, while the thin gray dashed lines represent game divisions.

It can be seen that the situation is very anxious, and the match winning probablity in the middle of the game fluctuates greatly throughout the match. The first set was easily won by Djokovic, but the second was tense . . . . . . This is consistent with the facts. Thus, the reliability of the algorithm is verified.

5.2.3 Leveraging Every Point with Counterfactual Framework

Next, we will employ a counterfactual framework[4] to delve into the analysis of the impact of each point in a tennis match. By defining 'leverage' as a measure of the magnitude of change in winning probability resulting from winning or losing a particular point, we will quantify the potential influence of each point on the outcome of each game/set, and even the entire match.

The following example shows the exact steps of our leverage quantisation process.

- **Current Scenario:** Player A is down 0-15 in the current game, his real-time winning probablity (abbreviated as **RTWP** in the latter.) is 69%.
- **Define Realistic Scenario:** If player A wins this point, the score becomes 30-30 and his RTWP becomes 71%.
- **Imagine the Opposite Scenario:** if player A loses this point, the score becomes 15-40 and his RTWP becomes 65%.
- **Calculating Leverage:** By comparing the change in A's RTWP in these two scenarios, the Leverage for this point is calculated to be 6%.

Note: The above data is fictional. The above process is shown schematically below:

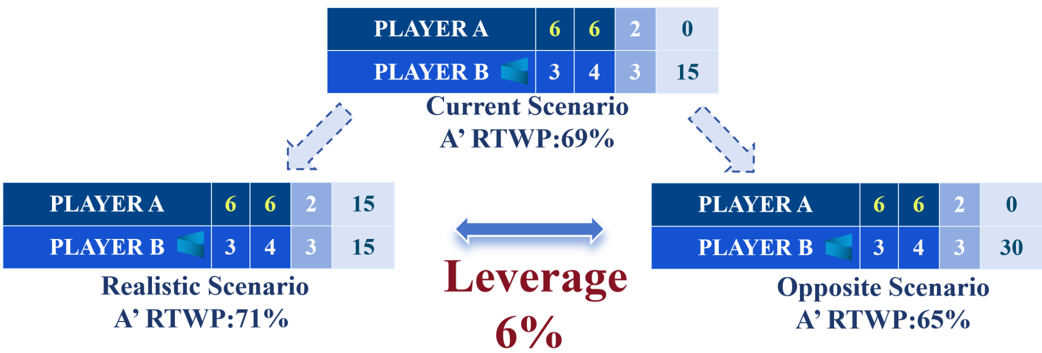


Figure 6: Leveraged Quantitative Schematic

We set L_i to represent the leverage of the i^{th} point. The positive or negative sign does not indicate the magnitude but only signifies **which player gained the point**. Throughout the subsequent analysis, we consistently calculate the leverage with respect to Player A. If $L_i > 0$, it means that player A won the ith point; if $L_i < 0$, it means that player B won the ith point; if $L_i = 0$, it means that the ith point did not result in a change in winning percentage.

If player$K$ wins the ith point, the player$K$ is called to have gained leverage, and vice versa, the player is called to have lost leverage. The leverage gained is treated as a separate variable and denoted by the symbol G. To wit:

$$G_i^A = \begin{cases} L_i, & L_i \geqslant 0 \\ 0, & L_i < 0 \end{cases}, G_i^B = \begin{cases} L_i, & L_i \leqslant 0 \\ 0, & L_i > 0 \end{cases}$$

5.2.4 Defining Momentum

We define momentum(M) as an exponentially weighted moving average of the leverage gained by a player[5].

$$M_i^K = \frac{G_i^K + (1-\alpha)G_{i-1}^K + (1-\alpha)^2 G_{i-2}^K + \cdots + (1-\alpha)^i G_0^K}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^t},$$

where $[G_0, G_1, \ldots, G_i]$ is the leverage of i recent points, and the points are listed in chronological order.α is the smoothing coefficient ,and $\alpha \in (0,1)$. Due to the fact that the impact of previous goals diminishes on the player's current impact as the game progresses, the importance of a certain points decreases exponentially as the game progresses, The leverage of the most recent goal has the greatest impact on momentum, and the sooner the point is scored, the less it impact on momentum.

Finally, by focusing on the fluctuations in momentum, we can capture "the flow of play".

5.3 Visualization of the Flow of Play

Utilizing our model, we have described the match flow of the Wimbledon Final in 2023. The parameter α is set to 0.33, by testing the leverage defined by the handicap winning percentage, it better reflects the fluctuation of the game. The visualization is presented as follows:

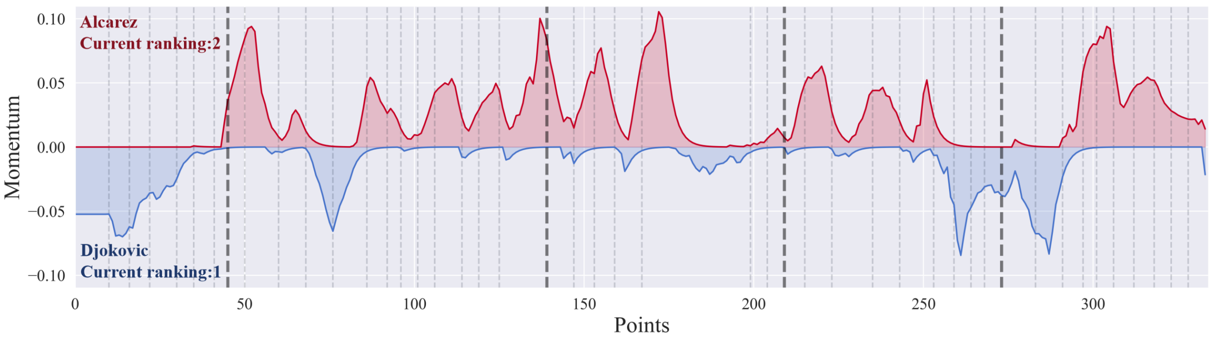


Figure 7: The Visualization of Match Flow. The thick gray dashed lines represent set divisions, while the thin gray dashed lines represent game divisions.The same applies to the subsequent figure.

Our results perfectly describe the flow of the match. We can correlate the fluctuations in momentum in the graph with the remarkable moments in the game:

Riding a dominating momentum after crushing the opponent in the first set, Djokovic had the opportunity to claim the second set. However, Alcaraz overcame the initial struggle and secured a crucial break in the second set, leading 2-1; Djokovic swiftly responded with a break of his own, but Alcaraz seized the opportunity in a tense tiebreak, winning the set with a brilliant backhand winner. In the third set, Djokovic struggled with a low momentum, while Alcaraz continued his high momentum, leading 2-1. In the fourth set, Djokovic broke Alcaraz twice, leveling the match.

Then, in the decisive fifth set, Djokovic missed a golden opportunity while leading, allowing Alcaraz to break and take a 2-1 lead. With Djokovic's momentum dropping, he once again violated rules by smashing his racket against the net post, subsequently trailing 3-1. Alcaraz remained resolute, maintaining his high momentum. When Djokovic hit a forehand into the net, Alcaraz secured his extraordinary victory at 3-1."

It can be found that the momentum we define shows the race flow well.

6 Task 2 : Whether the Swings of Play and Momentum are correlated

A tennis coach argues that swings in play and runs of success in a match are random and not related to momentum. In response to this view, we plan to determine whether the distribution of scores in a match is completely random by means of a Runs Test.

6.1 Run Test for Assessing Randomness in Match Scores Sequence

The runs test is commonly employed to examine whether a binary sample with two potential outcomes, such as X and Y, originates from a binomial distribution.

In the runs test, a sample is drawn from a population containing X and Y, assuming n X's and m Y's are drawn. Subsequently, X and Y are arranged in order of sampling to form a sequence. In this sequence, consecutive occurrences of X or Y are referred to as a run. The total number of runs is denoted as the run count r.

