Problem Chosen	2024	Team Control Number
C	MCM/ICM Summary Sheet	2422816

How Does the Momentum of Tennis Superstars Change the Course of A Match

Momentum is a phenomenon that is difficult to observe directly, but has the potential to influence the change of the game situation in real sports.

First, a review of the literature shows that momentum can be measured at both physiological and psychological levels (Ben Moss and Peter O'Donoghue, 2015). Thus, **momentum is defined as a combination of a player's performance in different situations.** Based on the given data, a player's performance in different situations of the match can be measured by the player's scoring performance and match condition. When constructing the indicator system, by measuring the receiving and serving performance, successful breaks and break point saved performance of both sides, it can make the indicator measurement system more objective.

By constructing the evaluation index system through Independence Weight Coefficient Method and visualizing the players' momentum by game, it can be found that **the direction of the players' momentum is consistent with the direction of the game situation**. From this, it is firstly determined that there is a correlation between the momentum and the game situation. Furthermore, a one-way ANOVA is conducted to analyze the relationship between players' momentum and their successive successes and swings. The test of significance and effect size analysis revealed that momentum is indeed related to it, indicating that momentum plays a significant role in the match.

As a result of the above research, the indicators related to momentum are selected, and the swing prediction model is constructed based on the Decision Tree algorithm. The prediction results of the model and the analysis of the importance of the features show that the momentum can be effectively captured by tracking the direction of the momentum, and the player's performance in handling the break point is a key factor in the situation shift.

In addition, we generalize the model concerning both different opponents and different matches. The results of the optimized model show that the model predicts better by capturing the details of match performance. The evaluation results show that the model has a **prediction accuracy of 0.931 and an F1 score of 0.934**, indicating that **the model can correctly predict situation reversals while capturing a wide range of situation fluctuation points.**

Finally, we have prepared a memo that explains the meaning of momentum and advises coaches and players on how to deal with changes in the game situation.

Keywords: Momentum; Swing; Independence Weight Coefficient Method; Decision Tree;

Content

1 Introduction

1.1 Background

During the 2023 Wimbledon Gentlemen's final, Carlos Alcaraz, a rising star from Spain, defeated the Grand Slam player Novak Djokovic in a remarkable battle. The swings in the game that occurred when one player seemed to take the lead are often attributed to 'momentum', a difficult-to-measure factor in a match.

1.2 Problem Restatement

Based on the background information and provided data, the following problems need to be addressed:

- **Problem I:** Develop a model to capture the flow of play, identify which player is performing better at a given time in the match, quantify the degree of their performance advantage, and visualize the match flow to illustrate the findings.
- **Problem II:** Define the meaning of “momentum” and assess the impact of “momentum” on swings in play and runs of success by one player.
- **Problem III: First**, develop the model to predict the swings in the match and analyze the most related factors. **Second**, compare the differential in past match “momentum” swings and advise a player on his further new matches against different opponents.
- **Problem IV:** Test the model on the other matches and report its predictive reliability on swings in the match, generalize it to other matches and identify factors that might need to be included in the future models.
- **Problem V:** Prepare a one-to-two-page memo summarizing the results of the role of “momentum” for coaches and advice on preparing players to respond to events that impact the flow of play during a tennis match.

1.3 Our Work

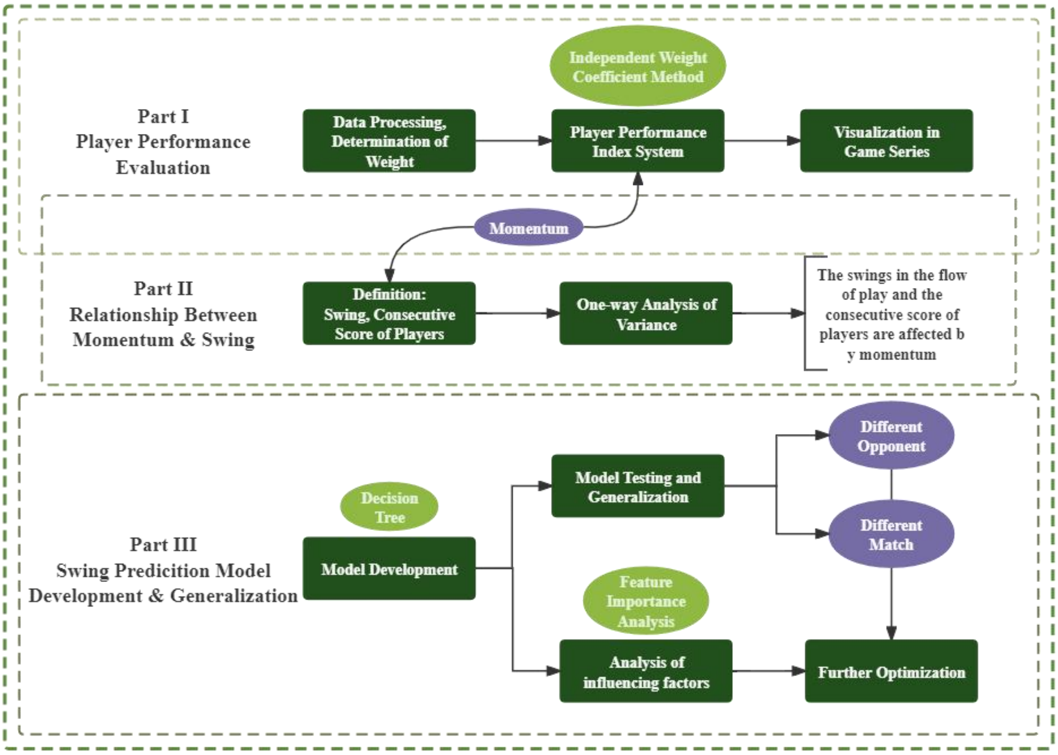

Figure 1-1 *Structure of Our Paper*

2 Assumptions and Justifications

To simplify the problem and make it more amenable to analysis and solution, consider introducing the following assumptions:

- Possible tactical shifts during a tennis match are ignored due to the variety and complexity of match tactics, which are difficult to quantify.
- The impact of serve speed, serve direction, return depth, and untouchable stroke category on a player's momentum are not considered. This approach enables the examination of other variables that may influence the match outcome.
- Assumes weather and course conditions are relatively stable and will not interfere with play.

