Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2424505

Index Analysis and Momentum Feature Judgment based on Random Forest and Decision tree

Summary

Unveiling the elusive force of momentum in tennis, this study ventures into the heart of the 2023 Wimbledon Gentlemen's final, embarking on a quest to decipher its influence on the tide of battle between titans Carlos Alcaraz and Novak Djokovic.

This paper first preprocesses the data, breaking down player performance into four distinct abilities and constructing a Random Forest model to weigh match data fields that encapsulate these abilities. Visualization of player abilities and overall performance through early rounds of the 2023 Wimbledon illustrates the match flow.

For the second question, a "momentum" model was developed, quantifying momentum changes through scores for base points, break points, hold points, and consecutive scoring. Cumulating these scores, we depicted the match's momentum shifts, highlighting control shifts and pivotal moments. Analysis of momentum's role revealed a correlation between maximum "momentum" scores and match victors, supported by permutation tests confirming momentum's significant effect.

The third question involved extracting basic and advanced features to define and predict match turning points. A Decision Tree model, applied to match number 1701, demonstrated a 97.8% accuracy in predicting such points, offering insights into factors influencing momentum shifts and providing tactical advice for players.

In the fourth question, further testing of the model on matches numbered 1403 and 1503 showcased its robust predictive strength in identifying upcoming turning points. The model's application to the 2020 Tokyo Olympic Games' table tennis matches assessed its generalizability across sports, revealing limitations and the need for future models to incorporate sport-specific nuances.

Finally, the paper concludes with a comprehensive summary of findings and recommendations for coaches, emphasizing psychological adjustment, diversified training scenarios, enhanced physical training, tactical adaptability, and comprehensive mental health strategies. These suggestions aim to prepare players for tactical and psychological adjustments under varied match conditions.

Keywords: Momentum, Random Forest, Permutation Test, Predictive Model, Decision Tree, Turning Point

Contents

1 Introduction

1.1 Problem Background

Tennis competition rules are more rigorous, and the competition process for five sets of three wins, divided into three rounds of sets, games, small points, the outcome of each set is determined by the game, the outcome of each game is determined by small points, small points determine that under normal circumstances priority than the opponent more than one ball to win, 40:40 situation limited than the opponent more than two balls to win. In a match, both players change servers after each game, triggering the deduction rule and awarding the other team's points.

In the 2023 Wimbledon men's singles final, 20-year-old Spanish rising star Carlos Alcaraz defeated 36-year-old Novak Djokovic in a match full of ups and downs and twists. In fact, the performance of tennis players is affected by objective factors such as momentum and subjective factors such as momentum, so it is very difficult to predict and quantify the performance of players.

In order to quantitatively evaluate the flow of tennis matches, our team uses data analysis to achieve the following goals:

- Create a real-time scoring model of the competition process, analyze the player's performance, and visualize the model to describe the competition process.
- Create a model indicator to evaluate whether the statement "momentum has an effect on tennis matches" is true.
- Identify the signs of favor conversion during the game and create a model to predict, analyze momentum-related factors and make profit recommendations.
- Evaluate the model's predictive power, improvement factors and degree of generalization.
- Write reports, summarize results, and make match recommendations to coaches and tennis players.

1.2 Our work

1. In this paper, the data parameters are divided into four categories of players' offensive ability, defensive ability, control ability and competitive state. We use the random forest model to calculate the contribution degree of each index, and get the relationship model between the player's score and each index.

2. We define and quantify momentum, set the calculation of momentum score based on the definition, and conduct random tests to analyze the relationship between momentum score and competition winner, and verify whether there is a significant correlation between the two. Then we use permutation test to generate random data sets by randomly permutation match result labels, calculate momentum scores under random conditions, get the consistency distribution with the winning side, and compare the original data to draw conclusions.

3. We define the turning point of favorable party transformation, capture the features that may affect the turning point, build a decision tree model, and use the captured features to judge whether the score point is the turning point, so as to obtain the connection between the turning point and the grasping feature and the judgment method of the turning point, then we test and evaluate the model through the confusion matrix. Make suggestions to players based on the

characteristics of the grab and its priority.

4. We use the decision tree model to evaluate and predict other tennis matches in Tokyo and the men's singles final of the Olympic Games.