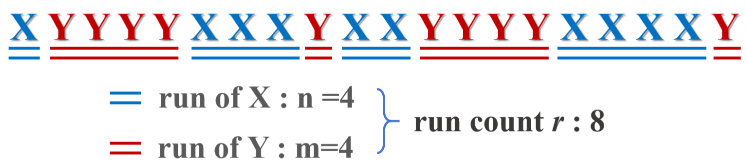


Figure 8: Runs Example Chart

In the XY sequence shown in Figure 8, there are $n = 4$ runs of X and $m = 4$ runs of Y, with a total of $r = 8$ runs.

- The null hypothesis H_0: The arrangement of X and Y is random, with the probability of drawing X each time as P, and the probability of drawing Y as $1P$.
- The alternative hypothesis H_1: The arrangement of X and Y is non-random, indicating no regularity in the probability of drawing X or Y.

If H_0 holds, based on the parity of r, there are different scenarios:

$$P(r = 2k) = \frac{2\begin{pmatrix} n-1 \\ \mathrm{k}-1 \end{pmatrix}\begin{pmatrix} m{-}1 \\ \mathrm{k}-1 \end{pmatrix}}{\begin{pmatrix} n+m \\ n \end{pmatrix}};$$

$$P(r = 2k+1) = \frac{\begin{pmatrix} m-1 \\ k-1 \end{pmatrix} + \begin{pmatrix} m-1 \\ k-1 \end{pmatrix} + \begin{pmatrix} n-1 \\ k \end{pmatrix}\begin{pmatrix} n-1 \\ k \end{pmatrix}}{\begin{pmatrix} n+m \\ n \end{pmatrix}}$$

The mathematical expectation of r is $E(r) = \dfrac{2mn}{m+n} + 1$, the variance of r is $V(r) =$

$\frac{2\mathrm{mn}(2\mathrm{mn}-\mathrm{m}-\mathrm{n})}{(\mathrm{m}+\mathrm{n})^2(\mathrm{m}+\mathrm{n}-1)}$, and the statistic$Z$ constructed from the number of trips follows a normal distribution, that is, $Z=\frac{r-E(r)}{\sqrt{V(R)}}\sim N(0,1)$.

We recorded the outcome of each point within each game throughout the entire match. If player A won the point, it was recorded as 1, and 0 otherwise. This resulted in sequences representing the distribution of scores within each game, and runs were calculated based on this. Subsequently, a runs test was conducted for each game. The results indicated that within a 90% confidence interval, the percentage of games rejecting the null hypothesis accounted for 45.11% of the total number of games.

This suggests that scoring in a game is not completely randomized.

6.2 Scores-Momentum Contingency Analysis

"In the previous section, we employed the runs test to investigate whether the scoring sequence of opposing sides in tennis matches exhibits randomness. The result was negative, indicating that the match dynamics are not random. Subsequently, to explore whether the fluctuations in the game are related to momentum, we conducted a contingency table analysis on the relationship between the score of each point and the momentum of the opposing sides.

We first counted the winners and losers of each point as well as the relationship between the momentum of playerA and playerB in terms of magnitude when competing for this point, as shown in the table below:

Table 2: Score vs. Momentum Table

	Player A Won the Point	**Player B Won the Point**
$\mathbf{M^A \geqslant M^B}$	n_{11}	n_{12}
$\mathbf{M^A < M^B}$	n_{21}	n_{22}

In the table, n_{ij}represents the frequency with which events in row i and column j occur simultaneously, where i represents the comparison of player A's momentum with player B's momentum, and $j$represents the winner of the match point. Let the original hypothesis be H_0, indicating that there is no correlation between score and momentum, and the alternative hypothesis be H_1, indicating that there is a correlation between score and momentum.

Assuming that there is no correlation between score and momentum, for each cell in Table 1 calculate the expected frequency E.

Since the total sample with all expected frequencies is greater than 5, the conditions for using the chi-square test are met. Therefore, we used the chi-square test to assess the difference between the observed and expected frequencies. Taking Wimbledon as an example, the calculated chi-square statistic $X=5.566$, indicating that the observed frequencies are significantly different from the expected frequencies; the significance level $P=0.018$, indicating the rejection of the original hypothesis H_0 at the 90% confidence level that the relationship between the winners and losers of each point and the magnitude of playerA's and playerB's momentum when competing for that point is correlated.

So **"momentum" does play an important role in the game**, and swings in play and runs of success are **not random**.

7 Task 3 : Identify Indicators Signaling a Change in Flow

7.1 GSRF Swing Predictor Model

First, we need to develop a model to predict swings in a match and look for the factors most associated with swings.So we developed the GSRF Swing Predictor Model.

Our model needs to be able to predict whether there will be swings at s points in the future, and determine its direction accurately - whether it will be a wave in favor of the player or a swing in favor of his opponent. To do this, we first introduce the concept of '**swing reversal points**', which have two different types.

7.1.1 The Swings Reversal Point

We use the '**scipy.signal.find_peaks**' function in Python to locate the peaks in the curves representing the momentum changes over time for both the player and the opponent. Between the peaks of the player and opponent, we identify the point where the momentum value is closest to 0 as the "swing" reversal point.

If the peak of the opponent comes before the player's peak, indicating the advantage is leaning towards the player, the reversal point between these two peaks is marked as a transition point favoring the player, labeled as "1". Conversely, if the player's peak comes before the opponent's peak, indicating the advantage is shifting towards the opponent, the point between these two peaks is marked as a reversal point favoring the opponent, labeled as "-1". Points that are not transition points are labeled as "0". The schematic representation of this process is illustrated in Figure 9.

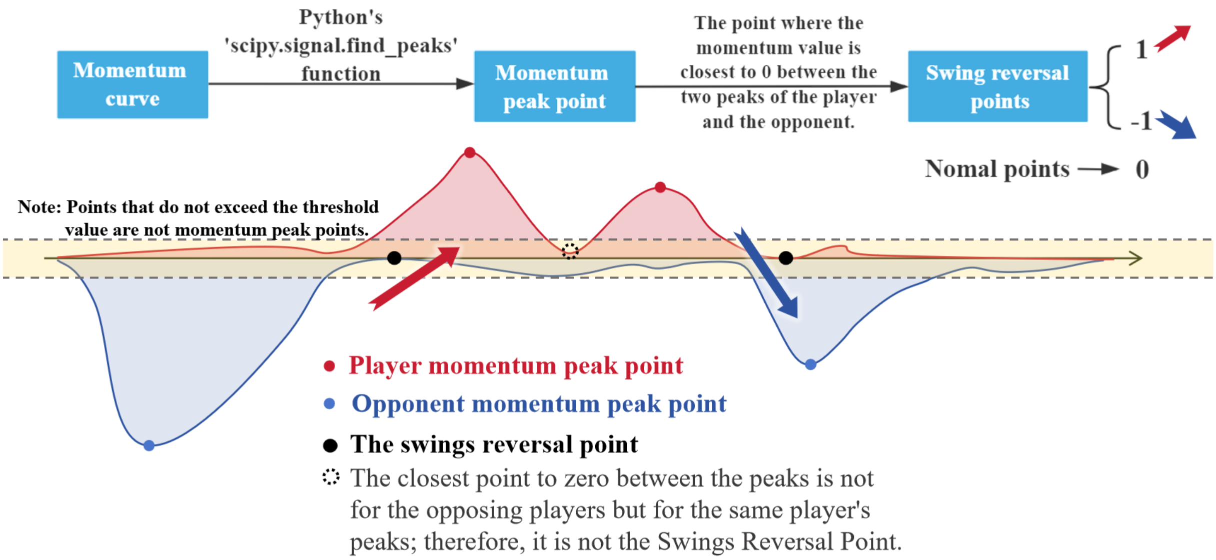


Figure 9: Rurning Swing Points Selection Process and Schematic Diagram

We verified the method using the 2023 Wimbledon final data, and the results are good, as shown in the figure below:

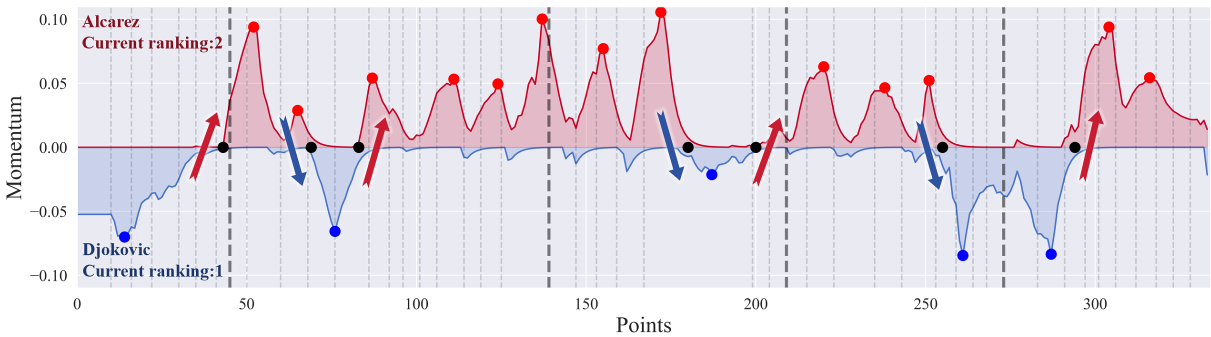


Figure 10: GSRF Algorithm Flow Diagram

To predict swings in future matches, we take a subset of players' on-court performance metrics and use the grid search Random Forest (GSRF)[6] algorithm to predict which of three scenarios each point in future belongs to.

Next, we will describe the selected indicator data and explain the algorithm flow.

7.1.2 The Selected Indicator

Through the given tennis dataset, we have selected two categories of indicators for prediction. One category pertains to individual player metrics, while the other involves shared metrics when the two players compete. These two categories of indicators are listed as follows:

The first category: p1_ace: Whether player 1 hits an "unreturnable" ball during the serve; p1_ace: p1 total aces; p1_break_pt: Break point: player 1 has the opportunity to win a point in opponent's service game; p1_double_fault: Player 1 committed two faults in serving, resulting in a lost point; p1_double_fault: p1 total double faults; p1_net_pt: Player 1 approached the net in this round; p1_net_pt_won: Player 1 won the point when approaching the net in this round; p1_score: Player 1's score in this game; p1_sets: Number of sets won by player 1; p1_unf_err: Player 1 made an unforced error in this round; p1_winner: Whether player 1 hit an "unreturnable" ball,**p1_games**: p1 won games; **p1_points_won**: p1 total points won; **p1_distance_run**: p1 distance run in this point; **brp1**: p1 break point conversion rate; **p1_distance_run_total**: p1 total distance run; **ntp1**: p1 net points won percentage.

Player 2 has exactly the same indicator type as Player 1, so it is not repeated.

The second category: **elapsed_time**: Duration of the point ; **speed_mph**: Serve speed ; **rally_count**: Cumulative number of hits ; rally_count: Total hits in a rally ; return_depth: Depth of returning, indicating whether the return lands near the baseline, midcourt, or frontcourt ; serve_depth: Depth of the serve, indicating whether the serve is placed in the front, mid, or backcourt of the tennis court ; serve_no: Which serve the player is in, the first or second ; serve_width: Direction of the serve ; server: Player serving.

Some of the indicators are original metrics from the dataset, while others (in bold) are derived through calculations based on the original metrics.

7.1.3 GSRF Algorithm

The GSRF algorithm consists of the Random Forest and Grid Search algorithm.

The Grid Search algorithm is a parameter optimization technique that iterates through a specified parameter grid. It trains and evaluates the model for each possible combination of parameters, ultimately outputting the best parameter set along with corresponding model performance metrics.

Random Forest, is a machine learning algorithm that trains based on multiple decision trees. During training, it randomly selects a subset of features for each tree. In the classification process, Random Forest aggregates the predicted classification results of each decision tree through averaging or voting, yielding the final result.

Unlike the regular Random Forest algorithm, the GSRF algorithm utilizes the best hyperparameter combinations to train and predict the Random Forest model. This significantly enhances the model's performance, mitigating issues such as overfitting or underfitting. The GSRF algorithm workflow is illustrated in Figure 11.

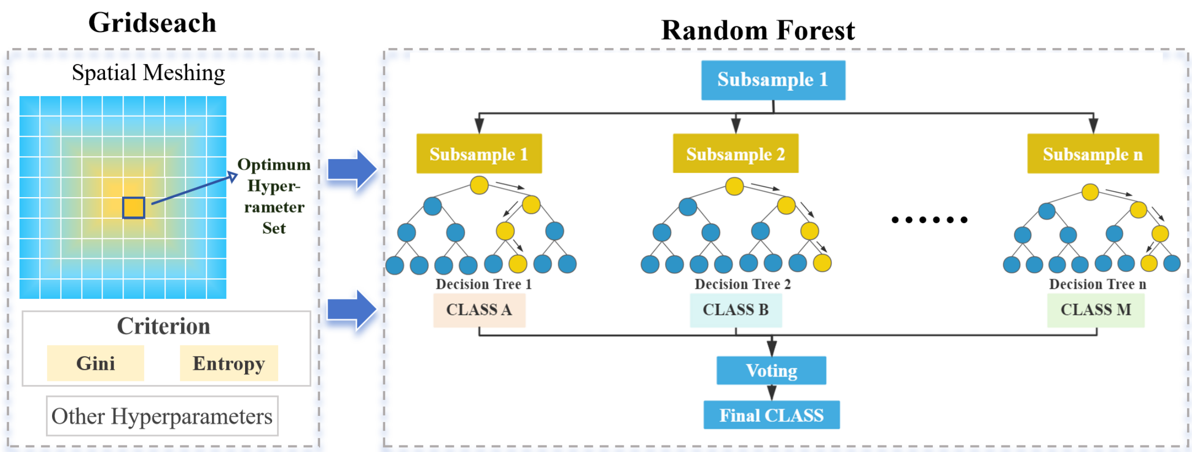


Figure 11: GSRF Algorithm Flow Diagram

7.1.4 The Predicted Swings Results of the Match

We utilize the aforementioned classification indicators as input parameters for the Random Forest model to predict future scenarios.

As an example, for a particular match, we split the dataset into training and testing sets in an 8:2 ratio. We employed the RandomForestRegressor algorithm from the ensemble module in the scikit-learn (sklearn) machine learning library, along with the GridSearchCV algorithm from sklearn. The optimal hyperparameter combination obtained from GridSearchCV is as follows:

Table 3: Table of Optimal Hyperparameter Combinations

Indicator Names	max_depth	min_samples_leaf	min_samples_split	n_estimators
Values	None	1	3	100

The prediction accuracy of the model is up to 94%, and the performance of the test set is also good. The confusion matrix diagram is drawn as follows:

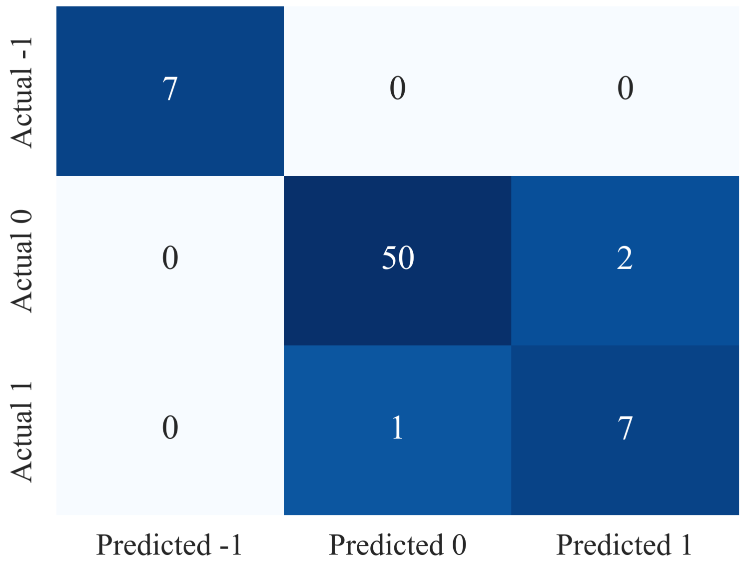


Figure 12: Confusion Matrix

As can be seen from the matrix, most of the samples were correctly categorized, only 3 samples were not correctly categorized, and none of the samples from any of the reversal point types were categorized into another reversal point type.