3 Data Processing

3.1 Default Data Processing

Data pre-processing can improve the effectiveness and efficiency of data calculation, and ensure that the results of data analysis have greater accuracy. In order to eliminate the influence of the different units on the variables, we need to standardize the data. Therefore, for the original data package "Wimbledon_featured_matches.csv", according to the situation of each game, we have done the following processing:

First, the five-set 36-game data and five-set 47-game data of the quarterfinals and the five-set 46-game data of the final match were selected for analysis. The screened quarterfinal data were used for generalization, and the screened final match data were used for the basic evaluation model as well as the prediction model. According to the ratio of 7:3, the last 14 games of the final match data of the final match were selected for the test of the prediction effect in terms of the number of games.

Next, the screened raw data were subjected to data cleaning to check whether there were missing values and outliers in the data.

Next, in order to assess the performance of the players at a certain time point, metrics were transformed. Each game is taken as a time unit, the number of total games is counted, and the number of games in the plate is made to correspond to the value of the number of total games. Considering that the game score of the tie-breaker is calculated differently from the game score of the ordinary games, the substitution process is performed for the score data of the two players in each game, so that 15=1, 30=2, 40=3, and AD=4.

Then, all the screened indicators in ***Table 3-1*** are standardized with the following formula of the range method.

$$x_{ij}^{'} = \frac{x_{ij} - min\,(x_j)}{\max\left(x_j\right) - \min\,(x_j)}$$

Finally, the preprocessed dataset is partitioned into a training set, a validation set, and a test set so that different subsets of data can be used in training and evaluating the model.

3.2 Index Calculation and Definition

In the selection of data indicators, in order to quantify the momentum, considering the fact that the reversal of the situation in tennis matches was affected by multiple factors, we referred to the relevant literature on the influence factors of tennis matches.

Zhang Jie selected eighteen technical indicators, including the number of aces, the number of double faults, the percentage of first-serve points, the percentage of first-serve points, the percentage of first-serve points, the percentage of break points saved, the percentage of first-serve points, and so on, to study the techniques and tactics of Novak Djokovic's Grand Slam matches of the season of 2021 from the service link, the return link, and the scoring link. Luo Yijie selected nine technical indicators such as winners, aces, unforced errors, first-serve win rate, second-serve win rate, serve-receive win rate, etc. to study the technical and tactical analysis of the "key points" of the world's outstanding men's tennis players in the hard court tournaments. Fu Rao analyzed 239 matches of the Australian Open in 2005-2006, and obtained three indicators with high influence on the winning and losing of the matches, which are in the following order: serve-return percentage, first-serve percentage and first-serve success rate.

In summary, based on the processed data, the 12 indicators in ***Table 3-1*** were calculated to be studied.

Tabel 3- 1 *Abbreviation and Definition*

Abbreviation	Definition
p_score_no1	Times of scores on first serve
p_score_no2	Times of scores on second serve
winner_shot	Times of wining shot
net_pt_won	Times of scores while at the net
break_pt_handle	Times of chances to win
max_gap_distance_run	The gap between the maximum distance run and the minimum distance run
unf_err	Times of unforced errors
whether_bpt	Whether or not there is a chance to win the game
game_continuous_score	Times of continuous scores
gap_game_continuous_score	The gap between the number of continuous scores

4 Dynamic Analysis of Player Performance in Game Situation

4.1 Overview

Considering the high winning rate of tennis serve players and referring to the authoritative ATP website, we selected seven indicators in ***Table 3-1*** to quantify the performance of players in terms of both match scoring and player status. In order to assess which player performs better at a certain time in the match and quantify the degree of his performance, we use the independent weighting coefficient method to objectively assign weights to the seven selected indicators, and assign lower weights to the indicators with higher complex correlation coefficients with the other indicators, avoiding the influence of subjective factors on the distribution of weights, and finally obtaining the composite score as the quantified player performance score.

4.2 Player Performance Evaluation Model

4.2.1 Correlation of Momentum and Player Performance

For tennis players, match performance can be measured in a combination of scoring ability, serve and serve receive performance, baseline and net performance, and durability and fitness, such as the number of points scored on game-winning shots and unforced errors.

Momentum i.e. the momentum or advantage gained by a side as a result of a series of events during a match. For the purposes of this paper, given the need to quantify momentum, we define momentum as the sum of factors that may cause a turnaround at a critical point in a match.

In a tennis match, a player's performance often directly reverses the situation at a critical moment, which is due to the nature of the tennis sport and the variable factors in the match. Based on the definition of momentum in this paper, the performance of both players can be visualized as the quantified momentum value. Considering the high winning rate of tennis serve players, we introduce the key scoring factors for the quantization of momentum, and reflect the players' performance in the serve segment as well as the return segment of the game with the momentum value.

Referring to the ATP authoritative website, we finally select the relevant indexes to measure the performance of the players in terms of the scoring situation and the players' status for the final court data. For the scoring situation, we selected the number of first serve reception points, the number of second serve reception points, the number of winners in the set, the number of successful breaks in the set, and the number of points in front of the net as the constituents of the player's performance score; for the player's status, we selected the maximum distance difference in the set and the number of unforced errors in the set as the constituents of the player's performance score. The seven indicators are as follows. The seven indicators are as follows:

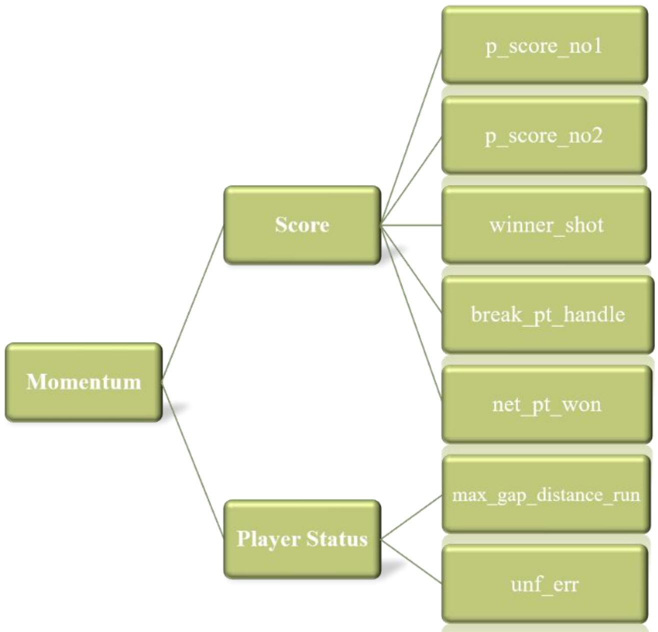

Figure 4-1 *Momentum indicator system*

The number of first serve reception points, the number of second serve reception points, the number of ace points in the set, the number of successful break points in the set, and the number of points at the net indicate that the player has a high efficiency of serve reception in the set, and the number of active points indicates that the player is able to hit the opponent's difficult path on the offensive end, and is able to grasp more initiative on serve, and the subsequent attack is more fluent and the defense is strong, so it is used as a positive indicator of the player's performance. Therefore, it is a positive indicator of the player's performance. The difference between the maximum running distance of the game visualizes the physical exhaustion of the player, and the high number of unforced errors indicates that the player is not in good condition, so it is used as a negative indicator of the player's performance.

