

## Study on Momentum in Tennis Matches

As the level of competition in tennis continues to rise, the outcomes of matches are influenced not only by technical factors but also by non-technical factors such as the psychological state of the athletes, their confidence and so on. We refer to this combination of factors as "momentum." This paper aims to explore the role of momentum in tennis matches and predict changes in momentum. In order to comprehensively analyze the impact of momentum, we developed five models including the **Analytic Hierarchy Process (AHP)**, **Support Vector Machine (SVM)**, **TOPSIS**, **regression analysis**, and **neural network models**.

**For Problem 1**, we established an **AHP model** to score player performances by constructing a criteria layer and judgment matrix. After consistency testing ( $CI=0.1081$ ,  $CR=0.0965<0.10$ ), we calculated the **weights of service point win rate, break point win rate, winning serve rate, double fault rate and break rate**. They were **0.1951, 0.0904, 0.1469, 0.3250, 0.2426**. Then, we evaluated the performance of players in each game and provided a visual chart of players' performance excellence according to these weights.

**For Problem 2**, we developed an **SVM-based model** with features reflecting momentum, such as consecutive game wins, performance on crucial points, and service errors, and optimized the regularization parameter and kernel function parameter. The model's performance metrics showed an accuracy score of 0.933, precision of 0.95, recall of 0.98, and f1-score of 0.96. These results demonstrated the effectivity of our model. Moreover, through the **Kolmogorov-Smirnov test** (KS statistic: 0.169, P-value:  $1.77e-05$ ), we confirmed that changes in momentum are not random but closely related to fluctuations in the match and victory.

**For Problem 3**, Based on AHP, we determined the weights for scores related to momentum, consecutive game wins, consecutive set wins, and consecutive match wins with **0.5396, 0.2970, and 0.1634**, respectively. Then, we scored momentum by using the **TOPSIS** method and integrating most indicators from the dataset. Through **linear regression analysis (fitting degree of 0.9562)** and **neural network models**, we scored and predicted changes in momentum, presented visual predictions of momentum, and identified the most relevant factors affecting players' momentum fluctuations: consecutive scoring, service quality, The score of the previous game, and rally count. We offered suggestions for playing against different opponents based on momentum.

**For Problem 4**, We achieved **predictive performance metrics**:  $R^2=0.953$ ,  $MAE=3.0696 \times 10^{-5}$ ,  $MBE=-1.6126 \times 10^{-5}$  by using neural networks to predict fluctuations in multiple matches. These results indicated that the model is excellent to predict match momentum fluctuations. Meanwhile, we discussed the model's general applicability to other matches from various perspectives.

**For Problem 5**, regarding the impact of momentum fluctuations on match outcomes, we provided suggestions for coaches in a memo, focusing on the role of momentum and the importance of adjusting players' momentum.

**Key Words:** Analytic Hierarchy Process; Support Vector Machine; Kolmogorov-Smirnov test; TOPSIS; Regression Analysis; Neural Network Models; Momentum

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Background	1
1.2	Restatement of the Problem	1
1.3	Our Work	2
<b>2</b>	<b>Assumptions</b>	<b>3</b>
<b>3</b>	<b>Solution to Problem 1</b>	<b>3</b>
3.1	Evaluating the Performance of Players : Analytic Hierarchy Process	3
3.1.1	Data Preprocessing	3
3.1.2	Notation	4
3.1.3	Establishing the Hierarchy	4
3.1.4	Constructing the Judgment Matrix	4
3.1.5	Consistency Check of the Judgment Matrix	5
3.1.6	Weight Calculation and Comprehensive Evaluation	5
3.1.7	Result Analysis	5
3.2	Visualization of Performance Level	6
<b>4</b>	<b>Solution to Problem 2</b>	<b>7</b>
4.1	Support Vector Machine Model	7
4.1.1	Model Construction and Evaluation	7
4.2	Kolmogorov-Smirnov Test	7
<b>5</b>	<b>Solution to Problem 3</b>	<b>8</b>
5.1	Evaluation of Momentum : Analytic Hierarchy Process	8
5.2	Technique for Order Preference by Similarity to Ideal Solution	10
5.3	Linear Regression	10
5.4	Neural Networks Model	13
5.5	Advice to Players Based on Momentum Fluctuations	17
<b>6</b>	<b>Solution to Problem 4</b>	<b>18</b>
6.1	The Model Prediction Performance	18
6.2	The Generalizability of the Model	18
<b>7</b>	<b>Conclusion</b>	<b>19</b>
	<b>References</b>	<b>20</b>
	<b>Appendices</b>	<b>20</b>

## Memo

To: Coach

From: Team 2425262

Date: February 5, 2024

Subject: The Effect of Momentum and the Suggestion Based on Momentum in Tennis Matches

Dear Coach,

I am honored to report to you the results of our mathematical modeling study on the phenomenon of "momentum" in tennis matches. Through in-depth analysis of historical tennis match data, we have explored the impact of "momentum" on the fluctuation of match outcomes and investigated various factors influencing "momentum." Additionally, we have delved into strategies for players to respond to events that affect the course of the game. Below are our findings and recommendations:

### 1. Definition and Quantification of Momentum

We define momentum as the force or power generated by continuous events (such as consecutive points won) in a match, influencing the trajectory of the game. By analyzing data from various matches, we have established a model for quantifying and measuring changes in momentum, including features like consecutive points won, consecutive game wins, point differentials within a game, break point conversion rates, serve error rates, and service game win rates.

### 2. Factors Influencing Momentum

Changes in momentum are closely related to various factors. We have identified factors such as consecutive points won, consecutive game wins, serve points won, set points won, successful key serves, serve errors, and break point conversion rates as influencing the evolution of momentum. In a match, momentum changes are not only a reflection of the score but are also associated with the actual circumstances of the game and the psychological aspects of the players.

### 3. Impact of Momentum

We have observed that momentum has a significant impact on the flow of a match. Fluctuations in momentum can shift the balance from one player to another, and the degree of a player's momentum and its fluctuations can affect the outcome of the game. Establishing momentum can have far-reaching consequences for the entire match, especially in critical sets where changes in momentum may directly determine the match's outcome. Therefore, cultivating strategies for players to adjust momentum is crucial.

### 4. Strategies for Player Preparation

a. Preparation Phase: Prior to a match, thorough analysis of the opponent's characteristics and performance is essential. Coaches can assist players in understanding their opponents, identifying strengths and weaknesses, and formulating tailored tactics to gain control of the match. Developing strategies to overturn momentum and seize dominance is crucial.

b. During the Match: Monitoring changes in momentum during the match is critical. Utilizing metrics such as consecutive points won, consecutive game wins, serve points won, set points won, successful key serves, serve errors, and break point conversion rates, coaches can guide players to be more aggressive when their momentum is favorable and prioritize control and stability when facing unfavorable momentum. Adjusting tactics in different situations, prioritizing actions to improve one's momentum, and preventing the opponent's momentum from rising are key considerations. Special attention should be given to crucial points or set points, as these are sensitive areas where momentum can swing dramatically. Simulating key situations in training helps players develop composure and effective responses during critical moments. Changes in momentum can impact a player's psychology. Coaches should pay attention to players' reactions to events affecting the flow of the game, such as consecutive serve errors or crucial mistakes during key points. Psychological training and techniques to maintain composure and confidence in the face of mistakes can help stabilize and gradually increase momentum, improving the chances of winning the match.