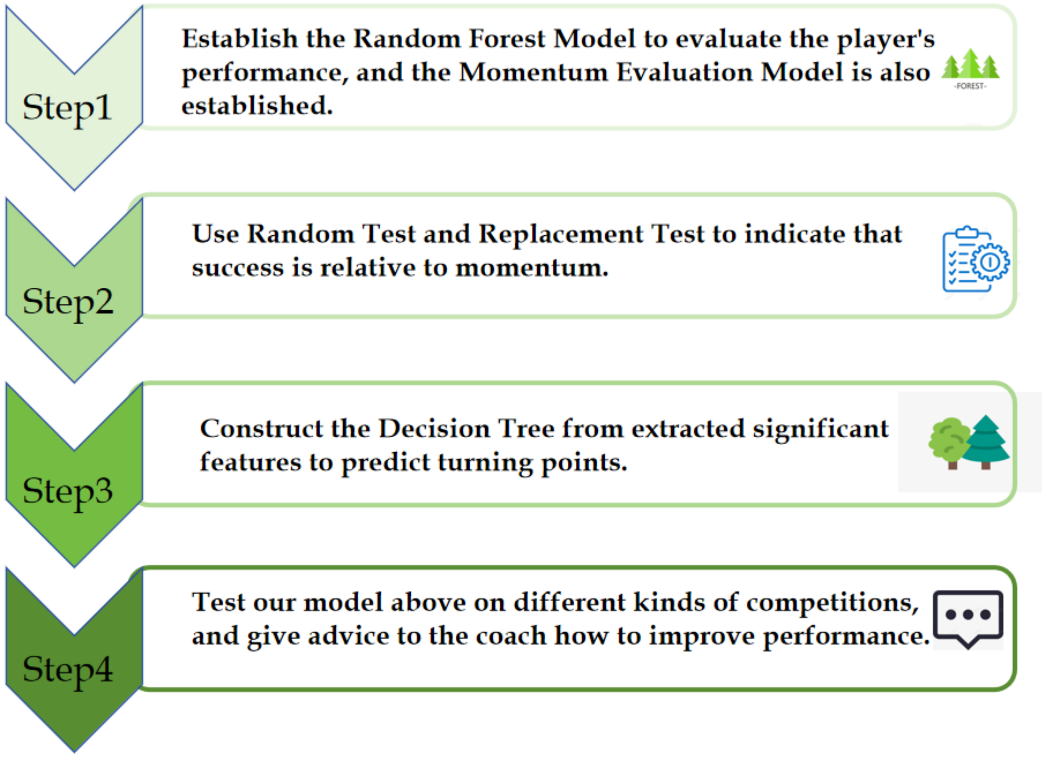


Figure 1: flow chart

2 Assumptions and Notations

2.1 Assumptions

- During the competition, the players' performance is influenced by the competition rules and their own performance.
- The indices of "momentum" calculation are linearly additive, without regard to possible nonlinear relationships or interactions between them
- In all matches, the index weights are fixed, regardless of the specific situation of the match, the difference between the strength of the players, the influence of the stage of the match and other factors.
- Ignore external factors that may affect the "momentum", such as the player's mental state, physical condition, weather conditions, crowd support, etc.
- A player's momentum accumulates continuously during a race.

2.2 Notations

Table 1: Notations Table

Notations	Definition
X_j	the j-th feature
J	the number of features
VIM_j	the contribution of the j-th feature
F	the composite performance score
F_i	the qbility index score
$S_{momentum}$	the momentum score
S_{base}	the basic score
S_{hold}	the guarantee score
S_{break}	the break score
$S_{continiue}$	the continuous score

3 Task1: Random Forest Model

3.1 Data Preprocess

We normalize non-0-1 data to avoid situations where one indicator has too much influence, such as service speed, travel distance, and parallel counts. At the same time, we quantified the label encoding of non-numerical data to simplify the difficulty of analysis, such as serve depth, serve width and return depth. Then we use the average value of the current server player to supplement the missing value of service speed. In order to simplify the solution of the objective function of the random forest model in the later stage, we added the minor score data to calculate the minor score of players on both sides. The result is a feature database with values between 0 and 1.

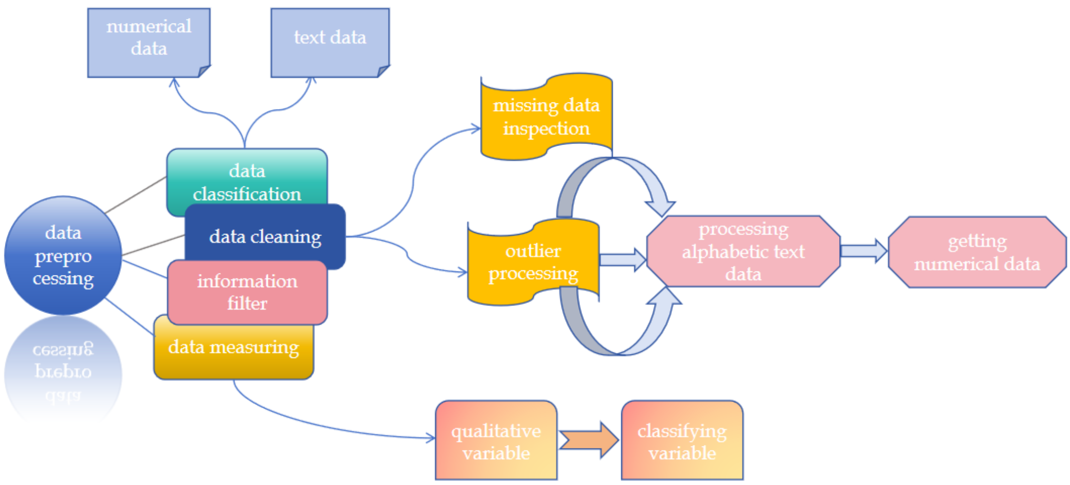


Figure 2: data Preprocess

3.2 Random Forest

Taking the player's performance as the target parameter, we use 14 data parameters as the features, and divide them into the four indicators, namely, offensive ability, defensive ability, control ability and competitive state.