4.2.2 Quantification of Momentum

After normalizing the data of the above screened players' performance related indexes, the independence weighting coefficient method is used to objectively assign ω_j, and the resulting composite score is taken as the players' performance score.

Step1: Calculate the compound correlation coefficients between the performance indicators of each player. Here, the seven momentum indicators in this paper are set as X1, X2, X3, X4, X5, X6 and X7.

$$R_j = \frac{\sum_{j=1}^{m}\left(X_j - \overline{X}\right)\left(\widetilde{X} - \overline{X}\right)}{\sqrt{\sum_{j=1}^{m}\left(X_j - \overline{X}\right)^2\left(\widetilde{X} - \overline{X}\right)^2}} (j = 1,2, \dots, 7)$$

If the compound correlation coefficient between the player performance indicator X_j and the other performance indicators is larger, the stronger the covariance with the other performance indicators, the less weight the indicator has. Below is the correlation chart of the player performance indicators:

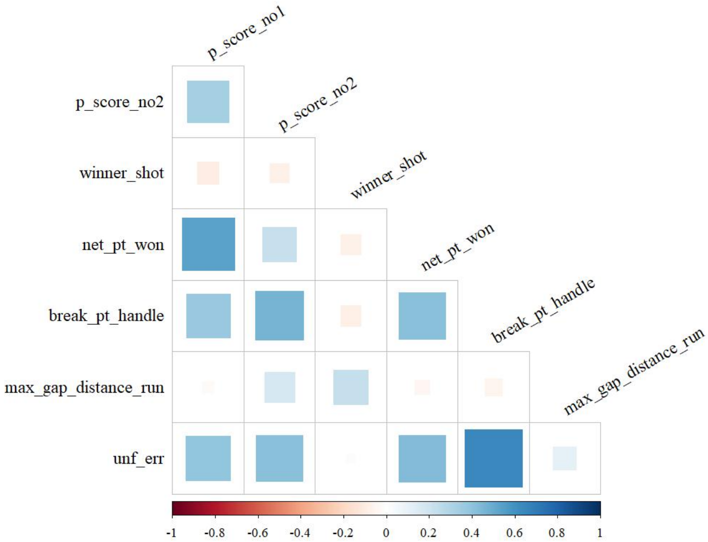

Figure 4-2 *Heat correlation plots for player performance indicators*

Step2: Select the inverse of the complex correlation coefficient and calculate the final weights after normalization.

$$\omega_j = \frac{\frac{1}{R_j}}{\sum_{j=1}^{m}\frac{1}{R_j}}$$

We obtain the following assignments for each momentum indicator after objective assignment through the independence weight coefficient method:

Table 4-1 *the Weigh of Independence Weight Coefficient Method*

Var.	Complex Correlation Coefficient R	Reciprocal of the Complex Correlation Coefficient 1/R	weigh(%)
p_score_no1	0.595	1.680	11.603
p_score_no2	0.552	1.813	12.517
winner_shot	0.267	3.752	25.908
net_pt_won	0.604	1.655	11.427
break_pt_handle	0.709	1.411	9.739
max_gap_distance_run	0.366	2.730	18.850
unf_err	0.694	1.442	9.956

Step3: Calculate the composite score of the Momentum Indicator as a quantification of the player's performance:

$$\text{Momentum} = 11.603\text{p\_score\_no1} + 12.517\text{p\_score\_no2} + 25.908\text{winner\_shot} + 11.427\text{net\_pt\_won} + 9.739\text{break\_pt\_handle} + 18.85\text{max\_gap\_distance\_run} + 9.956\text{unf\_err}$$

According to the defined assignment formula, the player performance scores of the two players for each set are obtained and Figure1 is drawn to visually compare the match performance of the two players.

4.3 Visualization of Player Performance

Table 4-2 shows the first seven games of the first set of the Wimbledon men's singles final as an example of data showing the combined scores of the two players in the first seven games of the first set as well as the winners of each game.

Table 4-2 *Scoring in the first set (first seven games)*

Game	1	2	3	4	5	6	7
C.A.	0.136	0.123	0.279	0.172	0.080	**0.299**	0.058
N.D.	**0.261**	**0.157**	**0.441**	**0.189**	**0.192**	0.165	**0.275**
game_victor	N.D.	N.D.	N.D.	N.D.	N.D.	C.A.	N.D.

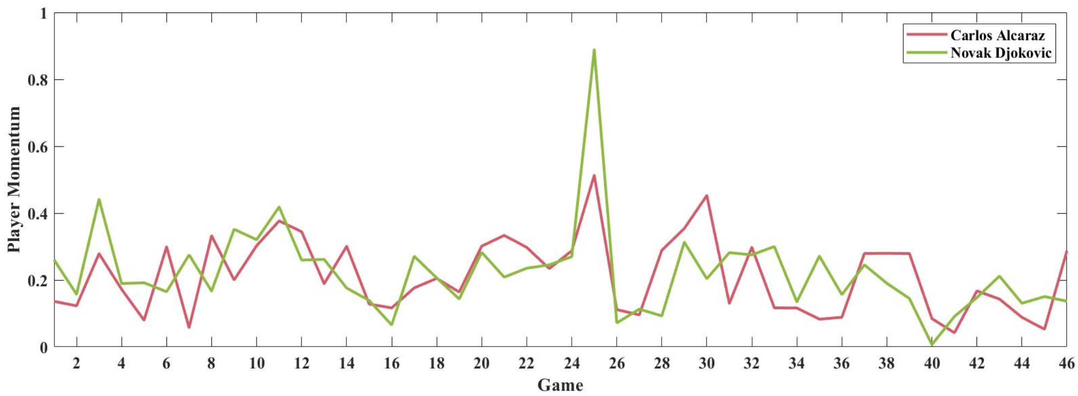

Figure 4-3 *the Tendency of "Momentum"*

We can visually compare the match performance of the two players in terms of momentum size through ***Figure 4-3***. In the first set of six games, Djokovic's performance points are significantly larger than Alcaraz's performance points that is, Djokovic's game performance is significantly better than Alcaraz, dominant, in the actual game Djokovic also finally won 6-1 easily.