Through this mathematical modeling study, we have gained a deeper understanding of the role of momentum in tennis matches. The recommendations provided are aimed at assisting players in responding effectively to events that impact the course of the game. We hope these suggestions will serve as valuable guidance for coaches in training sessions, contributing to the overall improvement of players' performance. We are eager to discuss and explore these findings further with you.

Hope that our models and these suggestions will be helpful to you !

Sincerely,  
Team 2425262

# 1 Introduction

## 1.1 Problem Background

With the rapid development of various competitive sports, the competition in all kinds of events is becoming increasingly fierce. Many athletes and coaches pay close attention to how to achieve victory in competitions. It is not solely the technical aspects that influence the outcome of a match. Due to the complexity of competitions, factors such as the athlete's mindset, physical condition, and winning streak during the match also significantly impact their performance, leading to a tendency for the match to favor a particular player. We refer to these factors as "momentum".

Taking professional tennis as an example, the 2023 Wimbledon Championships singles final provided a spectacular match between 20-year-old newcomer Carlos Alcaraz and Grand Slam player Novak Djokovic, with the match being full of twists and turns, ultimately concluding with a victory for Carlos Alcaraz. The records of this match and numerous other tennis matches indicate that a player's "momentum" greatly affects the outcome and direction of a match. Various events during a match can influence a player's "momentum". Our study focuses on the factors affecting a player's "momentum" based on extensive tennis match data, to provide coaches with insights for guiding players towards victory.

## 1.2 Restatement of the Problem

Considering the background information identified in the problem statement, the following problems need to be solved:

- Problem 1: Develop a model capable of determining which player performs better during a specific period of a match and to what extent their performance excels. Based on this model, provide a visualized description of the match flow.
- Problem 2: Establish a model to assess whether "momentum" affects the real-time fluctuations of the match and its outcome.
- Problem 3: Construct a model to predict the fluctuations of a match and its key influencing factors. Utilize the differences in "momentum" fluctuations from past matches to offer advice to athletes on dealing with various opponents.
- Problem 4: Evaluate the sensitivity of the predictive model developed in Problem 3 in forecasting match fluctuations and discuss the model's applicability to other sports.
- Write a memo. Summarize the research findings and offer recommendations to coaches, including the role of "momentum" and how athletes can prepare to handle events that affect the course of a tennis match.

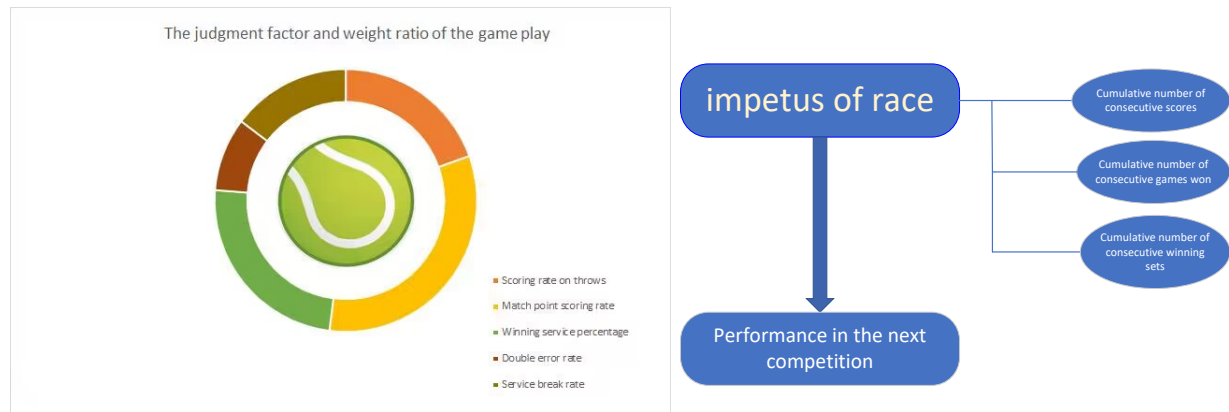


Figure 1: Judgment Criteria

### 1.3 Our Work

- Initially, we employed the Analytic Hierarchy Process (AHP) model to capture the excellence of players' performance during specific time intervals in matches. Through the hierarchical analysis of five indicators - serve point percentage, match point percentage, winning serve percentage, double fault rate, and break rate - we constructed a well-performing model for rating players' performance.
- We established Support Vector Machine models and conducted KS tests to validate the significant correlation between player momentum during matches and fluctuations in the game, as well as the outcome of sets.
- Utilizing AHP, we determined the weights for consecutive winning points, consecutive winning games, and consecutive winning sets in the context of momentum. Subsequently, employing the TOPSIS method, we scored momentum by considering a comprehensive set of indicators from the dataset. Following this, we utilized linear regression analysis and a neural network model to score momentum fluctuations and predict changes in momentum. The most relevant factors influencing player momentum were identified. Given the substantial impact of momentum changes on the game, we recommend players analyze momentum-related factors comprehensively when preparing for matches against different opponents.
- In the application of regression analysis and training neural network models, we conducted training and testing across multiple matches. Evaluation metrics demonstrate that our models exhibit strong performance in predicting game fluctuations and momentum changes. If the models exhibit suboptimal performance at times, adjusting or adding features can further enhance their performance. Moreover, a comprehensive analysis of the models' generalizability for other matches indicates overall good applicability.
- In consideration of the significant impact of momentum fluctuations on matches, we provide coaches with recommendations in the memorandum section, cover-

ing various aspects.

## 2 Assumptions

- Assumption 1: Assume that the external environmental conditions of a tennis match (such as weather, court conditions, audience, etc.) affect both players equally and do not confer an unfair advantage or disadvantage to either side.
- Assumption 2: Assume that in a tennis match, the server has a fixed advantage, with a higher probability of winning the point than the receiver.
- Assumption 3: Assume that the psychological factors during a tennis match (such as winning streaks, losing streaks, psychological impacts of service errors) have the same effect on all players, to facilitate the development and prediction of subsequent models.
- Assumption 4: Assume that each tennis player's physical fitness remains constant throughout each match, and the duration or intensity of the match has the same impact on all players' physical condition changes, for the ease of predictive model development.
- Assumption 5: Assume that tennis players' playing styles and strategies remain relatively consistent throughout the match and do not undergo fundamental changes due to the progress of the match, ensuring the stability of the model.
- Assumption 6: Assume that the provided dataset is relatively complete and accurate, ensuring the effectiveness of model analysis after data preprocessing.