Figure 3: mind map of indicator division

After classification, we invert the negative index value to convert it into a positive index and then construct a random forest model to solve the contribution degree of the secondary index to the target index, representing the variable importance score by VIM and the Gini index by GI. There are J features X_1,X_2,...,X_J. Each minor score is taken as the decision tree, so there are I decision trees in total, and I is the total number of minor score times of each player. Let's start by calculating the Gini index score VIM_J for each feature X_J.

The Gini index of node q of tree i is calculated as

$$GI_q^{(i)} = \sum_{c=1}^{|C|} \sum_{c' \neq c} p_{qc}^{(i)} p_{qc'}^{(i)} = 1 - \sum_{c=1}^{|C|} \left(p_{qc}^{(i)}\right)^2 \tag{1}$$

Where, C indicates that there are C categories, namely scoring and not scoring, and p_{qc} represents the proportion of category c in node q.

The importance of feature X_j in node q of tree i is the Gini index change before and after node q branching, which is

$$VIM_{jq}^{(\mathrm{Gini})(i)} = GI_q^{(i)} - GI_l^{(i)} - GI_r^{(i)} \tag{2}$$

Where $GI_l^{(i)}$ and $GI_r^{(i)}$ represent the Gini index of the two new nodes after branching, respectively.

If the node where feature X_j appears in decision tree i is set Q, then the importance of X_j in tree i is

$$VIM_j^{(\text{Gini})(i)} = \sum_{q \in Q} VIM_{jq}^{(\text{Gini})(i)} \tag{3}$$

If RF has I trees, then

$$VIM_j^{(\text{Gini})} = \sum_{i=1}^{I} VIM_j^{(\text{Gini})(i)} \tag{4}$$

Finally, all the obtained importance scores are normalized.

$$VIM_j^{(\text{Gini})} = \frac{VIM_j^{(\text{Gini })}}{\sum_{j'=1}^{J} VIM_{j'}^{(\text{Gini})}} \tag{5}$$

Using the data of p1 player, the random forest model is calculated and solved by python, and the contribution degree of each feature to the player's performance was obtained.

Table 2: the degree of contribution of the feature

Feature	Importance
speed_mph	0.344637917
p1_distance_run	0.274844723
serve_width	0.124565842
return_depth	0.064615154
serve_depth	0.042951032
game_victor	0.037170509
point_victor	0.031790555
p1_unf_err	0.021500137
p1_winner	0.019451708
p1_net_pt	0.013721211
set_victor	0.009028027
p1_net_pt_won	0.007327595
p1_ace	0.004245402
p1_break_pt_won	0.004150186

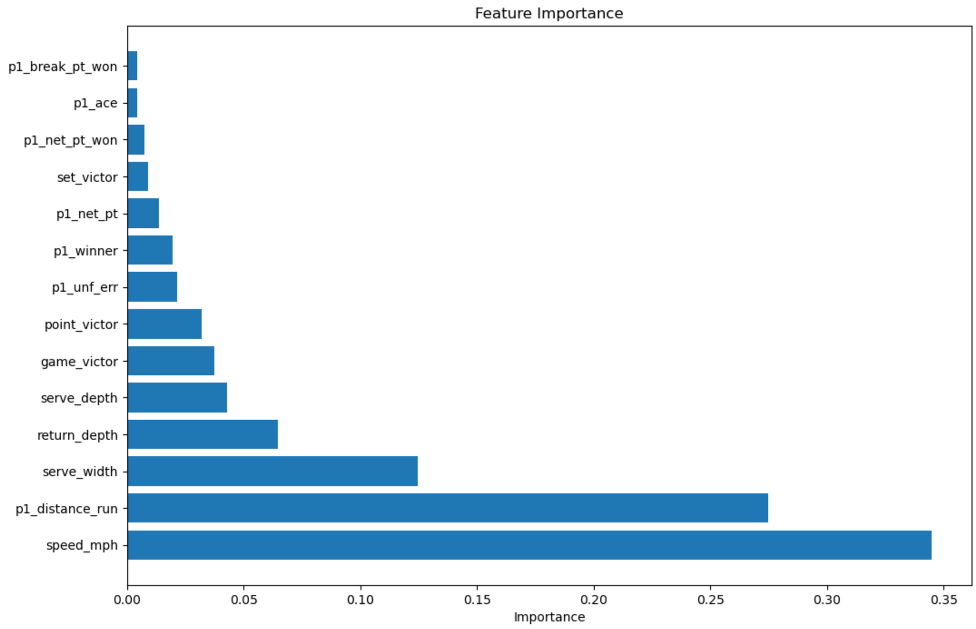


Figure 4: visualization of the contribution of features

According to the contribution degree of the features, the contribution degree of the four indicators to the performance of the players is determined.

Table 3: the contribution degree of four indicators

Index	Importance
offensive ability	0.344637917
defensive ability	0.274844723
control ability	0.124565842
competitive state	0.064615154

3.3 Competition Process

In the competition process, since the score is related to the moment node, the score replaces the time as a factor to promote the game, so as to analyze and visualize the comprehensive performance and ability performance of p1 and p2 players each time the score is scored.