In the second set, the situation is more complicated, to Alcaraz serve the first set 40-15 victory to carry out a new round of fierce battle, at this time Alcaraz performance score is higher than Djokovic; the third set Djokovic to AD-40 to get a break point, however, Alcaraz retained their own serve, which marks his performance score may be a favorable change. In the actual match it ended up being Alcaraz who won 7-6 in a steal of seven.

In the third set, Djokovic trailed 40-AD in the first game on his own serve; the fifth game was the most intense, with both players peaking at this point in the match, however, Alcaraz broke to win the set when the score was 40-40. Considering the physical exertion of the two players at this time also reached a high level, which set the stage for the subsequent reversal of the situation. In the actual match, Alcaraz won 6-1.

The fourth set of the match to Alcaras serve the first set of 40-30 victory to start, the first three sets of Alcaras performance points are higher than the performance of Djokovic 0.1 or so, but after the third set of Alcaras momentum is gradually downward trend, which also makes Djokovic in the late stages of the plate are easy to win the score, in the actual game finally Djokovic achieved a break of serve points to 6-3 40-15 to obtain the In the actual match, Djokovic broke to win the set 6-3 40-15.

In the fifth set, Alcaraz lost the first game on his own serve 40-AD, and the situation didn't seem favorable for Alcaraz at the beginning. However, Alcaraz was able to increase his points and in the third game he broke to win the set at 30-30. Both players fought hard in the later stages of the set, but Alcaraz finally took control and won 6-4.

4.4 Impact Analysis of Momentum on the Game Situation

4.4.1 Definition of Swing and Continuous Success Scenario

In tennis matches, the performance of a swing is unique. It usually occurs during a specific period, either through a set point, set point, or match point. Therefore, breakpoints can be used as a measure of swing. At the same time, given the dynamic nature of "swing", the accumulation of breakpoints throughout a game can better reflect the dynamics of the match: the more breakpoints accumulated, the more advantage a player has over his opponent's serve. This will affect the entire match. Therefore, the number of breakpoints per game is finally chosen to measure it. According to the question, a player's consecutive success can be understood as the player's ability to score consecutive points. So it can be reflected by the consecutive points earned by the player in each game.

4.4.2 Impact Analysis

In tennis matches, different players have different levels of serve receive, match tactics, and physical condition. On the one hand, the discrepancies in the performance of different abilities and tactics in the match will be reflected in the performance of the players. On the other hand, the physical exertion of the game on the players as the game time goes on will also have an impact on the situation. This leads to the hypothesis that the fluctuation of the situation and the successive successes of the players in tennis matches are not random, but are the result of the action of various reasons.

The hypotheses are further analyzed in relation to the definition of momentum in the previous section: momentum is a subjective reflection of a player's scoring performance and immediate status, while situation fluctuations and the number of consecutive successes of a player are objective reflections of the player's performance.

To summarize, the differences in momentum under situation fluctuation and consecutive successes are discussed below:

- Continuing with the example of the 2023 Wimbledon Open final, the number of breakpoints per set and the consecutive successive points data of the players are extracted. Let μ_i and θ_i be the mean values of momentum under different situation fluctuations and different continuous successes, respectively, and make the following assumptions:

$$\mathbf{H_{a0}}: \mu_1 = \mu_2 = \dots = \mu_i$$
$$\mathbf{H_{a0}}: \theta_1 = \theta_2 = \dots = \theta_i$$

- One-way ANOVA was performed. And F-statistics were calculated. The results were as follows:

***Table 4-3** ANOVA Result*

Var.	F-statistic	Sig.
Momentum[1]	18.429	0.000***
Momentum[2]	14.063	0.000***

In ***Table 4-*** , Momentum1 and Momentum2 represent the results of the ANOVA test under situation fluctuations and successive successes, respectively. The results show that there is a significant difference in momentum in the case of fluctuating situations and different continuous successes.

- The differences were further analyzed and the effect sizes were calculated as follows:

***Table 4-4** ANOVA Effect Size: Swings in Play*

		Point Estimate	95% Confidence Interval	
			Lower	Upper
momentum	Eta-squared	0.393	0.212	0.499
	Epsilon-squared	0.365	0.176	0.476
	Omega-squared Fixed-effect	0.362	0.174	0.473
	Omega-squared Random-effect	0.124	0.050	0.183

***Table 4-5** ANOVA Effect Size: Continuous Success Scenario*

		Point Estimate	95% Confidence Interval	
			Lower	Upper
momentum	Eta-squared	0.565	0.391	0.641
	Epsilon-squared	0.535	0.348	0.615
	Omega-squared Fixed-effect	0.532	0.346	0.613
	Omega-squared Random-effect	0.159	0.081	0.209

The two sets of effect sizes are analyzed separately. By looking at the values of η^2 and ω^2 for the two groups, it can be found that the corresponding values are greater than 0.2 in ***Table 4-*** and greater than 0.5 in ***Table 4-*** . This can indicate that momentum has a small effect on the change of situation fluctuations and a large effect on the continuous successful performances of the players.

5 Swings Prediction in the Flow of Match

5.1 Swing Prediction Based on Decision Tree

In predicting when the game will swing from one side to the other, it is necessary to analyze the players' skill level, physical and mental state, scoring, and other key characteristics in a comprehensive manner. In this way, the whole game can be more comprehensively portrayed and the fluctuation of the game situation can be captured in a more detailed way. In analyzing situation fluctuations, there is a clear definition of what is meant by "the situation switches from one side of the match to the other", i.e., the number of break points accumulated in each game.

Based on the previous arguments for momentum fluctuations, it is possible to draw the preliminary conclusion that swings can be affected by momentum and that the succession of points scored by different players can also reflect the situation in the match. Therefore, it is now attempted to incorporate momentum-related metrics and combine them with the difference in consecutive points scored in each game to construct a preliminary prediction model for situation change. This model reflects the dynamics of the match situation by predicting the accumulation of break points.