## 3 Solution to Problem 1

### 3.1 Evaluating the Performance of Players : Analytic Hierarchy Process

#### 3.1.1 Data Preprocessing

- Missing Value Handling: Check for missing data points in the dataset and use interpolation methods to fill in the gaps.
- Outlier Detection: Detect and handle outliers using the 3-sigma rule to identify anomalies and replace them with mean values for correction.
- Feature Engineering: Construct new features based on existing data, calculating each player's service point win rate, match point win rate, ace rate, double fault rate, and break point conversion rate for each set, and perform statistical analysis.

### 3.1.2 Notation

Table 1: Notation

Symbol	Definition
SP	Service Point Win Rate
MP	Match Point Win Rate
AR	Ace Rate
DF	Double Fault Rate
BP	Break Point Conversion Rate

### 3.1.3 Establishing the Hierarchy

After analyzing the data, filtering, and establishing new features, this paper identifies the factors affecting players' performance as: service point win rate, match point win rate, ace rate, double fault rate, and break point conversion rate for study. The decision-making hierarchy is structured as follows:

- Goal Layer: To assess the degree of excellence in performance of players during matches.
- Criteria Layer: Service point win rate, match point win rate, ace rate, double fault rate, break point conversion rate.
- Alternative Layer: Two players in a particular match.

### 3.1.4 Constructing the Judgment Matrix

For each criterion, a judgment matrix is constructed. Using the 1-9 scale of the Analytic Hierarchy Process (AHP), the pairwise comparison matrix for the degree of excellence in performance of players, involving service point win rate, match point win rate, ace rate, double fault rate, break point conversion rate, is as follows:

$$A = \begin{bmatrix} 1 & 2 & 3 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & 2 & 1 & \frac{1}{2} & 1 \\ 2 & 3 & 2 & 1 & 2 \\ 3 & 2 & 1 & \frac{1}{2} & 1 \end{bmatrix}$$



### 3.1.5 Consistency Check of the Judgment Matrix

To ensure the correctness and rationality of the weights obtained, a consistency check is required. There may be inaccuracies in the evaluation of some indicators, possibly leading to deviations in the eigenvalues of the judgment matrix. The judgment matrix should have overall consistency. Calculate the Consistency Index (CI) and Consistency Ratio (CR). When the  $CR < 0.1$ , the weights derived from the judgment matrix are deemed reasonable and effective, passing the consistency check. Otherwise, the judgment matrix needs to be adjusted until it passes the consistency test.

Using MATLAB software for the construction of the AHP model, the obtained

- Consistency Index (CI) = 0.1081,
- Consistency Ratio (CR) = 0.0965.

Since  $CR < 0.10$ , the consistency of judgment matrix A is acceptable.

### 3.1.6 Weight Calculation and Comprehensive Evaluation

Using MATLAB for AHP, the weights of factors influencing players' performance degree in matches are calculated using arithmetic mean, geometric mean, and eigenvalue methods, as shown in the table below for service point win rate, match point win rate, ace rate, double fault rate, break point conversion rate:

Table 2: Weights of Indicators

Indicator	Weight
Service Point Win Rate	0.1951
Match Point Win Rate	0.0904
Ace Rate	0.1469
Double Fault Rate	0.3250
Break Point Conversion Rate	0.2426

Based on the weights in the table above, we can construct a formula to calculate the performance level, where each factor is multiplied by its corresponding weight. Let SP be the service point win rate, MP the match point win rate, AR the ace rate, DF the double fault rate, and BP the break point conversion rate. Then, the calculation formula Y (representing the player's degree of excellence in performance in each set) can be expressed as:

$$Y = 0.1951 \times SP + 0.0904 \times MP + 0.1469 \times AR + 0.3250 \times DF + 0.2426 \times BP$$

### 3.1.7 Result Analysis

Through the aforementioned Analytic Hierarchy Process (AHP) model, we can determine the performance level of each player based on their service point win situation,

match point win situation, ace performance, service errors, and ability to break the opponent's service (i.e., achieving break points). The numerical value of the performance level indicates the extent of their excellence (the higher the value, the better the performance). Through hierarchical analysis, we can identify at which specific moments in the match a player performs better.

### 3.2 Visualization of Performance Level

Drawing upon the conclusions above, we utilize charts to visualize the composite performance scores of each player over time, illustrating how each player's performance evolves throughout the match. We have selected the 43-game match spanning four sets between Carlos Alcaraz and Nicolas Jarry for this analysis. The performance of each player in every game is analyzed. The following graph allows us to draw the conclusion:

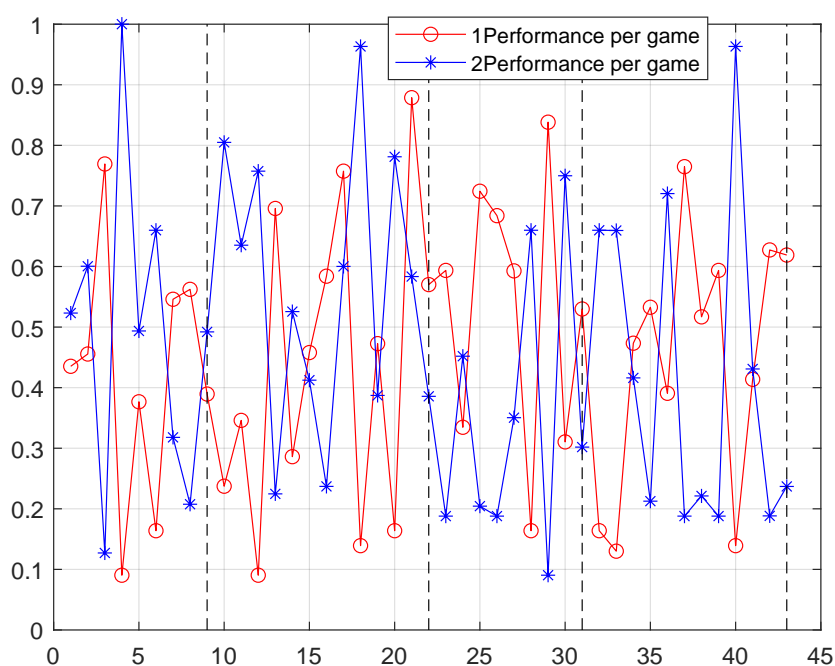


Figure 2: Visualization of Performance Level

In the match, the victories for the first through fourth sets were secured by Player 1, Player 2, Player 1, and Player 1, respectively, with the game scores (Player 1's winning games vs. Player 2's winning games) for each set being: 6:3, 6:7, 6:3, 7:5.

The provided graph reveals the performance scores of both players per game, highlighting a notably superior performance by Player 2 in the fourth set compared to the third, corroborated by the scores of these sets. Notably, the game preceding the third set was won by Player 1, and the game preceding the fourth set by Player 2, suggesting a preliminary hypothesis of a correlation between a player's performance and their immediate past performance.

## 4 Solution to Problem 2

### 4.1 Support Vector Machine Model

#### 4.1.1 Model Construction and Evaluation

We employ a machine learning model, the Support Vector Machine (SVM), using data features related to "momentum" to simulate match scenarios. If the simulated outcomes in the test set closely match the actual scenarios, it suggests a non-random relationship between "momentum" and match dynamics and victories.