From the random forest, we can also obtain the importance of various metrics, as shown in the following figure:

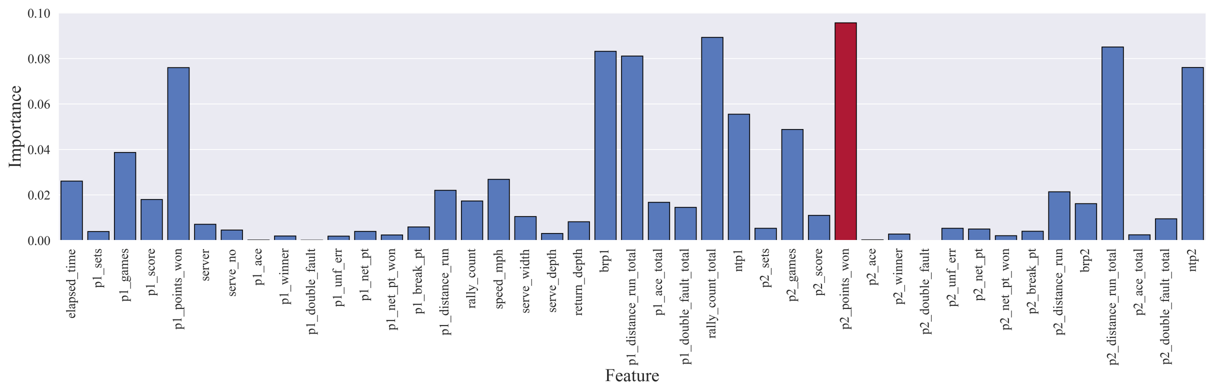


Figure 13: Bar Chart of Metric Importance

As seen in the graph above, the total score of opponents is the most relevant metric for swings in the match.

7.2 Player Performance Factor Analysis Model

The influencing factors obtained from the analysis of the GSRF Swing Predictor Model established in the previous section, which predicts the direction of the game, are complex and diverse. This complexity makes it less convenient for coaches to guide players and understand the momentum of the game. Therefore, we conducted a factor analysis[7] on the indicators that have a significant impact on "swing" to extract independent common factors, enabling players to clearly capture the factors influencing the direction of the game.

We selected the top 15 indicators in terms of feature importance and conducted the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test[8] using SPSS.The test results are as follows:

Table 4: Kaiser-Meyer-Olkin (KMO) and Bartlett's Test

KMO Sampling Adequacy Measure		0.708
Bartlett's Sphericity Test	Approximate Chi-Square	11512.67
	Degrees of Freedom	105
	Significance	0

According to the table, the statistic for Bartlett's sphericity test is 11512.67, and the corresponding probability (P-value) is 0. At a 95% significance level, the null hypothesis should be rejected, indicating a significant difference between the correlation matrix and the identity matrix. Meanwhile, the KMO value is 0.708. According to Kaiser's standard for interpreting KMO, it suggests suitability for factor analysis, indicating factors that influence swing[9].

Next, we utilized SPSS and conducted a principal component analysis for extraction. The scree plot obtained is shown in Figure 14. It was observed that when the number of extractions reached 4, the contribution rate became quite small. Therefore, we decided to extract 4 common factors.

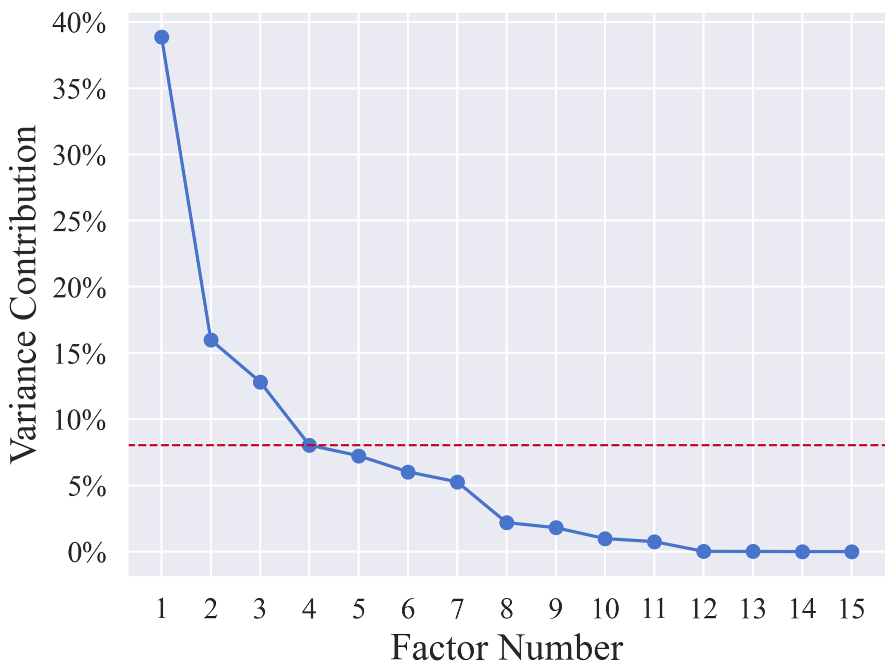


Figure 14: Scree plot of Principal Component Analysis

We utilized the Kaiser normalization and varimax rotation method to obtain the rotated component matrix, as shown below:

Table 5: The Rotated Component Matrix

Indicator \ Extracted Factor	1	2	3	4
p1_points_won	0.992	0.051	-0.035	0.008
rally_count	0.991	0.084	-0.030	0.019
p1_distance_run_total	0.991	0.076	-0.030	0.014
p2_distance_run_total	0.990	0.097	-0.030	0.024
p2_points_won	0.986	0.100	-0.035	0.032
brp1	0.569	0.502	0.000	0.111
brp2	0.072	0.884	0.073	0.232
ntp1	0.368	0.753	-0.007	0.070
ntp2	-0.060	0.732	0.068	0.052
p2_distance_run	-0.061	0.056	0.957	-0.034
p1_distance_run	-0.027	0.067	0.954	-0.061
p2_games	-0.104	0.097	-0.047	0.880
p1_games	0.233	0.286	-0.067	0.808
elapsed_time	-0.048	-0.100	0.185	0.211
speed_mph	-0.021	-0.140	0.067	0.253

Next, we analyze the four factors in the component matrix.

Factor 1 loads heavily on $p1\_points_won$, $p2\_points_won$, $rally\_count$, $p1\_distance_run_total$, and $p2\_distance\_run\_total$. It involves aspects such as shot count and running distance, representing the player's stamina and energy expenditure during the match. We name this factor the **Expenditure Factor**.

Factor 2 primarily involves break point conversion rates ($brp1$, $brp2$) and net point conversion rates ($ntp1$, $ntp2$). As it can reflect the player's technical proficiency during the match, we name it the **Technical Factor**.

Factor 3 loads significantly on $p1\_distance_run$ and $p2\_distance\_run$, primarily reflecting the players' running patterns during the match, likely related to court coverage and movement tactics. We name it the **Running Distance Factor**.

Factor 4 loads heavily on $p1\_games$ and $p2\_games$, mainly focusing on the players' game situations during the match. We name it the **Game Situation Factor**.

The coach can provide guidance to the players by comparing the abilities of their players with those of the next opponent in these four factors.