In terms of model selection, a decision tree is considered to categorize and predict the cumulative number of break points. Decision tree is a machine learning algorithm that can be used for classification prediction, which is able to make layer-by-layer determination and decision through data features. Its handling of mixed data types and non-linear relationships is more able to reasonably capture the factors that transform the situation, and the prediction is more reliable.

Based on the 2023 Wimbledon men's singles final 1701 data, the process of constructing the situation fluctuation prediction model is as follows:

Step 1: Slice the data set, 70% for the training set, 30% for the test set.

Step 2: Based on the processing ability of continuous and discrete attributes of the data, the Gini index is chosen to divide the attributes of the dataset:

$$gini(T) = 1 - \sum_{j=1}^{n} p_j^2$$

where p_j is the probability of occurrence of class j.

Step 3: Let the minimum number of internal nodes to be categorized be 2 and the minimum number of leaf nodes to be sampled be 1. Establish the decision tree classification case as follows:

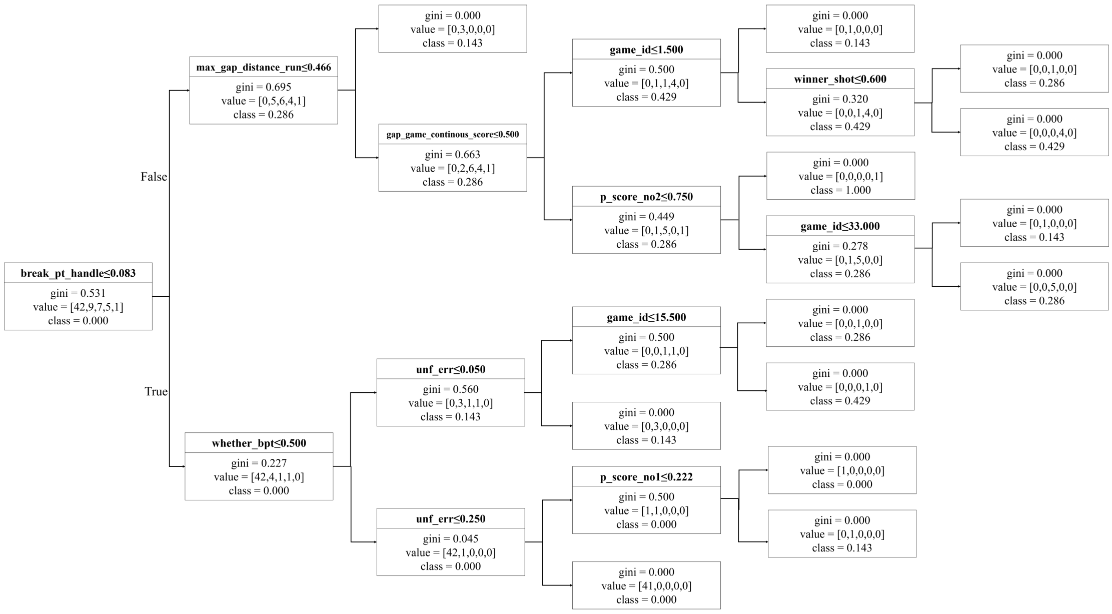

Figure 5- 1 *Decision Tree Classification Diagram*

Step 4: The constructed model was cross-validated with five folds and the performance metrics of the five tests were averaged to evaluate the overall performance of the model. The results of the model evaluation are as follows:

Table 5- 2 *Model Evaluation Result*

	Accuracy	Recall	Precision	F1
Trainset	1	1	1	1
Cross-Validationset	0.705	0.705	0.722	0.71
Testset	0.786	0.786	0.773	0.765

According to the evaluation results, the model performs well overall. The accuracy and recall are 0.786, indicating that the model is able to correctly classify most of the samples and categorize them into the correct classification. The precision rate is slightly lower than the accuracy rate, with some misclassification. The F1 score is 0.765, indicating that the model performs well in balancing the accuracy and coverage of prediction. In the actual prediction of fluctuations in the tennis situation, further adjustments should be made to incorporate the characteristic contributions of each indicator.

5.2 Insights into Factors Influencing Swings

Feature importance refers to a measure of how much each feature contributes or is important to splitting the data and making decisions when constructing a decision tree model. It is calculated by accumulating the contribution of each feature at the split point in the decision tree. The resulting feature importance is calculated as follows:

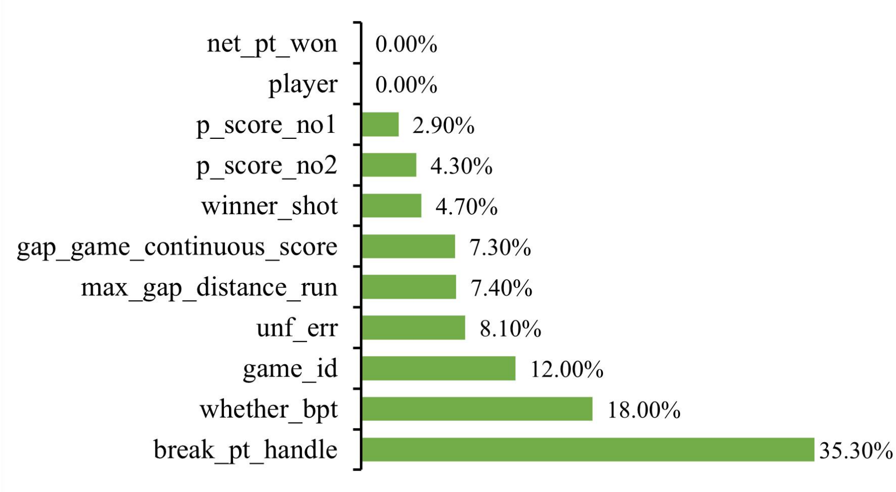

Figure 5- 2 *Result of Feature Importance*

In ***Figure 5-*** , there are differences in the contribution of each indicator feature to the decision tree classification prediction. Since the judgment on the prediction of situation fluctuation comes from the contribution of each feature at the split point, the magnitude of the feature's importance also determines its importance for situation reversal. This is analyzed below in terms of the magnitude of contribution:

- Firstly, the level of players' handling of break points is the most important factor influencing the situation from one side to the other, with a share of 35.3%. This is followed by the number of break points and the number of sets played.
- Next are the number of unforced errors, the maximum running distance difference per game, and the consecutive point difference, respectively. This suggests that a player's serve-receive status and physical exertion also have an impact on whether or not the situation turns around.
- In addition, the number of winning strokes, performance in receiving first and second serves also had a small effect on the turnaround.
- It is worth noting that performance in winning shots at the net has no effect on turnover. This can be explained by the randomness of the luck factor of winning at the net.