SVM, a binary classification model, employs a linear classifier defined in feature space with maximum margin. It incorporates kernel techniques, rendering it effective for nonlinear classification. Ideal for small to medium datasets and high-dimensional problems, SVM is adept at handling nonlinear relationships, making it suitable for assessing "momentum's" impact on winning probabilities in tennis matches.

Feature construction prior to SVM model development includes analysis of point-by-point outcomes, serve statistics, score changes, and consecutive scoring streaks. Features reflecting "momentum" such as the number of consecutive games won, performance on crucial points, score differentials, and serving errors significantly impact model performance.

After dividing the dataset into training and testing sets, the SVM model is trained to find an optimal hyperplane for separating categories, with kernel functions exploring decision boundary shapes. Parameter adjustment and cross-validation techniques are employed to enhance model performance, selecting optimal regularization parameter C and kernel function parameters. Model efficacy is evaluated using accuracy, precision, recall, and F1 score metrics on the test set. Based on the model evalua-

Metric	Accuracy	Precision	Recall	F1-Score
Value	0.933	0.95	0.98	0.96

Table 3: Model Performance Metrics

tion metrics, it is evident that the model performs well in simulating match outcomes, demonstrating both validity and accuracy.

Consequently, we can preliminarily conclude that "momentum" is indeed related to match fluctuations and victories, rather than being a random factor.

### 4.2 Kolmogorov-Smirnov Test

Based on the SVM model, we draw a preliminary conclusion. This paper further verifies the correctness of this conclusion using the Kolmogorov-Smirnov (KS) test, suggesting that "momentum" in matches correlates with fluctuations and victories rather than being random. The KS test, a non-parametric statistical test, determines if two samples derive from the same distribution, or if a sample comes from a reference distribution, by comparing cumulative distribution functions (CDFs) and evaluating

significant differences based on their maximum divergence. This study uses the two-sample KS test to compare the distribution differences between actual "momentum" scores and simulated scores, thus testing their correlation.

Features such as consecutive scoring, score gaps, and serving errors were used. Initially, two hypotheses were defined: the null hypothesis (H0) posits no effect of "momentum" on match outcomes, while the alternative hypothesis (H1) suggests a significant effect. Using Python for the KS test, we compared the empirical distribution of match outcomes with "momentum" to the expected distribution under random conditions, yielding:

Statistic	Value
KS statistic	0.16884079601990054
P-value	$1.7714855777124417 \times 10^{-5}$

Table 4: KS Test Results

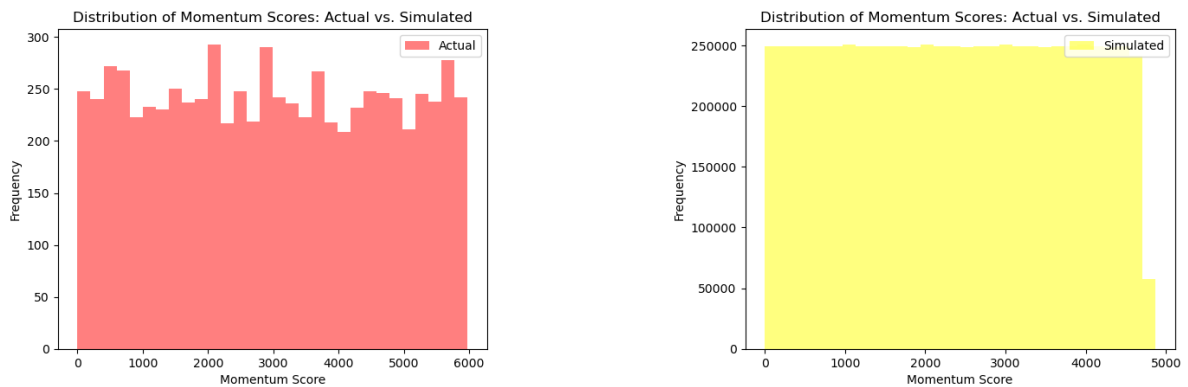


Figure 3: Visualizing the Actual Momentum Score with Simulated Game Score Distribution

Given the small p-value ( $<0.05$ ), the null hypothesis is rejected, indicating non-random momentum shifts significantly differ from random simulations, **further proving "momentum" plays a role in match fluctuations and outcomes.**

## 5 Solution to Problem 3

### 5.1 Evaluation of Momentum : Analytic Hierarchy Process

#### 1. Hierarchy Structure

After analyzing and filtering the data, and establishing new features, this study categorizes the factors influencing a player's "momentum" into three aspects: consecutive points won, consecutive games won, and consecutive sets won. The hierarchy structure for decision-making is outlined as follows:

- Objective Layer: Evaluate the player's "momentum" in the game.

- Criterion Layer: Consecutive points won, consecutive games won, consecutive sets won.
- Alternative Layer: Two players in a specific match.

## 2. Constructing Judgment Matrices

For each criterion in the criterion layer, a judgment matrix is constructed. Using the 1-9 scale of the Analytic Hierarchy Process (AHP), the pairwise comparison matrices for consecutive points won, consecutive games won, and consecutive sets won are represented as:

$$B = \begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix}$$

## 3. Consistency Check of Judgment Matrices

To ensure the correctness and rationality of the obtained weights, a consistency check is performed. Inconsistent evaluations of a criterion may lead to deviations in the matrix's eigenvalues. The consistency of the judgment matrix is evaluated by calculating the Consistency Index (CI) and Consistency Ratio (CR). When  $CR < 0.1$ , the weights are considered reasonable and effective. The calculated values for the provided matrix

Consistency Index (CI)	Consistency Ratio (CR)
0.0046	0.0088

Since  $CR < 0.10$ , the consistency of matrix is deemed acceptable.

## 4. Calculating Weights and Comprehensive Evaluation

Using Matlab software for the Analytic Hierarchy Process, the weights for the factors influencing a player's "momentum" in the game are calculated. The weights obtained using the arithmetic mean, geometric mean, and eigenvalue methods are as follows:

Criterion	Weight
Consecutive Points Won	0.5396
Consecutive Games Won	0.2970
Consecutive Sets Won	0.1634

Table 5: Weights Obtained using Analytic Hierarchy Process

Define these three indicators as variables in the following table,:

Then the formula "momentum" (representing the degree of "momentum" of the player) can be expressed as:

$$\text{Momentum} = 0.5396 \times X_1 + 0.2970 \times X_2 + 0.1634 \times X_3$$

Variable	Representation
Consecutive Points Won	$X_1$
Consecutive Games Won	$X_2$
Consecutive Sets Won	$X_3$

Table 6: Variable Representations

## 5.2 Technique for Order Preference by Similarity to Ideal Solution

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a multi-criteria decision analysis method used for evaluating and ranking alternative solutions. It effectively utilizes the information from raw data and provides precise reflections of the differences between various evaluation scenarios. The fundamental process involves normalizing the original data matrix, employing the cosine method to identify the optimal and worst-case scenarios within a finite set of alternatives, calculating the distances between each evaluation object and the optimal/worst-case scenarios, determining the relative proximity of each evaluation object to the optimal scenario, and utilizing this information as the basis for evaluating superiority or inferiority. In this context, TOPSIS is employed to identify moments in a match when a momentum shift is more likely to occur.