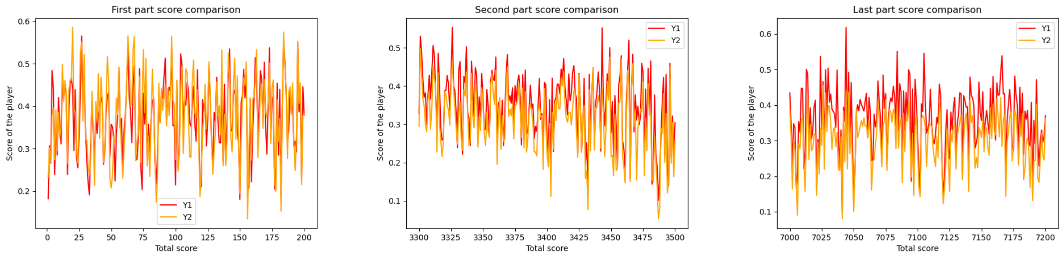
$$F = \sum_{j=1}^{J} (VIM_j * X_j) \tag{6}$$

Where F stands for the player's overall performance score.

$$F_i = F * VIM_i \tag{7}$$

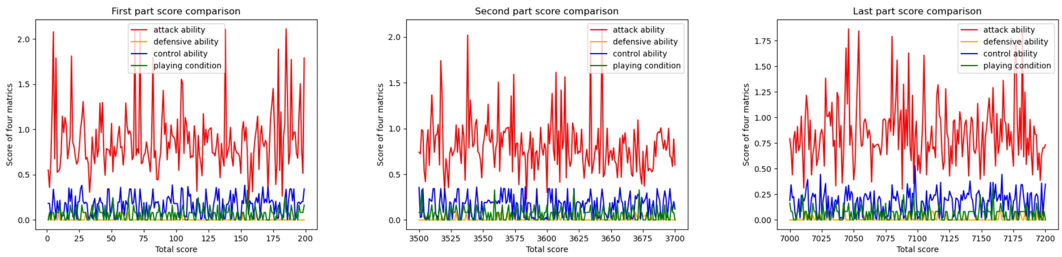
Where, F_i represents the player's ability performance score, and VIM_i represents the contribution degree of the player's four ability indicators to the comprehensive performance.

Too much data makes it difficult to fully visualize all competition processes, so we use a random forest model to intercept data from the early, middle and late stages of the match for visualization.Take the comprehensive performance of both players and the performance of p1's ability index as examples.



(a) early stage (b) metaphase (c) anaphase

Figure 5: comprehensive performance



(a) early stage (b) metaphase (c) anaphase

Figure 6: ability performance

Since the competition stipulates that the players of both sides rotate the server, and the server's win rate is larger, the evaluation of comprehensive performance will produce a situation of stage rise and stage decline, which is in line with the actual situation.

4 Task2: Random Test and Replacement Test

4.1 Momentum Measurement

Since momentum is the energy brought about by movement and a series of activities, and is related to momentum, momentum is measured and quantified through base points, break points, guaranteed points, and consecutive points[1].

Table 4: the contribution degree of four indicators

goal	influencing factor	explain	point
momentum	base point	win a small point	1
	break point	win a point on your opponent's service game	5
	guaranteed points	win a point in your own service game	3
	consecutive point	win more than three points in a row	2

The player's "momentum" point at each scoring point

$$S_{momentum} = S_{base} + S_{break} * x_{ifb} + S_{hold} * x_{ifh} + S_{continiue} * x_c \tag{8}$$

Where, $S_{momentum}$ is the momentum score, S_{base} is the base score, S_{break} is the break score, which means that a player wins a game at 40:40.S_{hold} is the guarantee score, which means that a player wins a game as the server. $S_{continiue}$ is the continuous score, x_{ifb} represents whether the break is made, x_{ifh} represents whether the guarantee is made, and x_c represents several consecutive points.

"Momentum" point is cumulative, so a player's total "momentum" point in a competition is the sum of the "momentum" point on each scoring point

$$S_{total} = \sum_{i=1}^{n} S_{momentum} \tag{9}$$

Where n is the total number of points scored in the competition, and $S_{momentum}$ is the "momentum" point on the i-th score.

The model is solved.

Table 5: momentum point

set_no	game_no	point_no	server	complex_momentum_p1	complex_momentum_p2
1	1	1	2	0	0
1	1	2	2	6	0
1	1	3	2	12	0
1	1	4	2	12	1
1	1	5	2	18	1
1	1	6	2	18	2
1	1	7	2	18	3
1	1	8	2	24	3
1	1	9	2	24	4

The "momentum" of the visualization changes with the flow of the competition.