5.3 Comparison of Momentum Differences

5.3.1 Comparative Analysis of Momentum Differences

In order to compare the momentum difference of a player in matches against different players, the data of the semifinal and quarterfinal Djokovic matches are introduced into the decision tree model constructed in the previous section, and the following two matches are obtained to obtain the rankings of the top five feature contributors for each factor:

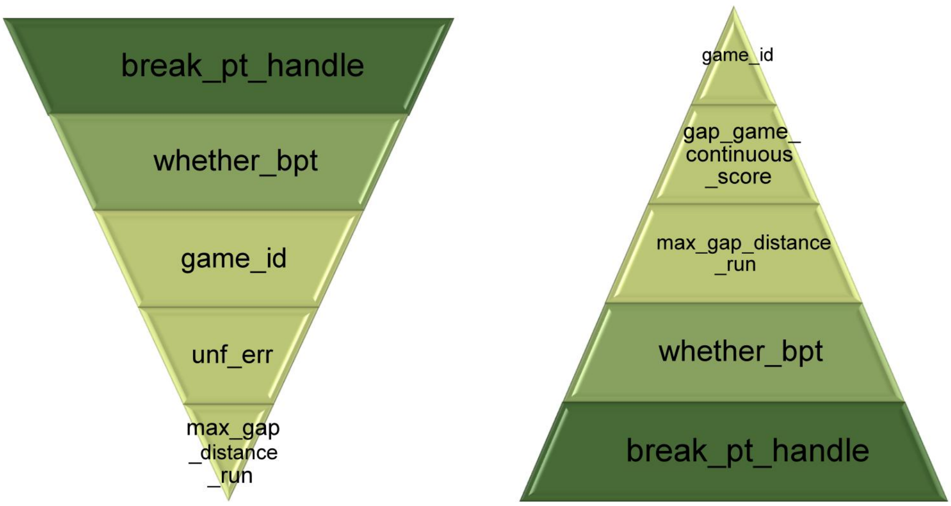

Figure 5-4 *Top five factors for feature contribution*

There are differences in the contribution of each indicator feature to the decision tree classification predictions, and the magnitude of the feature's importance also determines its importance for situation reversal. In both the quarterfinals and semifinals, the most relevant factor for situation fluctuation was the player's ability to handle breaks of serve, followed by whether or not there was a break point, which is consistent with the results of the model building analysis above. Considering the difference in momentum between the two matches, the characteristic contribution of the same variable also differs. The difference in momentum is visualized in the factors that rank differently in the characteristic contributions of the two, such as the first serve reception score, the number of winners, and the difference in maximum running distance.

5.3.2 Game Tactics Advice for Players

Based on the difference in momentum during a match, this paper makes the following recommendations for the match strategy of one player versus different players:

- In order to increase the percentage of first serve points, a player should focus on improving the speed and angle of the serve, while maintaining consistency and accuracy of the serve. Receiving first serves requires players to have quick reaction time and good technical skills. Players can improve their anticipation of the serve and their hitting skills, as well as adopt appropriate tactics to deal with the opponent's serve.
- Players can also reduce the maximum distance difference by optimizing their running routes and hitting patterns. Improve the ability to cope with different situations by practicing different hitting tactics and styles. Improve feel training to reduce the number of unforced errors.
- It is vital to keep a cool head whether you are in the lead or behind. Adjust

your tactics at the right time according to the changing situation of the game. Maintain a high level of concentration at all times during the match, especially at critical moments. Do not allow external factors to interfere with your concentration, and do not be influenced by your opponent's performance.

- During the game, make good use of the field and environment to create opportunities. For example, use the uneven nature of the field, wind direction and other factors to control the bounce and direction of the ball.

6 Model Testing and Application

6.1 Predictive Reliability Test

The reliability test of model prediction can assess the model's ability to generalize beyond the training data and ensure the stability and credibility of its prediction in different situations. In order to test the generalization ability of the prediction model, based on the constructed prediction model, we predicted the turn of the situation of the quarter-final match between Novak Djokovic and Andrey Rublev, and compared the real value with the predicted value. The prediction results are plotted as follows:

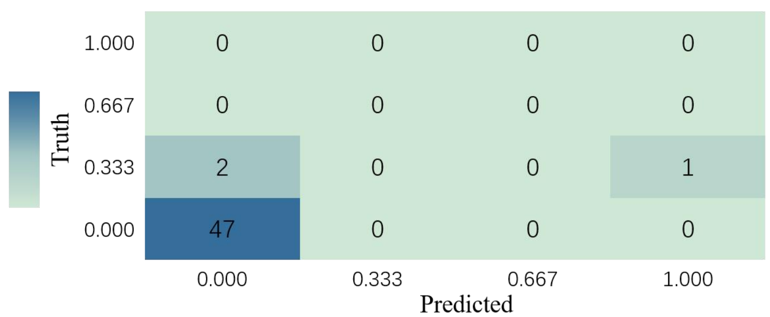

Figure 6- 1 *Confusion Matrix Heatmap*

Figure 6- 1 shows the confusion matrix between the true and predicted values. The results show that the model is not very effective in predicting the situation changes in the course of the game. Combined with the previous analysis of the differences in the "momentum" of different players in the case of Novak Djokovic, it can be seen that due to the differences in the strength and condition of the opponents, the impact of different factors on the reversal of the situation is different in each game. Therefore, we consider to use the data of the first half of the game to predict the second half of the game, and the prediction results of the test set are shown in ***Figure 6- 2***.