1. Based on the results from AHP, criteria relevant to momentum shifts, specifically consecutive points won, consecutive games won, and consecutive sets won, were selected.

2. Standardization of data was performed on the chosen three indicators to ensure uniform scales.

3. The standardized data was used to construct a decision matrix.

4. Ideal and negative-ideal solutions for momentum shifts were identified.

5. Distances between each solution and the ideal/negative-ideal solutions were calculated.

6. Comprehensive scores for each solution were computed based on the distances, and subsequent ranking was carried out.

By leveraging the weights assigned to three indicators through the Analytic Hierarchy Process (AHP), we employ the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method to determine the momentum scores for each game set. These scores are intended for subsequent applications in linear regression analysis and neural network modeling.

The following are the variables required for the subsequent models:

## 5.3 Linear Regression

Based on the prior AHP weight analysis and TOPSIS method to obtain momentum scoring data, we employ linear regression analysis to establish a model for predicting momentum fluctuations in tennis matches. Linear regression is suitable for studying

Abbreviation	Explanation
p1_ac_win_game	Player 1 Consecutive Winning Games
p1_ac_point	Player 1 Consecutive Winning Points
p1_ac_set	Player 1 Consecutive Winning Sets
p1_score	Player 1 Performance Rating in Question 1
p1_ac_score	Player 1 Momentum Rating

Table 7: Translations of Abbreviations and Explanations

the relationship between independent variables and the dependent variable (momentum score).

#### Variable Definitions:

Dependent Variable (Y): Momentum score (p1\_ac\_score).

Independent Variables (X): Various factors that may influence momentum fluctuations, including the number of consecutive winning games, consecutive points scored, consecutive winning sets, performance scores from Question 1, serving performance, match point scores, winning serve situations, serve fault situations, break situations, and other relevant features.

#### Data Preprocessing:

Ensure data quality, handle missing values, and address outliers. Standardize or normalize numerical variables to bring them to similar scales, aiding in model convergence. Assign values to qualitative variables:

winner\_shot\_type: B->1, F->2

return\_depth: D->1, ND->2, NA (mode: ND)

serve\_depth: CTL->1, NCTL->2, NA (mode: NCTL)

serve\_width: B->1, BC->2, BW->3, C->4, W->5, NA (mode: C)

#### Data Splitting:

Divide the dataset into training and testing sets. The training set (first two sets of the match) is used for model training, and the testing set is used to evaluate its performance.

#### Model Training:

Apply linear regression on the training set, using defined independent variables to predict momentum fluctuations.

The linear regression equation is given by:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \epsilon$$

Coefficients:

$$b_0, b_1, b_2, \dots, b_n$$

Independent Variables:

$$X_1, X_2, \dots, X_n$$

Error Term:

$$\epsilon$$

### Model Evaluation:

Evaluate the model's performance on the testing set using metrics such as mean squared error (MSE), R-squared, or other relevant indicators. Examine coefficients to understand the impact of each independent variable on momentum fluctuations.

The F-test is a test of the overall significance of the regression equation, specifically assessing the collective significance of all variables on the dependent variable.

$F(29, 256)$ ,  $\text{Prob} > F$  indicates significance through a joint significance test.

Source	SS	df	MS	Number of obs	=	286
Model	.000662888	29	.000022858	F(29, 256)	=	192.79
Residual	.000030352	256	1.1856e-07	Prob > F	=	0.0000
				R-squared	=	0.9562
				Adj R-squared	=	0.9513
Total	.00069324	285	2.4324e-06	Root MSE	=	.00034

Figure 4: Linear Regression F-test

The analysis of the results from the F-test reveals a significant p-value of 0.000, indicating statistical significance at a high level. The null hypothesis of a regression coefficient being zero is rejected. Additionally, the model's goodness of fit, represented by  $R^2$ , is 0.95, suggesting excellent performance. Therefore, the model generally meets the requirements.

Regarding the performance on variable collinearity, all Variance Inflation Factors (VIF) are below 10, indicating the absence of multicollinearity issues. The model construction is deemed satisfactory.

```
estat hettest ,rhs iid
reusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: pl_score rally_count speed_mph pl_distance_run pl_points_won A1 o.A2 B1 o.B2 C1 o.C2 D1 o.D2 D3 o.E1 E2 F1
o.F2 G1 o.G2 H1 H2 o.H3 I1 o.I2 J1 o.J2 K1 o.K2 L1 o.L2 o.M1 o.M2 N1 o.N2 O1 O2 O3 O4 o.O5 P1 o.P2 Q1 o.Q2

chi2(26)      =    28.21
Prob > chi2    =    0.3481
```

Figure 5: Linear Regression Test

We utilized MATLAB software to establish a linear regression analysis model. Based on the results of the model (partially shown in the appendix), along with the evaluation and testing, preliminary conclusions can be drawn. The model exhibits a good predictive performance for fluctuations in matches.

Notably, the variables  $P1\_points\_won$ ,  $serve$ , and  $point\_victor$  show the highest correlation with the momentum of players. The impact of  $Rally\_count$  on player momentum is at the second level of significance. Their respective  $P > |t|$  values are 0.000, 0.004, 0.008, and 0.022. Therefore, the conclusion can be drawn that

- the factors influencing player momentum fluctuations include the player's consecutive winning points, the quality of serves, whether the player scored in the previous round, and  $Rally\_count$ .



pl_ac_score	Coef.	Std. Err.	t	P> t	Beta
pl_score	-.0005649	.0011428	-0.49	0.621	-.075855
rally_count	-.0001835	.0000795	-2.31	0.022	-.1838786
speed_mph	5.91e-06	7.45e-06	0.79	0.429	.0438205
pl_distance_run	.0000101	.0000127	0.79	0.428	.0609062
pl_points_won	.0000119	1.87e-06	6.37	0.000	.3555957
A1	-.0005963	.000207	-2.88	0.004	-.1915123
A2	0	(omitted)			0
B1	.0001591	.0002894	0.55	0.583	.0481459
B2	0	(omitted)			0
C1	.00096	.0003594	2.67	0.008	.3076963
C2	0	(omitted)			0
D1	-.0006211	.0005072	-1.22	0.222	-.1412044
D2	0	(omitted)			0
D3	-.0004732	.0006358	-0.74	0.457	-.0718672
E1	0	(omitted)			0
E2	-.0005356	.0009013	-0.59	0.553	-.0350468
F1	-.0001076	.0006235	-0.17	0.863	-.0138567
F2	0	(omitted)			0
G1	.0003384	.0004203	0.81	0.421	.0745813
G2	0	(omitted)			0
H1	-.0002286	.0003208	-0.71	0.477	-.0574685
H2	-.0001075	.0004896	-0.22	0.826	-.01385
H3	0	(omitted)			0
I1	-.0002324	.0006671	-0.35	0.728	-.0230652
I2	0	(omitted)			0
J1	.0003385	.0003688	0.92	0.360	.0656299
J2	0	(omitted)			0
K1	.0000201	.0003247	0.06	0.951	.0035105
K2	0	(omitted)			0
L1	.0002355	.0007569	0.31	0.756	.032638
L2	0	(omitted)			0
M1	0	(omitted)			0
M2	0	(omitted)			0
N1	-.0002673	.000904	-0.30	0.768	-.0299788
N2	0	(omitted)			0
O1	.0000432	.0003304	0.13	0.896	.009414
O2	.0002445	.0003239	0.75	0.451	.0561223
O3	.0001338	.0003032	0.44	0.659	.0343339
O4	8.14e-06	.0002633	0.03	0.975	.0024362
O5	0	(omitted)			0
P1	-.0002203	.0002194	-1.00	0.316	-.0642193
P2	0	(omitted)			0
Q1	.0001035	.0001842	0.56	0.575	.0319745
Q2	0	(omitted)			0
_cons	.0025324	.0016062	1.58	0.116	.