Utilizing the **Expenditure Factor**, the coach can analyze the stamina gap between players and opponents, enabling targeted training for the players' fitness levels and endurance before the match. The **Technical Factor** reflects a player's technical proficiency, including break points and volleys, aiding the coach in devising more effective strategies. For instance, if a player excels in volleys, encouraging them to employ net strategies more frequently could be beneficial. The **Running Distance Factor** is associated with court coverage and movement tactics. Coaches can optimize players' running strategies based on these factors to enhance control over opponents and reduce self-pressure. The **Game Situation Factor** primarily focuses on players' game scores, assisting in analyzing situations and guiding players to flexibly adjust strategies during matches.

In essence, strengthening mental and physical resilience is crucial against diverse opponents, and focusing on improving break points and volley techniques gives players a substantial edge.

8 Task 4 : Assess the Generalization of Model

8.1 Generalization in women's tennis matches

To assess the generalization of our model, we tested it in the 2023 Wimbledon Women's Singles final. The test results are as follows:

We obtained data on the 2023 Wimbledon women's singles final that is consistent across all metrics with the dataset given in this question(source:https://www.wimbledon.com/index.html). Women's tennis differs from men's tennis matches in that two out of three sets are taken. Therefore, we modified the rules related to the relevant sets for Monte Carlo simulation to solve the real-time winning probability of the match. The calculated momentum for the whole match is shown below:

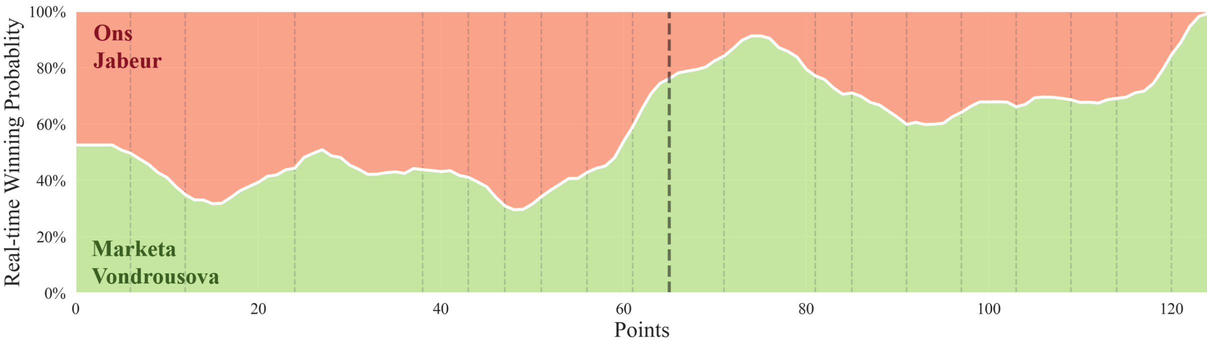


Figure 15: Wimbledon Final Real-Time Predicted Win Probability Swings Chart.The thick gray dashed lines represent set divisions, while the thin gray dashed lines represent game divisions.

For Marketa Vondronsova, the match was a windfall.

Using the model from Task 3 again, the momentum fluctuations for this match are illustrated in the following graph:

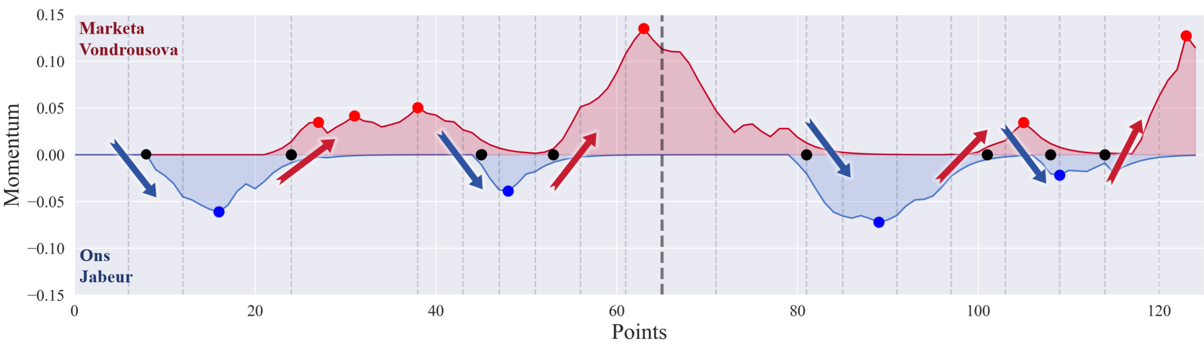


Figure 16: Momentum Fluctuations and Transition Points in Fluctuations.

Similarly, the prediction of short-term NME transition points has the following results on the test set.

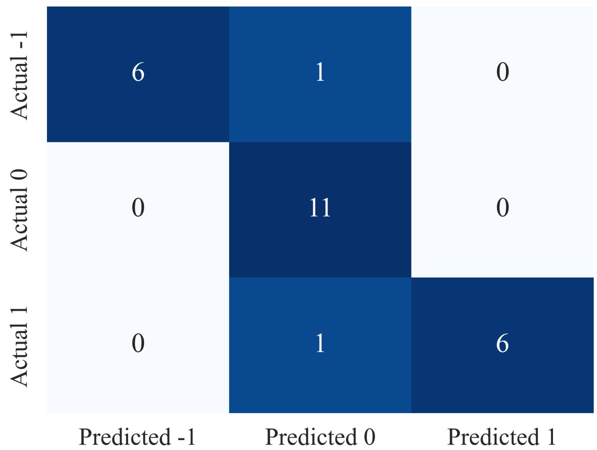


Figure 17: Women's Singles Confusion Matrix

It is not difficult to find that there are fewer samples with wrong scores, indicating that the application of our model to women's tennis matches is feasible. And the important indexes obtained by GSRF model for women's tennis game appeared a little difference with men's tennis game. For example, the most important factor affecting the women's tennis game is the total distance run (as shown in Fig. 18), while the men are the total score of the opponent, which may be related to the difference in physical strength between men and women.

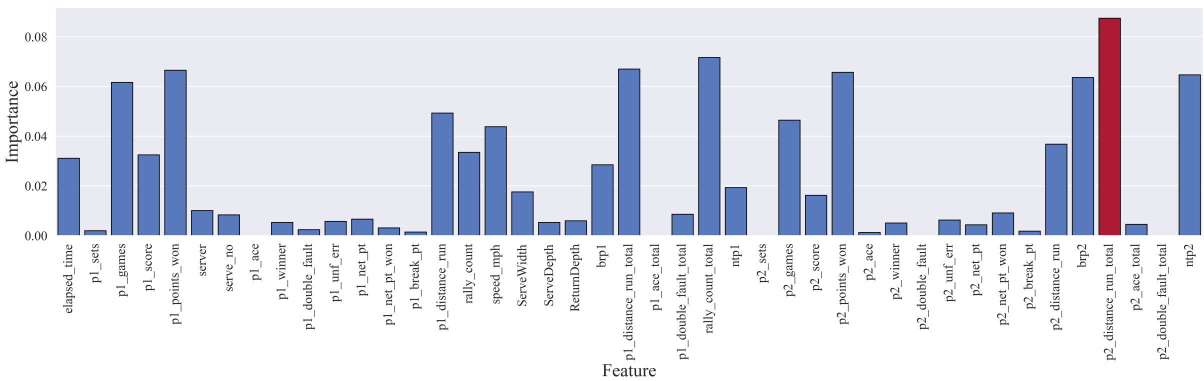


Figure 18: GSRF Algorithm Flow Diagram

8.2 Discussion on the impact of the playing surface.

Research[10] indicates that, with a consistent return speed of 100 km/h and a landing angle of $30°$, on grass, the rebound angle is approximately $36°$ at a speed of about 73 km/h; on hard courts, it's around $41°$ at a speed of about 67 km/h; while on clay courts, the rebound angle is about $44°$ at a speed of about 57 km/h. Different launch angles and rebound heights are also observed.

In summary, playing surface types significantly impact the game. Athletes need to adapt their techniques and physical fitness to the surface characteristics for better performance. Studying momentum variations on different surfaces is essential.