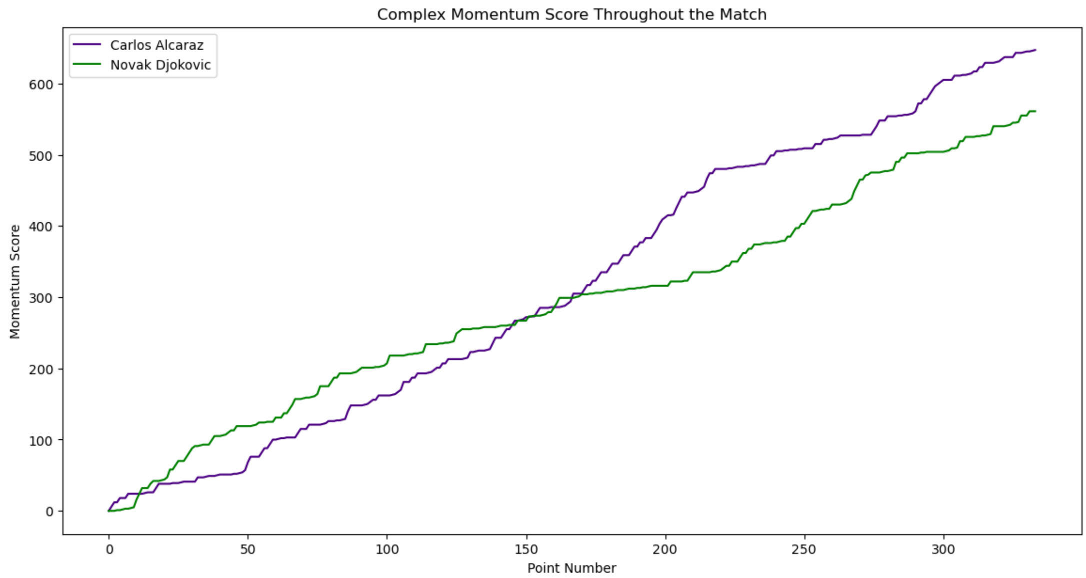


Figure 7: the change in "momentum" of a set

4.2 Random Test

If there is a significant correlation between "momentum" points and competition wins, this may indicate that "momentum" has a real impact on the outcome of the competition.

Get the "momentum" point data from an Excel file containing the details of each score point. The winner is inferred by looking at the last score of each competition. Suppose the player with the highest "momentum" on the last scoring point is the highest "momentum" scorer. For each match, the point of the maximum "momentum" achieved by two players during the match is calculated. Use a scatter plot to show the maximum "momentum" point for both players. Use color to distinguish between the maximum "momentum" score and the actual winner.

Suppose that in the competition, the "momentum" points of Player 1 and Player 2 at i-th point are $M_{p1}^{\max}$ and $M_{p2}^{\max}$ respectively, and the maximum "momentum" points of player 1 and player 2 in the competition are:

$$\begin{aligned} M_{p1}^{\max} &= \max_i \left(M_{p1}^{(i)} \right) \\ M_{p2}^{\max} &= \max_i \left(M_{p2}^{(i)} \right) \end{aligned} \tag{10}$$

The winner W can be determined by comparing the "momentum" point of the last scoring point:

$$W = \begin{cases} \text{player 1, if } M_{p1}^{(\text{last})} > M_{p2}^{(\text{last})} \\ \text{player 2, otherwise} \end{cases} \tag{11}$$

Where $M_{p1}^{(\text{last})}$ and $M_{p2}^{(\text{last})}$ are the "momentum" points of Player 1 and Player 2, the last point in the competition, respectively.

By analyzing the association between the maximum "momentum" point and the winner of the match, we can assess the impact of "momentum" on the outcome of the match. If the winner of the greatest "momentum" point is usually the winner of the competition, this may indicate that "momentum" plays a role in the comprtition.

Table 6: the association between maximum ”momentum” points and competition winners

set_no	game_no	winner	max_m_p1	max_m_p2	max_m_winner	m_match_winner
1	1	player1	24	5	player1	TRUE
1	2	player2	26	38	player2	TRUE
1	3	player2	38	47	player2	TRUE
1	4	player2	41	82	player2	TRUE
1	5	player2	47	93	player2	TRUE
1	6	player2	51	105	player2	TRUE
1	7	player2	51	113	player2	TRUE
2	1	player2	57	119	player2	TRUE
2	2	player2	100	125	player2	TRUE
2	3	player2	103	157	player2	TRUE

Looking at the complete table, the highest ”momentum” scorer is the winner of the game, which can be visualized by using a scatter plot.