Figure 6: Linear Regression Result

We visualized the trend of changes in P1\_ac\_score and P1\_score by creating a line graph. In this graph, apart from four anomalous points, the general trend of P1\_ac\_score and P1\_score is similar. This observation confirms the presence of a 'momentum' factor influencing performance in the game. Additionally, it is noteworthy that P1\_score generally exceeds P1\_ac\_score.

## 5.4 Neural Networks Model

Based on the results of the aforementioned model, we further established a machine learning model: a neural network model for regression analysis and predicting momentum scores. This model addresses various factors that may influence momentum fluctuations, including consecutive winning streaks, consecutive scoring instances, consecutive winning sets, performance scores from Question 1, serving point outcomes, match point outcomes, successful serve scenarios, serve errors, break situations, and other relevant features. Utilizing the neural network, we aim to make reasonable and accurate predictions for the momentum fluctuations of players.

A neural network is a multi-layer feedforward network trained using the back-propagation algorithm, and it is one of the most widely used neural network models.

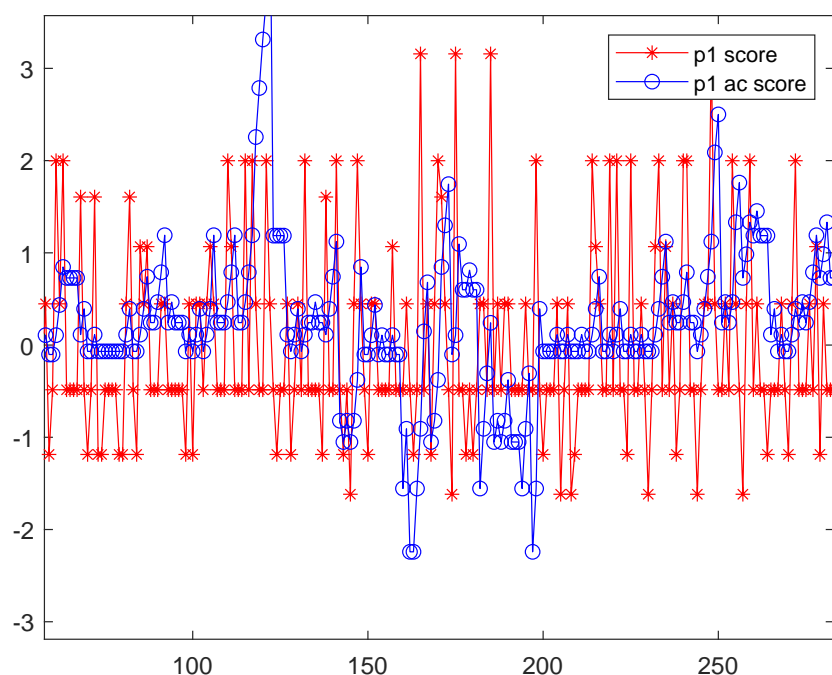


Figure 7: Linear Regression

It consists of layers of nodes, including an input layer, one or more hidden layers, and an output layer. The learning rule for a backpropagation (BP) neural network involves utilizing the steepest descent method. It continuously adjusts the network's weights and thresholds through backpropagation. If the output of any single node exceeds the specified threshold, the node is activated, and the data is sent to the next layer of the network. Otherwise, the data is not transmitted to the next layer, minimizing the sum of squared errors in the network.

#### **Data Loading and Understanding:**

Load the provided match dataset and understand the data structure, variable types, and feature distributions. Define the target variable as momentum score.

#### **Data Preprocessing:**

Handle missing values and outliers to ensure data quality. Standardize numerical features to ensure they are on a similar scale. Perform one-hot encoding for categorical variables.

#### **Data Splitting:**

Divide the dataset into training and testing sets.

#### **Feature Engineering:**

Explore the correlations between various features to identify which ones may influence momentum fluctuations. Select the most relevant features, including consecutive winning streaks, consecutive scoring instances, consecutive winning sets, performance scores from Question 1, serving point outcomes, match point outcomes, successful serve scenarios, serve errors, and break situations.

#### **Neural Network Model Construction:**

Build a neural network model using a deep learning framework. Define the number of neurons and activation functions for the input layer, hidden layers, and output layer. Choose an appropriate loss function, typically mean squared error (MSE) for regression problems.

- Initialize the BP neural network. Use random numbers in the (0,1) range to assign initial values to all connection weights. Set the error function  $e$ , allowed precision  $\epsilon$ , and maximum learning times  $L$ . Randomly select a sample for learning and set its expected output.
- Calculate the input and output of each node in the hidden layer. Utilize the expected and actual outputs of the network to compute the partial derivatives of the error function for each neuron in the output layer. Use the connection weights from the hidden layer to the output layer, along with the outputs of both layers, to calculate the partial derivatives of the error function for each neuron in the hidden layer.
- Adjust the connection weights in the neural network.
- Compute the error function for the network output.
- Assess the relationship between the output error  $E$  and the specified  $\epsilon$ . If  $E < \epsilon$  or the current learning times exceed the maximum learning times  $L$ , end the algorithm. Otherwise, select the next learning sample and return to step two for the next learning iteration.

### **Model Training:**

Input the training set into the neural network, updating model parameters using the backpropagation algorithm to minimize prediction errors. Set appropriate learning rates and training epochs.

### **Model Validation:**

Utilize a test set (several matches of a specific player) to validate the model's performance and assess its generalization ability. Examine the loss values on both the training and test sets to ensure there is no overfitting.

### **Model Fine-Tuning:**

Adjust the hyperparameters of the neural network, including learning rate and the number of neurons in the hidden layers, based on the validation results

### **Model Result:**

Below are the results of the three sets of neural network model training and testing. It can be observed that the predicted curves align well with the actual curves, indicating a good fit of the model in predicting momentum. The evaluation of the model will be further discussed in the next section.