While we believe our model is applicable across surfaces, unfortunately, we couldn't find match data for various surfaces in the limited time, which will be part of our future work.

9 Sensitivity Analysis

Short-term Motivation Parameter m_i

The short-term motivation parameter m_i was estimated by us from the dataset with some bias. We perturbed m_i, ($\pm 5\%$, $\pm 10\%$), and the calculated changes in the Wimbledon final winning percentage and momentum calculations are shown in the results in the following figure:

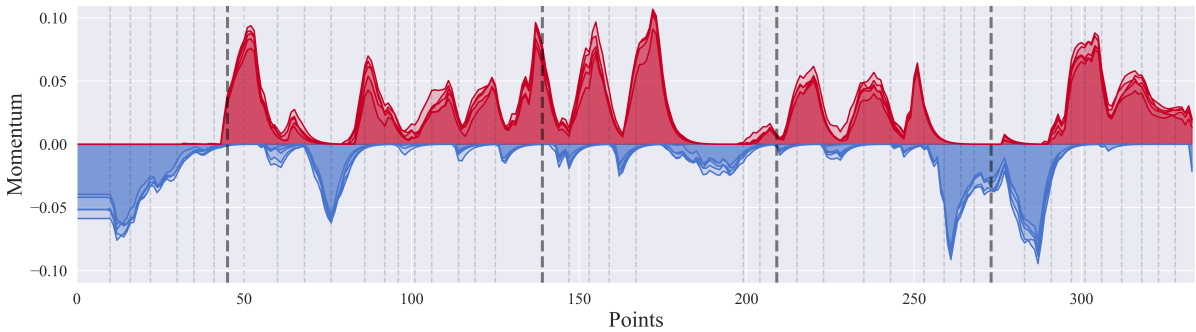


Figure 19: The Momentum Variation with respect to m_i Graph

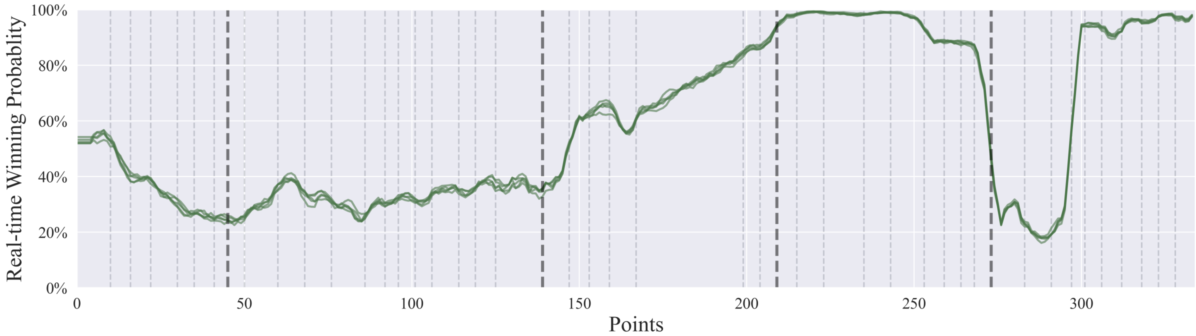


Figure 20: Score Quantity Sensitivity Analysis

From the above figure, it can be observed that the win rate and momentum are not sensitive to the variations in m_i and remain relatively stable. From the above figure, it can be observed that the win rate and momentum are not sensitive to the variations in m_i and remain relatively stable.

Parameters in the GSRF model: Score quantity s

In the GSRF model, the parameter is the given score quantity s, and sensitivity analysis is required. By varying the score quantity from 1 to 10, we found that the accuracy remains above 85%. We selected the top five stable indicators in terms of feature importance and plotted the changes in their feature importance. The figure below illustrates the sensitivity to s.

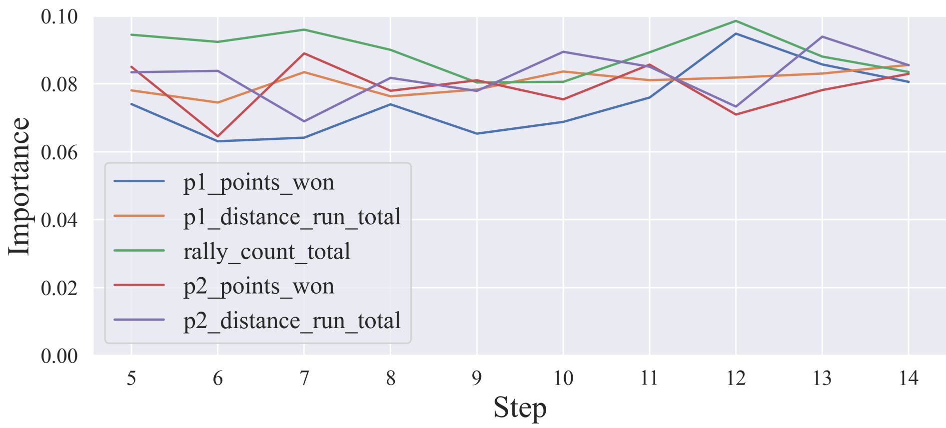


Figure 21: The Win Rate as a Function of m_i Graph

From the figure, it can be observed that within a short-term (i.e., not too long a score quantity), the changes in feature importance are not sensitive, indicating the robustness of the GSRF model.

10 Strength and Weakness

10.1 Strength

- For different sports, design rules for calculating win rates, incorporating key indicators to enhance the model's applicability.
- The GSRF model demonstrates a high accuracy rate of 94% in predicting match fluctuations, with very few misclassified samples.

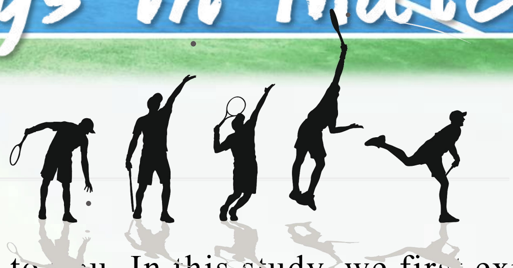
10.2 Weakness

- There hasn't been an attempt to employ machine learning models to measure the magnitude of changes in winning probability based on whether a match is won or lost.

11 Further Discussion

- For different sports, design rules for calculating win rates, incorporating key indicators to enhance the model's applicability.
- Utilize the GSRF model to predict match fluctuations and the significance of indicators, providing athletes with more detailed performance improvement strategies.

Memo
Managing Momentum Swings in Matches



To:Tennis Team
From:Team#2425454
Date : February 5

Dear Coach,

It is my honor to present our research findings to you. In this study, we first explored the relationship between "momentum" and changes in the game situation. We then further developed models capable of predicting fluctuations in the game situation and identified key factors influencing match outcomes.

Momentum refers to the psychological advantage a player gains by consecutively winning points or matches. It provides players with a sense of control in the game, exerting pressure on opponents and increasing their likelihood of making mistakes. When players have momentum, they play more dynamically, reducing fear and inhibition, leading to a more aggressive and confident playing style.

Next, let me explain how to use momentum to win more tennis matches and how to prepare players to cope with events that influence the course of tennis matches.

Taking Control from the Start

The first and arguably most crucial step in gaining momentum in tennis is to have a strong start. The initial points of a match are pivotal, setting the tone for the rest of the game. Winning the early points instills confidence and control, putting pressure on your opponent.

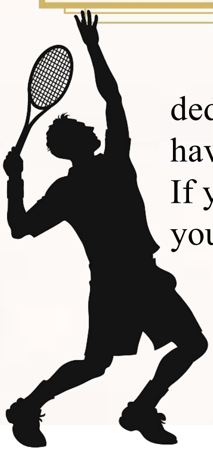
Focus on Your Footwork

Footwork is another essential aspect of gaining momentum in tennis. Your movement on the court can have a significant impact on your ability to control the game.

Command the Net

Controlling the net is a key strategy in establishing momentum in tennis. This involves approaching the net and executing volleys and overhead shots. Being at the net forces your opponent to make powerful shots and quick decisions.