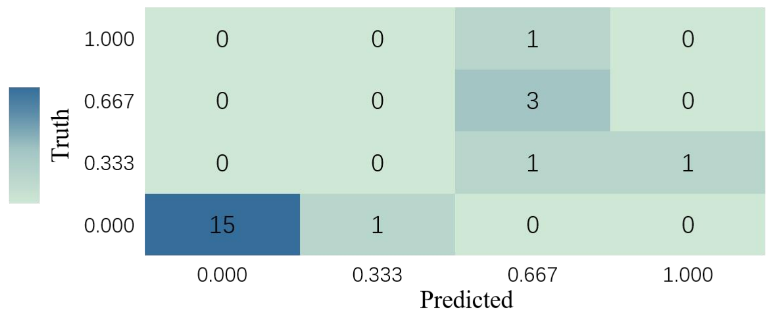

Figure 6- 2 *Confusion Matrix Heatmap (Optimized)*

It can be seen that by re-matching the characteristic importance of the influencing factors of this game, the model's ability to capture points with a large degree of turnaround in the situation is significantly improved. The specific fitting results of the model are as follows:

Table 6- 1 *Model Evaluation Result*

	Accuracy	Recall	Precision	F1
Trainset	1	1	1	1
Cross-Validationset	0.74	0.74	0.733	0.718
Testset	0.818	0.818	0.809	0.806

From ***Table 6- 2***, it can be seen that the overall performance of the model is significantly improved, with a prediction accuracy and recall of 0.818. The F1 score is 0.806, which is an improvement on the original model in terms of the balance between accuracy and coverage. It can be shown that the model is able to capture most of the situation rotations and predicts them well.

6.2 Further Application in Other Tennis Games

In order to further assess the model's ability to be applied, the model was generalized to the prediction of fluctuating tennis tournament situations of other players for evaluation. In order to ensure the reliability of the data, the predicted matches were filtered with the WTA world rankings. The final prediction is based on the match between Daniil Medvedev and Christopher Eubanks in the second match of the fifth round, and the prediction results are as follows:

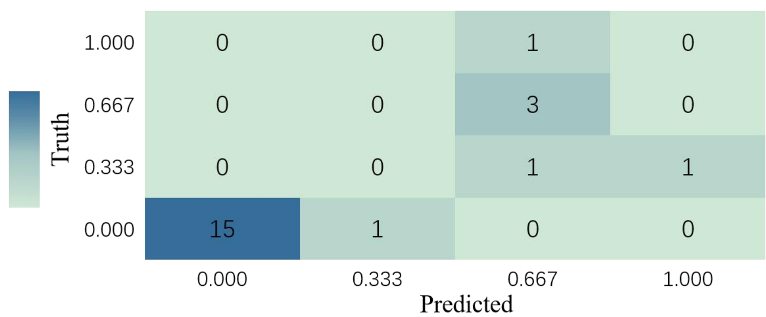

Figure 6- 3 *Confusion Matrix Heatmap (Optimized)*

Table 6- 3 *Model Evaluation Result*

Accuracy	Recall	Precision	F1
0.931	0.931	0.959	0.934

An analysis combining ***Table 6- 3*** and ***Figure 6- 3*** reveals that:

- The model is able to accurately predict most of the situation turning points with a prediction accuracy of 0.931.
- Although there are still cases of misclassification, it is more sensitive in recognizing the fluctuating situations of the situation.
- The F1 score of 0.934 indicates that the model is able to correctly predict situation reversal situations while capturing a wide range of situation fluctuation points.

6.3 Further Optimization

By exploring the model's prediction performance against different opponents and different matches, and analyzing it with the actual match situation, the conclusions of the model study and suggestions for future modifications are as follows:

- **Model applicability conditions:** different opponents in different tournaments have different opponents, and there are subtle changes in the factors affecting the match situation, so the generalization of the model needs to be reevaluated and analyzed based on the strengths of different opponents.
- **Feature Engineering Improvement:** First of all, in tennis, the environments of different tournaments are different, which has an impact on the player's play. Subsequently, further introduction of venue and weather difference indicators can be considered. Secondly, the physiological state of the players also has some influence on their playing status, so their heart rate, body temperature and other indicators can also be included in the model.

7 Strengthens and Weaknesses

7.1 Strengths

- Referring to a large number of related literature in the REFERENCE and ATP authoritative website to select indicators, we took the factor of high winning rate of serving into consideration and selected seven indicators for the two players in the final court from the aspects of scoring situation and player's status. Introducing the key scoring factor and reflecting the performance of the serve side, the index system is more comprehensive and it can accurately assess the performance of the players at the critical moment.
- The algorithm is efficient and makes the solution logic more perfect. The independent weight coefficient method is used to objectively assign weights

to the momentum-related indicators for each game, and the momentum is quantified by calculating the composite score, while avoiding the influence of subjective factors on the weight assignment. We choose the decision tree model, which has the advantages of robustness as well as being easy to understand and implement.

- Our model effectively answered all of the questions we needed to solve initially. It was fast and could handle large quantities of momentum data, but also had the flexibility we desired. The fundamental strengths of our model are the characteristics of convenience and strong applicability. The results of the prediction model are similar to the evaluation results of famous players and media, which is convincing.

7.2 Weaknesses

- The independence weight factor method is an objective assignment based on limited data, and the reasonableness of the assignment often has to be considered in the context of multiple and complex factors in reality.
- The use of the bureau as the time unit is somewhat limited in terms of indicator selection and model choice
- A more abundant data resource can guarantee a better result in our models. The results of the model depend heavily on the given statistics. Therefore, if a large amount of information is lost, the results may not be reliable.

8 Memo

To: Tennis coach
From: Team#2422816
DATE: February 5, 2024
Subject: Momentum in tennis

Dear Sir,

By your request, we establish a customized model to quantize momentum and determine when the flow of play is about to change from favoring one player to the other.