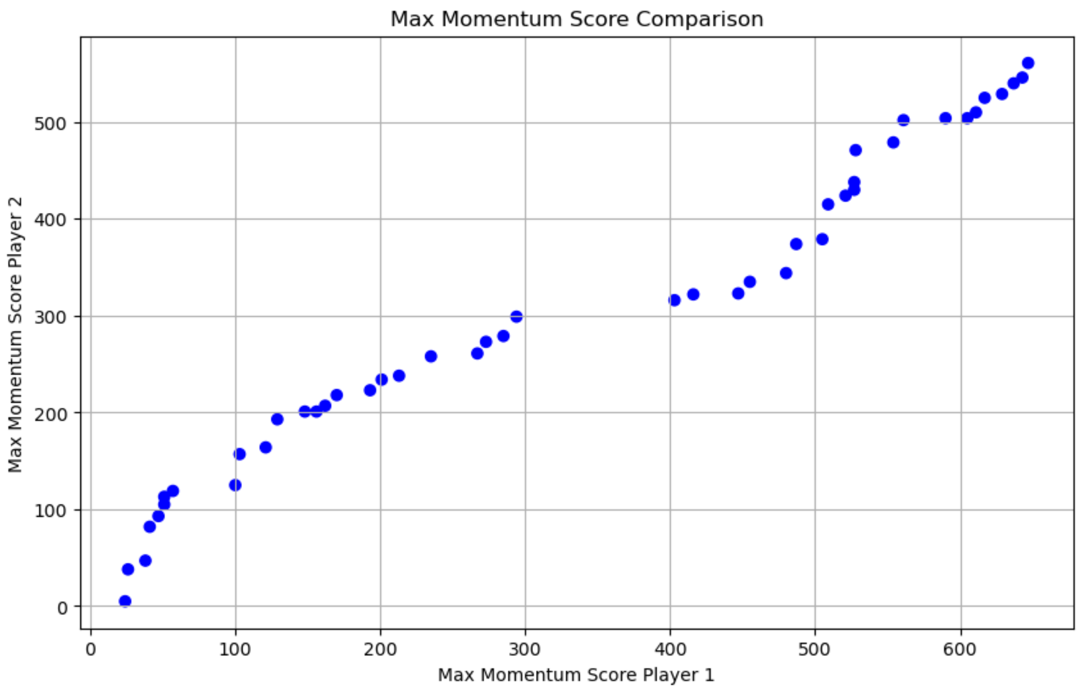


Figure 8: the degree of match between the highest ”momentum” scorer and the winner (coincidence is blue, non-coincidence is red)

All points are blue, indicating that the highest ”momentum” scorer overlaps exactly with the winner.

4.3 Replacement Test

4.3.1 Basic Idea

In order to perform a statistical test to see if changes in ”momentum” points affect victories beyond random fluctuations, we can use a nonparametric test such as the permutation test. This method does not depend on the specific distribution of data, which is more suitable for this situation.

The basic idea of the permutation test is to generate a large number of random permutations by redistributing the labels of the data points, and then calculate the statistics under some random permutations, for example, the proportion of the maximum momentum scorer who agrees with the winner. By comparing the actual observed statistics with the distribution of statistics generated by random permutations, we can assess the likelihood that the actual observed statistics will occur in a random situation, and thus determine whether the change of momentum point has a statistically significant effect on victory

4.3.2 Model Building and Solving

The purpose of the substitution test is to redistribute the attribution of momentum scoring data points, that is, to randomly assign a certain scoring point to either of these two players, thereby creating many possible match outcome scenarios. By comparing the statistical results in these random scenarios with the actual observed results, we can assess the particularity of the actual results and thus determine whether the relationship between momentum points and competition wins is beyond the range of random fluctuations.

The steps are as follows:

1. Define the null hypothesis (H_0) and the alternative hypothesis (H_1):

H_0:There was no difference between the sample groups.

H_1:There was differences between the sample groups.

2. Calculate the raw statistic:

Choose an appropriate statistic T to measure the difference between the sample groups. For example, the difference can be the difference in the mean, the median, or any other measure that fits the problem.

Two sample groups A and B are used, where A and B represent the matching degree of score points and maximum momentum of player 1 and player 2 respectively. The original statistic T_{obs} can be the difference between the means of the two groups:

$$T_{obs} = A_{mean} - B_{mean} \tag{12}$$

3. Generate random permutation:

Multiple randomly arranged data sets are generated by randomly exchanging group labels of sample data points. These random permutations represent the data permutations that might be observed if the null hypothesis were true.

4. Calculate the statistics for each permutation:

For each randomly arranged data, calculate the same statistic T_{perm} as the original statistic.

5. Calculate P-value:

P-value is the position of the observed statistic T_{obs} in the distribution of the statistic generated by all random permutations. It represents the probability that an observed statistic or a more extreme statistic will occur if the null hypothesis is true.

$$P = \frac{m}{M} \tag{13}$$

Where m represents the number of T_{perm} greater than T_{obs}, and M represents the number of all permutations.

6. Draw a conclusion:

Decide whether to reject the null hypothesis based on the P-value reaching α pre-set significance level a (usually 0.05). If the P-value is less than or equal to α, the null hypothesis is rejected, indicating that the difference between the sample groups is statistically significant.

7. Visualization:

Histograms and raw statistics showing the result of the permutation test:

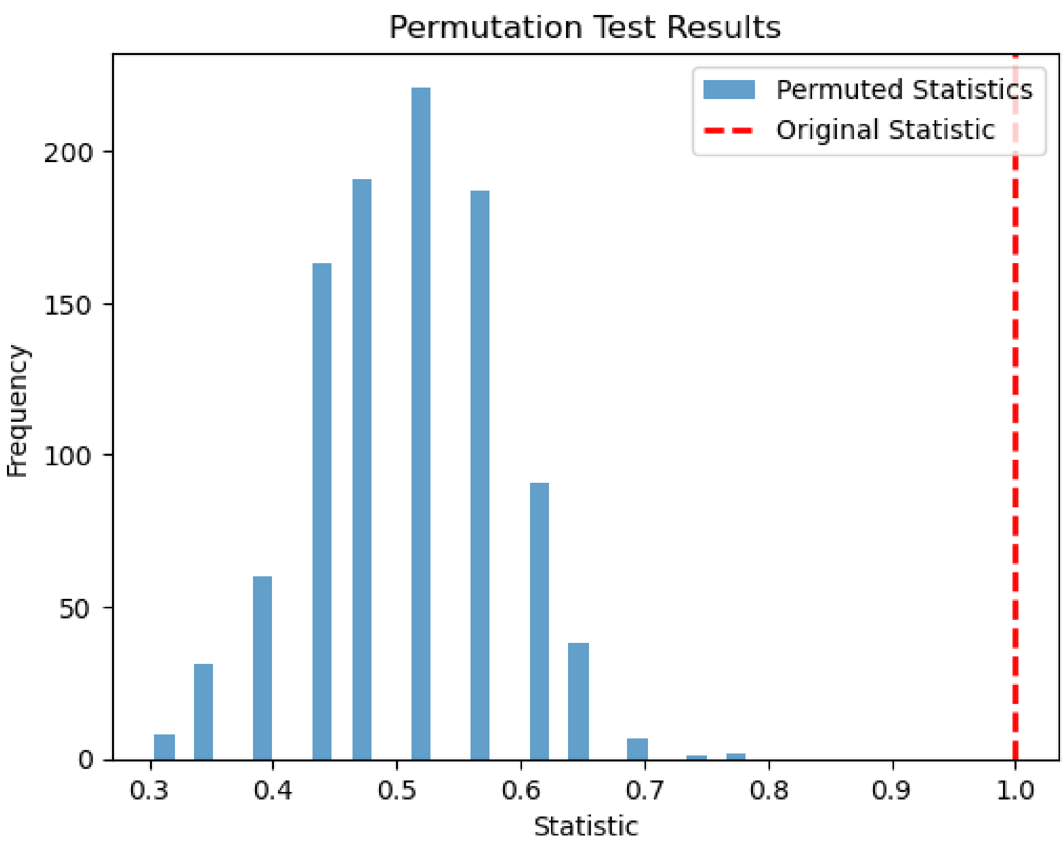


Figure 9: result of replacement test

Histogram: This histogram shows the distribution of statistics calculated under all the random permutations generated during the permutation test, such as the "momentum" score to the winner consensus ratio. Each bar represents the number of times in all permutations that the statistic falls within a specific range. This distribution reflects the distribution of statistics that we would expect to see under completely random conditions.

Raw statistic: This red dotted line represents the value of the raw statistic calculated from the actual data. In the context of the "momentum" point, this is probably the largest proportion of "momentum" point in the raw data that agree with the actual winner.

The raw statistic is on the far side of the histogram, which means that under random conditions, it is very rare to observe a statistic like the one in the raw data. Therefore, we can assume that the phenomena represented by the raw statistics are not likely to have occurred by chance, but are statistically significant. Therefore, the situation in which the ”momentum” point coincides with the actual winner is unlikely to have occurred by chance.

4.3.3 Results and Conclusions

The P-value is used to evaluate the probability that the original obsrvation will occur in a random situation. In our case, a P-value of 0.0(given possible rounding errors in the actual calculation, the actual P-value may be non-zero but very close to zero) indicates that it is difficult (or nearly impossible) to observe a proportion of the maximum ”momentum” point that agrees with the actual winner as high as the original data in the case of random permutations. This suggests that the original observation is unlikely to have occurred by chance, and that the consistency between the ”momentum” point and the actual winner of the match is statistically significant.

The results of the substitution test support the idea that ”momentum” plays a role in tennis matches, rather than being completely random. This suggests that ”momentum” shifts in a game and the winning streak of a player on one side are unlikely to be just random events, but rather related to the actual dynamics and outcome of the competition.

5 Task3: Decision Tree Model

5.1 Feature Extraction

For this question, to develop a model to predict these turning points in a competition, we first need to extract the feature data as the independent variable, define the turning point as the dependent variable, and then use the machine learning algorithm to make the prediction.

The extraction steps are as follows:

1. Extract Basic Features:

- match_id: Identify each match
- set_no,game_no,poim_no: Identify the competition's set number,game number, and scoring point number.
- server: Mark the server.
- p1_score, p2_score: Record the score of both sides.

2. Extract Advanced Features:

• break_point:

Break point is a key match dynamic indicator and often signals a potential momentum shift. This feature is extracted to capture the ability of the non-serving side to win points in the service competition, which is often considered a significant advantage in the competition.

extraction_method: According to the instructions of the data dictionary, use the score and server information to identify the break point, and mark whether each scoring point is a break point.

• break_score:

This feature indicates which player has won the point at the break point. This helps analyze how players perform under pressure and how they control the competition at crucial moments.

extraction_method: After identifying the break point, check the point_victor column to determine which player won the points at the break point, and create break_point_win_p1 and break_point_win_p2 columns for both players.

• score_row:

Score_row reflects a player's current state and momentum. High consecutive scores are usually associated with strong control of the competition.

extraction_method: Calculate the number of consecutive scores won by each player (consecutive_points_p1 and consecutive_points_p2).

• momentum_difference:

The momentum difference is a measure of the momentum of the match between two players. A significant change in the momentum differential could signal a turning point in the competition.

extraction method: Calculate the difference in momentum score , which is obtained by subtracting the momentum score of two players.