Remember, momentum is fleeting and can shift in an instant. Maintaining mental resilience ensures you can seize it when it works in your favor and reclaim it when it doesn't. Mastering the art of momentum in tennis can give you a significant advantage over your opponents, leading to more victories and ultimately elevating your level of play.



Thank you for reading, as the same love tennis people, we sincerely admire the dedication and efforts of tennis players. We hope that our models and recommendations will have a positive impact on your training and competition, and help you achieve better results. If you have any more questions, feel free to email us and we'll be happy to discuss them with you.

Look forward to seeing your team achieve more victories in future competitions!

Yours Sincerely,
Team #2425454

References

[1] Agnieszka M Madurska. “A set-by-set analysis method for predicting the outcome of professional singles tennis matches”. In: *Imperial College London, Department of Computing: London, UK* (2012), pp. 1–98.

[2] Ana Šarčević;Damir Pintar;Mihaela Vranić;Ante Gojsalić. “Modeling In-Match Sports Dynamics Using the Evolving Probability Method”. In: *Applied Sciences* (2021), pp. 4429–4508.

[3] Jan R Magnus and Franc JGM Klaassen. “On the advantage of serving first in a tennis set: four years at Wimbledon”. In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 48.2 (1999), pp. 247–256.

[4] Karl Christoph Klauer, Thomas Jacobsen, and Gerd Migulla. “Counterfactual processing: Test of an hierarchical correspondence model”. In: *European Journal of Social Psychology* 25.5 (1995), pp. 577–595.

[5] Seidl Robert;Lucey Patrick. “Live Counter-Factual Analysis in Women’s Tennis using Automatic Key-Moment Detection”. In: *MIT Sloan Sports Analytics Conference*. (2022).

[6] Brandon Giles, Stephanie Kovalchik, and Machar Reid. “A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis”. In: *Journal of sports sciences* 38.1 (2020), pp. 106–113.

[7] Noora Shrestha. “Factor analysis as a tool for survey analysis”. In: *American Journal of Applied Mathematics and Statistics* 9.1 (2021), pp. 4–11.

[8] Zhanneta Kozina et al. “General and individual factor structure of complex preparation of young tennis players of 10-12 years”. In: *Journal of Physical Education and Sport* 20 (2020), pp. 1242–1249.

[9] Shabrina Tri Asti Nasution, Devi Ayu Putri, and Thezar Fiqih Hidayat Hasibuan. “AN ANALYSIS FACTOR THROUGH KAISER-MEYER-OLKIN: DISCLOSING AUDIT TENURE IN BANKING COMPANIES LISTED ON INDONESIA STOCK EXCHANGE 2019–2022”. In: *Jurnal Ekonomi* 12.01 (2023), pp. 1694–1700.

[10] Rod Cross. “Measurement of the speed and bounce of tennis courts”. In: *Sports Technology* 3.2 (2010), pp. 112–120.

Report on Use of AI

1. OpenAI ChatGPT(GPT 3.5)
   Query1:<p1_games
p1_points_won
p1_distance_run
brp1
p1_distance_run
ntp1
p1_ace
p1_ace
p1_break_pt
p1_double_fault
p1_double_fault
p1_net_pt
p1_net_pt_won
p1_score
p1_sets
p1_unf_err
p1_winner

Put this data together into a paragraph, these are the individual athlete indicators data. The first six of them are our own construction, meaning respectively: cumulative wins up to this point
Cumulative total score up to this point
The distance the point runs
Service break rate
Total distance run by the player up to that point
Scoring rate near the net

Please tidy it up for me>
   Output:<The following is a compilation of the individual metrics data you provided for athletes:

p1_games: Cumulative wins up to this point
p1_points_won: indicates the total score as of this point
p1_distance_run: indicates the run distance of this point
brp1: Break rate
p1_distance_run: The total distance run by the player up to this point
ntp1: near net scoring rate
p1_ace: indicates the total number of ACES
p1_break_pt: Total break points
p1_double_fault: indicates the total number of double errors
p1_net_pt: indicates the total number of online accesses
p1_net_pt_won: indicates the total number of Internet access scores
p1_score: indicates the total score
p1_sets: indicates the total number of disks
p1_unf_err: Total number of unforced errors
p1_winner: Total number of winning points
Note that "p1_distance_run" appears twice, if this is an error, please provide the correct variable name.>

2. OpenAI ChatGPT(GPT 3.5.)
   Query1:<Please provide me with suggestions for modification>
   Output:<I changed "p1_distance_run" to "p_distance_run_total" to distinguish it from "p_distance_run". If this is still a problem, provide the correct variable name.>

3. OpenAI ChatGPT(GPT 3.5)

Query1:<The model built needs to be able to predict whether a "swing" will occur at s points in the future and its direction - whether it will be a swing in favor of the athlete or a swing in favor of his opponent. To do this, we first defined three possible scenarios for each point and the concept of swing turning points.>

Output:<Your description is already quite clear, but if you'd like a refined version, you might consider the following expression:

"Our model aims to predict whether there will be 'fluctuations' in s future points of a match and accurately determine their direction—whether they favor the athlete or their opponent. To achieve this, we have clearly defined three possible scenarios for each point and introduced the concept of a 'fluctuation turning point.'">

4. OpenAI ChatGPT(GPT 3.5)

Query1:<What does rally mean in tennis>

Output:<In tennis, "rally" refers to the continuous exchange of shots between two or more players. During a match, players engage in a back-and-forth of hitting the ball, responding to each other's serves and returns, forming a coherent sequence of shots. This continuous exchange of shots typically takes place in different areas of the court until one player fails to return the opponent's shot or makes an error. "Rally" reflects the intense competition and the application of skills and tactics in a tennis match, making it a common element of the game.>

5. OpenAI ChatGPT(GPT 3.5)

Query1:<What can you call a model that predicts swings in the match based on GSRF>

Output:<1. **GSRF Swing Predictor Model**
2. **Match Dynamics Forecasting with GSRF**
3. **OptiSwing Model Powered by GSRF**
4. **GSRF Momentum Forecast Model**
5. **PrecisionSwing: A GSRF-based Model**

When choosing a name, make sure it clearly conveys the model's purpose and the features based on GSRF, highlighting the functionality and advantages of the model.>

6. OpenAI ChatGPT(GPT 3.5)

Query1:<What are the characteristics and norms of memo writing>

Output:<The writing of a memo should have the following characteristics and guidelines:

1. **Conciseness:** Memos should be brief and to the point, avoiding lengthy and unnecessary details. Information should be presented directly, enabling readers to quickly grasp the main content.

2. **Clarity and Specificity:** Memos should be clear and specific in conveying information. Avoid ambiguity and provide details that help readers understand the purpose and context of the memo.

3. **Structured Format:** Memos typically follow a structured format, including headings such as "To," "From," "Date," and "Subject." This format helps organize information and makes it easy for readers to locate key details.

4. **Professional Tone:** Maintain a professional and formal tone in memo writing. Use clear language and avoid slang or overly casual expressions.

5. **Actionable Content:** If applicable, include actionable items or recommendations in the memo. Clearly state what actions, if any, need to be taken by the recipients.

6. **Relevance:** Ensure that the content of the memo is relevant to the recipients and pertains to the subject at hand. Avoid including unnecessary information that may distract from the main message.

7. **Correct Grammar and Spelling:** Check and double-check for proper grammar, punctuation, and spelling. Memos should be error-free to convey professionalism and attention to detail.

8. **Date and Reference:** Include the date of the memo and any relevant references or attachments. This helps provide context and allows for easy cross-referencing.

9. **Audience Consideration:** Tailor the memo to the specific audience, considering their background, knowledge, and needs. This ensures that the information is presented in a way that resonates with the recipients.>

7. OpenAI ChatGPT(GPT 3.5)
   Query1:<Format cloth>
   Output:<The format of a memo typically includes the following elements:

1. **Header Information:**
   - Date: The date of the memo.
   - To: The primary recipient of the memo.
   - From: The author of the memo.
   - Subject: The subject or purpose of the memo.