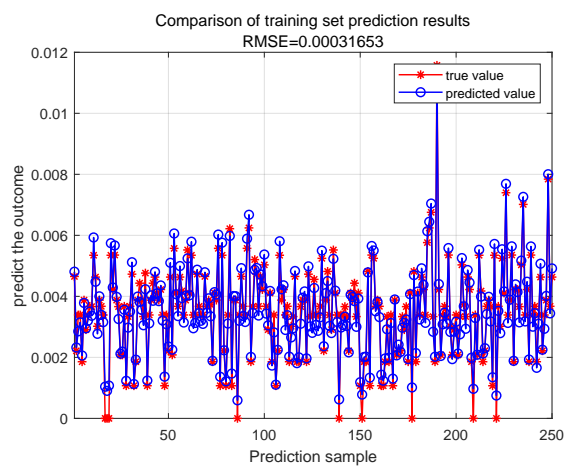


Figure 8: Neural Network Model Prediction Results 1

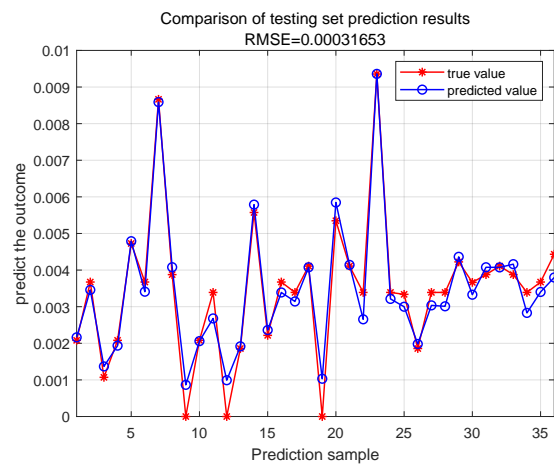


Figure 9: Neural Network Model Prediction Results 2

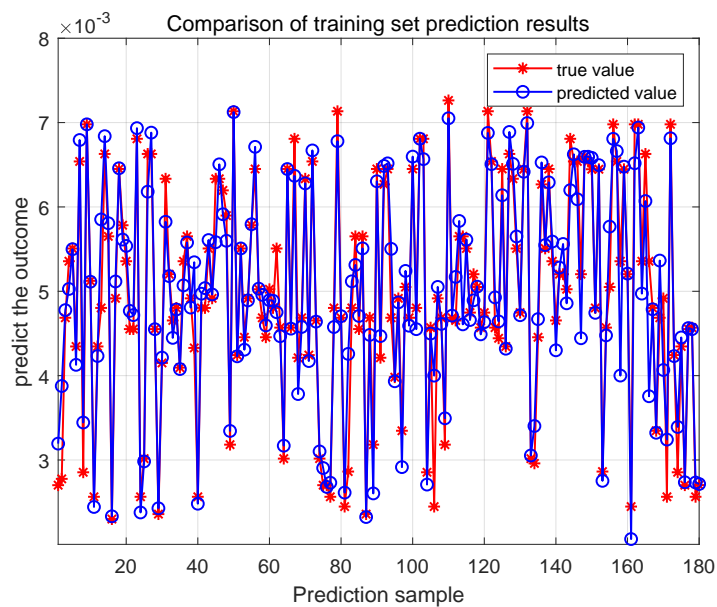


Figure 10: Neural Network Model Prediction Results 3

## 5.5 Advice to Players Based on Momentum Fluctuations

In light of the variations in momentum fluctuations observed in past matches, we propose the following steps and analytical methods for a player to improve their winning chances in new matches against different opponents:

### 1.Comprehensive Analysis of Opponents:

Perform an in-depth analysis of each opponent, considering their ranking, playing style, strengths, weaknesses, and historical match data. This analysis aids in determining the characteristics, tactical orientations, and playing styles of each opponent.

### 2.Utilization of the Momentum Model:

- Utilize our model to conduct comparative analyses of the historical "momentum" of different opponents' matches.
- Identify common patterns or situations where specific opponents are more prone to momentum changes leading to victory.
- Develop specific tactics tailored to each opponent based on their characteristics and historical momentum analysis.
- Implement different strategic approaches during various stages of the match, such as key points or service games (e.g., adopting a more aggressive approach when having favorable momentum, focusing on control and stability when facing unfavorable momentum, adjusting tactics in different situations, with priority given to enhancing personal momentum).
- Real-time monitoring of the opponent's momentum changes during the match allows for dynamic tactical adjustments based on actual game situations.
- Employ real-time data and feedback to assist players in adapting more flexibly to the opponent's performance and the dynamics of the match.
- Use our model to categorize opponents based on the characteristics of their momentum fluctuations.
- Gain a better understanding of similarities and differences between different opponents.
- Develop more specific counterstrategies tailored to each opponent category.

These suggested steps and analyses, leveraging our momentum model, aim to equip players with a strategic advantage against diverse opponents, enhancing their adaptability and responsiveness throughout matches.

## 6 Solution to Problem 4

### 6.1 The Model Prediction Performance

In Section 5.4, we presented the prediction results of the neural network model based on AHP, TOPSIS, and linear regression analysis. Below, we provide detailed information on the assessment performance of the model in predicting momentum fluctuations in multiple matches, utilizing quantitative metrics to evaluate the predictive effectiveness of the neural network.

- $R^2$  (Coefficient of Determination):  $R^2$  is used to gauge the goodness of fit of the model, comparing predicted values with predictions based solely on the mean. A higher  $R^2$ , closer to 1, signifies greater model accuracy.
- MAE (Mean Absolute Error): The MAE represents the average of absolute errors, reflecting the actual situation of prediction errors. A smaller MAE value indicates higher model accuracy.
- MBE (Mean Bias Error): The MBE measures the average difference between predicted values and true values, assessing the average deviation between predicted and actual values.

Metric	Training Set Data	Testing Set Data
$R^2$	0.9556	0.95304
MAE	0.00025564	0.00030696
MBE	$6.488 \times 10^{-11}$	$-1.6126 \times 10^{-05}$

Table 8: Evaluation Metrics of the Neural Network Model

The above indicators indicate that the model exhibits excellent predictive performance for momentum fluctuations. The  $R^2$  result is very close to 1, and both the mean absolute error and mean bias error values are quite small. Therefore, the model demonstrates outstanding capabilities in forecasting momentum fluctuations in matches.

In cases where the model performance is suboptimal, one may consider augmenting or adjusting features to enhance the model. For instance, incorporating additional match statistics such as first-serve success rate, winners-to-unforced errors ratio, and introducing new contextual features related to matches, such as historical head-to-head records between players, ranking changes, recent performance, etc., into the model construction can result in a more precise prediction of momentum changes.

### 6.2 The Generalizability of the Model

**1. For women's matches,** our model exhibits good generalizability. After testing and validation on the dataset, we found this is attributed to the high similarity in the impact of factors such as serve points won, set points won, key serve points won, double

fault rate, and break point conversion rate on momentum, as well as the mental aspect for tennis players. Gender differences have a relatively small impact on these momentum-influencing factors, hence indicating a high level of universality.