Analyzing the result predicted, we have many significant discoveries and recommendations, which can help you significantly improve the efficiency of tactical adjustments in the face of changing situations.

i) We define momentum as the sum of factors that may cause a turnaround at a critical moment in a match. We use the independent weighting coefficient method to assign weights to obtain a composite score as the quantified momentum, and the defined formula for calculating the momentum per game is as follows:

$$\begin{aligned}\text{Momentum} = {} & 11.603\text{p\_score\_no1} + 12.517\text{p\_score\_no2} + 25.908\text{winner\_shot} \\ & + 11.427\text{net\_pt\_won} + 9.739\text{break\_pt\_handle} \\ & + 18.85\text{max\_gap\_distance\_run} + 9.956\text{unf\_err}\end{aligned}$$

ii) The hypothesis is formulated that in tennis, situation fluctuations and successive successes of players are not random, but are the result of the action of various causes. A one-way ANOVA was conducted to explore the differences in momentum under situation fluctuations and successive successes, and we found significant differences in momentum under situation fluctuations and different successive successes. It is worth noting that we quantified the effect size of the difference. By analyzing the two sets of effect sizes, it can be concluded that momentum has a small effect on the change in situation fluctuations and a large effect on the successive successful performances of the players.

iii) To gain a deeper understanding of the role of momentum, we use a decision tree algorithm to predict when the flow of play is about to change from favoring one player to the other. By analyzing the contribution of features, the key factors of momentum include the number of points scored on the first serve, the player's own experience, the number of points scored on the second serve, and the ability to handle breakpoints. The results of the constructed decision tree model is shown below:

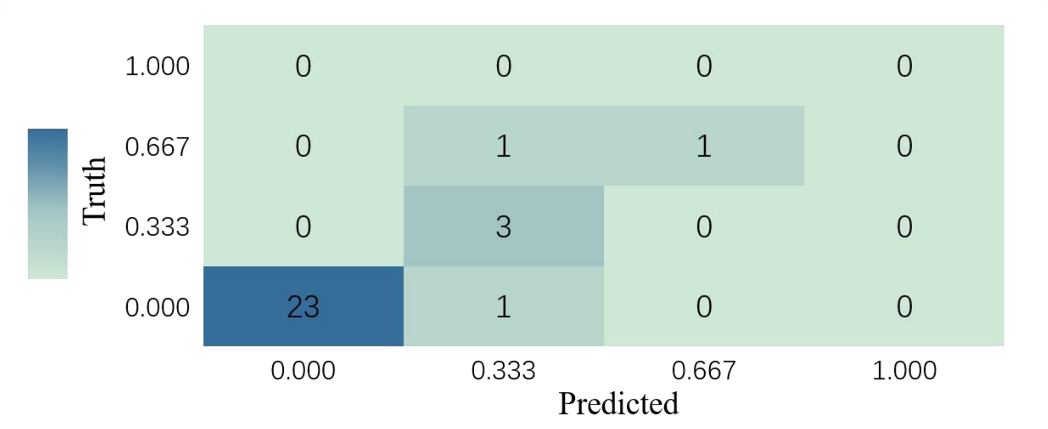

Figure 9-1 The results of the constructed decision tree model

iv) **The effect of Momentum:** the number of first-serve reception points, the number of second-serve reception points, the number of ace points in the set, the number of successful break points in the set, and the number of points in front of the net, the player has a high efficiency of serve reception in the set, and he is able to take the initiative in the service game. The maximum distance difference in this set visualizes the physical exertion of the player, and the high number of unforced errors indicates that the player is not in good condition.

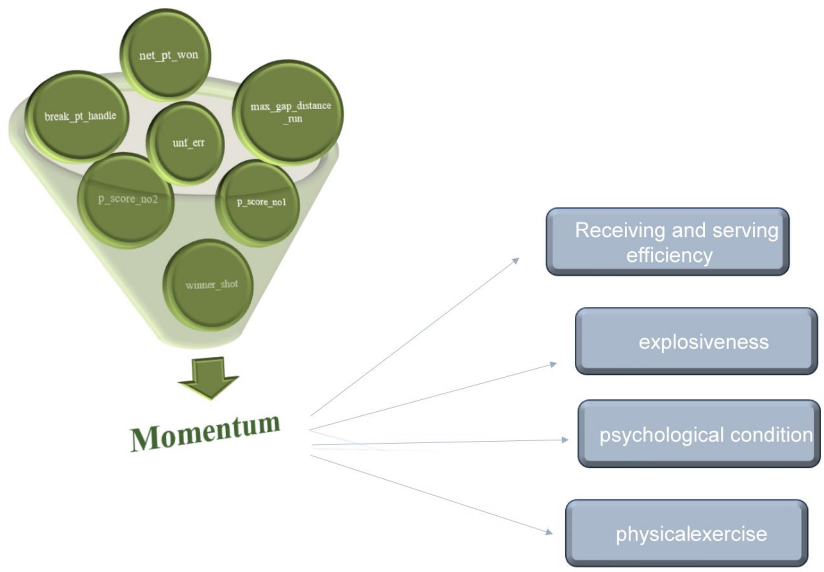

Figure 9-2 The effect of Momentum

v) Substituting different numbers of sets and players into our constructed prediction model, the model predicts the in-game fluctuations better.

As to how to respond to events that affect tennis events that affect the course of a match, after looking up relevant information, we provide several approaches as follows.

i) To increase the percentage of first-serve points, players should focus on improving the speed and angle of the serve, while maintaining the stability and accuracy of the serve. During training, players can increase their serve by practicing different serving styles and tactics.

ii) In order to seize the breakpoint opportunities, players should focus on improving their breaking skills and anticipation. In training, players can practice by simulating match scenarios to improve the success rate and accuracy of their breaking skills.

iii) Players should focus on improving their endurance and physical fitness level, and strengthen the training of their leg muscles. In addition, players can optimize their running routes and hitting methods to reduce the maximum running distance difference. Players should adjust their physical exertion according to the progress of the match to ensure that they have enough energy to cope with the critical moments.

iv) Whether you are in the lead or behind, it is vital to keep a cool head. Adjust your tactics according to the changing situation of the game. Maintain a high level of concentration at all times during the match, especially at critical moments. Do not allow external factors to interfere with your concentration. In addition, don't be influenced by your opponent's performance.

v) During the game, take advantage of the field and environment to create opportunities. For example, use the uneven nature of the field, wind direction, and other factors to control the bounce and direction of the ball.

The suggestions above will help you better utilize momentum in games. We sincerely hope that our study will benefit you.

Yours sincerely,
Team#2422816

9 References

[1] Ben Moss & Peter O’Donoghue (2015) Momentum in US Open men’s singles tennis, International Journal of Performance Analysis in Sport, 15:3, 884-896, DOI: 10.1080/24748668.2015.11868838

[2] Zhang Jie. Research on the application of Djokovic Grand Slam in the 2021 season based on the entropy-gray correlation method [D]. Shanghai Normal University.2023.

[3] Luo Yijie. Technical and tactical analysis of the "key points" of the world's outstanding men's tennis players [D]. Hunan Normal University.2012.

[4] Fu Rao; Sun Satellite;. Establishment and application of the strength evaluation method of the three indicators of hard-ground women's tennis doubles [J]. China Sports Science and Technology 0479-83.