   Example:
   ```
   Date: [Date]
   To: [Recipient]
   From: [Author]
   Subject: [Subject]
   ```

2. **Introduction (Opening Paragraph):**
   - Briefly introduce the purpose and background of the memo.
   - Provide an overview to let the reader know the main content of the memo.

   Example:
   ```
   Dear [Recipient],

   I am writing to [Purpose]. This memo outlines [Overview].
   ```

3. **Body:**
   - Paragraphs elaborating on the main content, ensuring logical flow.
   - Use headings, numbering, or other formatting to emphasize important information.

   Example:
   ```
   1. Overview
   [Content]

   2. Recommendations
   [Content]
   ```

4. **Conclusion:**
   - Summarize the main points or recommendations of the memo.
   - Emphasize expected actions or decisions.

   Example:
   ```
   ```

In conclusion, [Summarize points/recommendations]. I recommend [Specific recommendations].
```

5. **Appendix or Supplementary Information (if needed):**
   - Include additional data, charts, images, etc., to support main points.

   Example:
   ```
   Appendix A: [Additional Information Title]
   [Content]
   ```

6. **Closing Remarks:**
   - Conclude the memo, thanking the reader for their time.
   - Provide further contact information for discussions regarding any questions or concerns.

   Example:
   ```
   Thank you for your attention to this matter. If you have any questions or concerns, please feel free to contact me at [Contact Information].
   ```

7. **Signature:**
   - Provide the author's signature and position.

   Example:
   ```
   Sincerely,
   [Signature]
   [Position]
   ```

Ensure the overall layout of the memo is clear, use concise language to ensure readers can quickly understand and take necessary actions.>

8. OpenAI ChatGPT(GPT 3.5)
   Query1:<To show the fluctuation data of a game, where can you find such data>
   Output:<To showcase fluctuations in a game through data, you can obtain real-time data from various sources. Here are some common data sources:

1. **Official Sports League Websites:** Most sports leagues provide real-time match data on their official websites, such as the NBA, NFL, soccer leagues, etc. These websites typically offer scores, statistics, and real-time updates on the progress of the game.

2. **Sports News Websites and Apps:** Sports news websites and apps often provide real-time match updates and statistical data. Platforms like ESPN, CBS Sports, Yahoo Sports, etc., are common sports news sources offering real-time scores, player statistics, and more.

3. **Professional Data Providers:** Some professional sports data providers, such as Sportradar, Stats Perform, etc., offer comprehensive real-time sports data services. These companies cater to media, betting companies, and other organizations, and may provide APIs for developers.

4. **Broadcasting Platforms:** On official or affiliated broadcasting platforms for certain sports events, you can find real-time match data. This may include live commentary, score statistics, and other information.

5. **Social Media:** Social media platforms, especially Twitter, can be a source of real-time game
   ```

data. Official sports accounts, reporters, teams, and players often provide live updates during matches.

Regardless of the data source you choose, ensure that the data is reliable, updated in real-time, and aligns with the information you want to convey. Some data sources may require registration or the use of API keys for data access. When using such data, adhere to relevant usage agreements and legal regulations.>

9. OpenAI ChatGPT(GPT 3.5)
   Query1:<Run analysis of the sequence of the scoring order of the opposing teams

What can we call this section>
   Output:<The title for this section can be "Sequential Run-Length Analysis of Match Scores in Tennis." This title explicitly expresses the theme of the section, involving a run-length analysis of the sequential scores of opposing sides in tennis matches to gain insights into scoring patterns and trends during the game.>

10. OpenAI ChatGPT(GPT 3.5)
   Query1:<
Counterfactual frameworks are often used to estimate the impact of an event on an outcome. In tennis, it is possible to quantify the impact of each shot on the outcome of a match by introducing leverage.

What can the section be called>
   Output:<The title for this section can be "Leverage Analysis in Tennis: Quantifying the Impact of Each Point." This title explicitly conveys the theme of the section, involving the analysis of the potential impact of each point on the match outcome using a counterfactual framework and the concept of leverage.>

11. OpenAI ChatGPT(GPT 3.5)
   Query1:<What are the game statistics for all the games prior to the game the player is currently competing in>
   Output:<The data encompassing a player's performance in all the preceding matches before the current one is commonly referred to as "Pre-match Data" or "Historical Data." This incorporates a player's previous performances, including scoring, assists, goal differentials, and other relevant statistics. Such data holds significant reference value for analyzing and predicting a player's performance in the current match.>

12. OpenAI ChatGPT(GPT 3.5)
   Query1:<Short term Motivation and Confidence Boost What can this bar be called for an increase in the scoring rate of the serve>
   Output:<This section can be titled "Impact of Short-term Motivation and Confidence Boost on Serve Success Rate," highlighting the influence of short-term motivation and confidence enhancement on the success rate of serves.>

13. OpenAI ChatGPT(GPT 3.5)
   Query1:<momentum>
   Output:<The term "Momentum" has different definitions in various fields. Here are some common meanings:

1. **Physics:** In physics, momentum is a property of an object in motion, defined as the product of its mass and velocity. It is a conserved quantity in a conservative system, meaning that in the absence of external forces, the total momentum of a system remains constant.

2. **Financial Markets:** In the financial domain, momentum is often used to describe the trend of asset prices. Momentum in asset prices refers to the rate of change over a period, typically expressed as a percentage change or return over a specific time frame.

3. **Business and Management:** In business and management, momentum might refer to the competitive strength and growth trajectory of an organization or company in the market. A company with positive momentum in the market suggests continuous success in sales, market share, or innovation.

4. **Sports:** In sports, momentum is commonly used to describe the performance trend of a team or athlete during a game. A player or team with momentum is considered to have a continuous string of success in the competition.

In essence, "Momentum" signifies a trend or development in a particular direction. Its interpretation may vary depending on the context in which it is used.>

14. OpenAI ChatGPT(GPT 3.5)
Query1:<leverage>
Output:<"Leverage" is a financial and economic term commonly used to describe the practice of borrowing or using debt capital to increase investment returns. In the realms of investment and finance, leverage can enhance returns, but it also amplifies risks. Leverage can be achieved through debt financing, derivatives, and other means with the goal of increasing the rate of return on an investment.

Moreover, in different contexts, "leverage" can also be used to describe the concept of achieving greater effects through technical or strategic means, such as in the business context of "strategic leverage." In the sports domain, "leverage" may refer to the tactics or strategies employed during a game to achieve better results. In different contexts, the term may carry slightly different meanings.>

15. OpenAI ChatGPT(GPT 3.5)
Query1:<When an athlete is good in a game we call it hothand, but what can you call an athlete's state over time>
Output:<Over an extended period, a player's performance is commonly referred to as their "steady state" or "long-term performance." The term contrasting with "hothand" might be "coldhand," used to describe a state of consistently poor or unstable performance over the long term. Both terms are often employed to characterize the performance trends of athletes in games. It is essential to note, however, that these terms can be contentious at times, as in sports psychology, it is sometimes challenging to substantiate whether these phenomena truly exist.>

16. OpenAI ChatGPT(GPT 3.5)
Query1:<A tennis coach is skeptical that “momentum” plays any role in the match. Instead, he
postulates that swings in play and runs of success by one player are random. Use your model/metric to assess this claim.

What is the opposite of this coach's view>
Output:<The opposite viewpoint to the tennis coach's skepticism about the role of "momentum" in the match would be the belief that "momentum" does indeed play a significant role in the game. This opposing perspective suggests that swings in play and runs of success by one player are not entirely random but are influenced by momentum, which could be attributed to various factors such as confidence, skill, or psychological aspects affecting player performance.>

17. OpenAI ChatGPT(GPT 3.5)
Query1:<What can I call the chapter that uses contingency tables to analyze the relationship between score and momentum>
Output:<This section can be titled either "Correlation Analysis of Scores and Momentum" or "Contingency Table Analysis of Scores and Momentum.">