**2. Concerning tournaments**, our model maintains strong generalizability. We analyzed differences between various tournaments, including rules, playing surfaces, seasons, etc. As long as these differences are not substantial, our momentum prediction model can still be effectively applied.

**3. Regarding different court surfaces**, our model demonstrates excellent generalizability. Considering the model's adaptability to other surfaces (hard court, indoor court, etc.), fine-tuning is required to accommodate variations in game dynamics caused by factors like ball speed, bounce height, affecting serve points won, set points won, key serve points won, and serve errors, all of which influence player momentum.

**4. For different sports**, our model provides valuable insights and requires parameter adjustments for further utilization. In the case of other sports such as table tennis, an analysis of rule disparities and game dynamics peculiar to each sport is necessary. Different sports have unique factors influencing momentum, necessitating a thoughtful consideration. The model can be applied to other sports with appropriate adjustments, including redefining or adding new features to capture the specific characteristics of different sports.

## 7 Conclusion

- The Analytic Hierarchy Process (AHP) model was employed to capture the excellence of players' performance during specific time intervals in matches. Through the hierarchy analysis of five indicators - serve point percentage, match point percentage, winning serve percentage, double fault rate, and break rate - we constructed a well-performing model for rating players' performance.
- The momentum of players during matches is highly correlated with fluctuations in the game and the outcome of sets. We established a Support Vector Machine model and conducted KS tests to validate this conclusion.
- Using AHP, TOPSIS, linear regression analysis, and a neural network model, we scored momentum fluctuations and predicted changes in momentum. The key factors influencing player momentum were identified as consecutive winning points, serve quality, whether the player scored in the previous round, and Rally\_count. Considering momentum changes, we recommend players analyze opponents' historical performance and momentum changes, and adjust strategies in real-time to enhance their own momentum against different opponents.
- After training and testing in multiple matches, excellent evaluation metrics demonstrate that our model has strong performance in predicting game fluctuations and momentum changes. If the model performs suboptimally at times, adjusting

or adding features can further improve its performance. Additionally, a comprehensive analysis of the model's generalizability for other matches indicates good overall applicability.

- Considering the impact of momentum fluctuations on matches, we provide coaches with suggestions in the memorandum section from various perspectives.

## References

- [1] Bai, Z. T. Analysis of athletes' physical energy consumption and relevant indicators of match outcomes in men's singles tennis competitions. *Sports Science and Technology Bulletin*, 2022, 30(06), 61-65. DOI:10.19379/j.cnki.issn.1005-0256.2022.06.018.
- [2] Zhang, R. Prediction of tennis match outcomes and player analysis [D]. Yunnan University. 2022. DOI:10.27456/d.cnki.gyndu.2022.000807.
- [3] Xue, Z. Y. Technical statistics and analysis of Nadal's singles matches at the 2019 French Open [J]. *Sports Vision*, 2022, (02), 96-98.
- [4] Huang, Z. J. Analysis of the game theory of serve and return tactics in tennis matches between Federer and Nadal [D]. Zhejiang University. 2020. DOI:10.27461/d.cnki.gzjdx.2020.003747.
- [5] Yang, Z. M. Establishment of a predictive equation for men's singles tennis match outcomes [J]. *Journal of Beijing Sport University*, 2010, 33(04), 143-145. DOI:10.19582/j.cnki.11-3785/g8.2010.04.039.

## Appendices

### Appendix A: Meaning of Linear Regression Variables



A1	server==	1.0000		
A2	server==	2.0000		
B1	serve_no==	1.0000		
B2	serve_no==	2.0000		
C1	point_victor==	1.0000		
C2	point_victor==	2.0000		
D1	game_victor==	0.0000		
D2	game_victor==	1.0000		
D3	game_victor==	2.0000		
E1	set_victor==	0.0000		
E2	set_victor==	1.0000		
F1	p1_ace==	0.0000		
F2	p1_ace==	1.0000		
G1	p1_winner==	0.0000		
G2	p1_winner==	1.0000		
H1	winner_shot_type==	0.0000		
H2	winner_shot_type==	1.0000		
H3	winner_shot_type==	2.0000		
I1	p1_double_2ault==	0.0000		
I2	p1_double_2ault==	1.0000		
J1	p1_un2_err==	0.0000		
J2	p1_un2_err==	1.0000	O1	serve_width== 1.0000
K1	p1_net_pt==	0.0000	O2	serve_width== 2.0000
K2	p1_net_pt==	1.0000	O3	serve_width== 3.0000
L1	p1_break_pt==	0.0000	O4	serve_width== 4.0000
L2	p1_break_pt==	1.0000	O5	serve_width== 5.0000
M1	p1_break_pt_won==	0.0000	P1	serve_depth== 1.0000
M2	p1_break_pt_won==	1.0000	P2	serve_depth== 2.0000
N1	p1_break_pt_missed==	0.0000	Q1	return_depth== 1.0000
N2	p1_break_pt_missed==	1.0000	Q2	return_depth== 2.0000

Figure 11: Meaning of Linear Regression Variables

## Report on Use of AI

1. OpenAI ChatGPT (Feb 3, 2024 version, ChatGPT-3.5,)

Query1: <Tell me how the neural network model works>

Output1: <The principle of a neural network model is based on artificial neurons and hierarchical structure, simulating the functioning of the human nervous system to solve various machine learning and deep learning tasks. Here are the key principles of a neural network model:

**Artificial Neurons:** The basic unit of a neural network is the artificial neuron, also known as a node or neuron. Each neuron receives multiple inputs, calculates the weighted sum of these inputs, and applies an activation function to produce an output. This output can be passed to subsequent neurons or serve as the final output.

**Weights and Biases:** Each input connected to a neuron has an associated weight, representing its importance. Neurons calculate the weighted sum of inputs by multiplying each input by its corresponding weight and adding a bias.

**Activation Function:** The activation function determines whether a neuron activates (produces an output). Common activation functions include Sigmoid, ReLU (Rectified Linear Unit), tanh (hyperbolic tangent), etc. Activation functions introduce non-linearity, allowing neural networks to model complex non-linear relationships.

**Layers:** Neural networks typically consist of multiple layers, including the input layer, hidden layers, and output layer. The input layer receives raw data, hidden layers learn feature representations, and the output layer generates final predictions or results.

**Forward Propagation:** During forward propagation, input data passes through the network from the input layer to the output layer. Each neuron calculates its weighted sum and applies an activation function, passing the result to the next layer. This process continues until reaching the output layer.

**Backpropagation:** Backpropagation is a crucial step in training neural networks. During training, the difference between predictions and true values is computed using a loss function. The backpropagation algorithm propagates this error backward through the network, calculating gradients for each parameter using the chain rule, and then updating them to minimize the loss through gradient descent.

**Training and Learning:** Training a neural network involves iteratively performing forward and backward propagation. Through this process, the model gradually improves its ability to fit the training data by adjusting weights and biases.

**Generalization:** The goal of a neural network is to learn patterns from the training data to make accurate predictions or classifications on unseen data. A good neural network should exhibit good generalization, being able to handle new, unknown data effectively.>