3. Mark Turning Point

Identifying the turning point in the competition is the key to predicting the dynamic change of the competition. A turning point may be an indicator of a change in game momentum, strategy, or mental state.

This can be done by setting a threshold, such as when the change in the score_row exceeds a certain value, such as 3 points, it is considered a turning point.

5.2 Decision Tree

We have extracted the feature data of all the matches before. Now, we use the match No. 1701 as an example to conduct the actual decision tree classification model.

We analyze the extracted feature and turning point data through the decision tree model, and classify the feature as the evaluation basis to determine whether the score point is the turning point.

The steps are as follows:

1.The data is divided into training set and test set, and the ratio is 7:3.

2.Introduce the decision tree model, set the number of features captured as D, calculate the Gini coefficient of all features, and select the feature with the smallest Gini coefficient as the classification standard of the first layer; After the first layer classification, the Gini coefficients of the remaining D-1 features are calculated, and the feature of the smallest Gini coefficient is used as the classification standard of the second layer; and so on, until the exact score point that is the turning point is identified (class=1) and the Gini coefficient for all classes is zero (Gini=0).

$$\text{Gini} = \sum_{i=1}^{C} p_i * [1 - p_i] \tag{14}$$

Where C is the number of classes and p_i is the probability that a sample will be classified into class i.

3.Obtain the priority of all feature categories

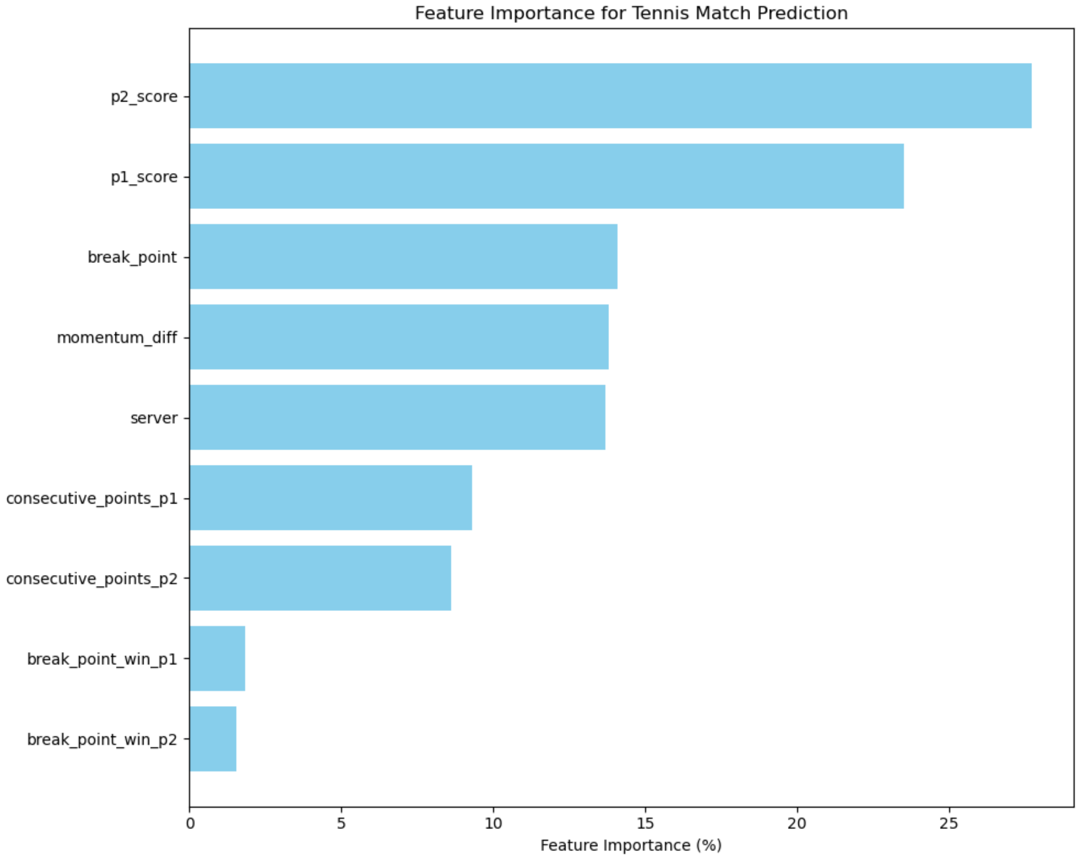


Figure 10: feature importance

The upper bar chart shows the proportion of importance of each feature.

4.The training set and test set are constantly learned through python to determine the most suitable classification criteria.

Table 7: parameter set

Parameter Name	Parameter Value
Training Time	0.004s
Data Split	0.7
Data Shuffling	No
Cross-Validation	No
Criterion for Node Splitting	Gini
Best Feature Splitting Criterion	Best
Maximum Feature Ratio for Splitting	None
Minimum Samples for Internal Node Splitting	2
Minimum Samples for Leaf Node	1
Minimum Weight of Samples in Leaf Node	0
Maximum Number of Samples in Leaf Nodes	50
Maximum Tree Depth	10
Impurity Threshold for Node Splitting	0

5.Visualize the decision tree model

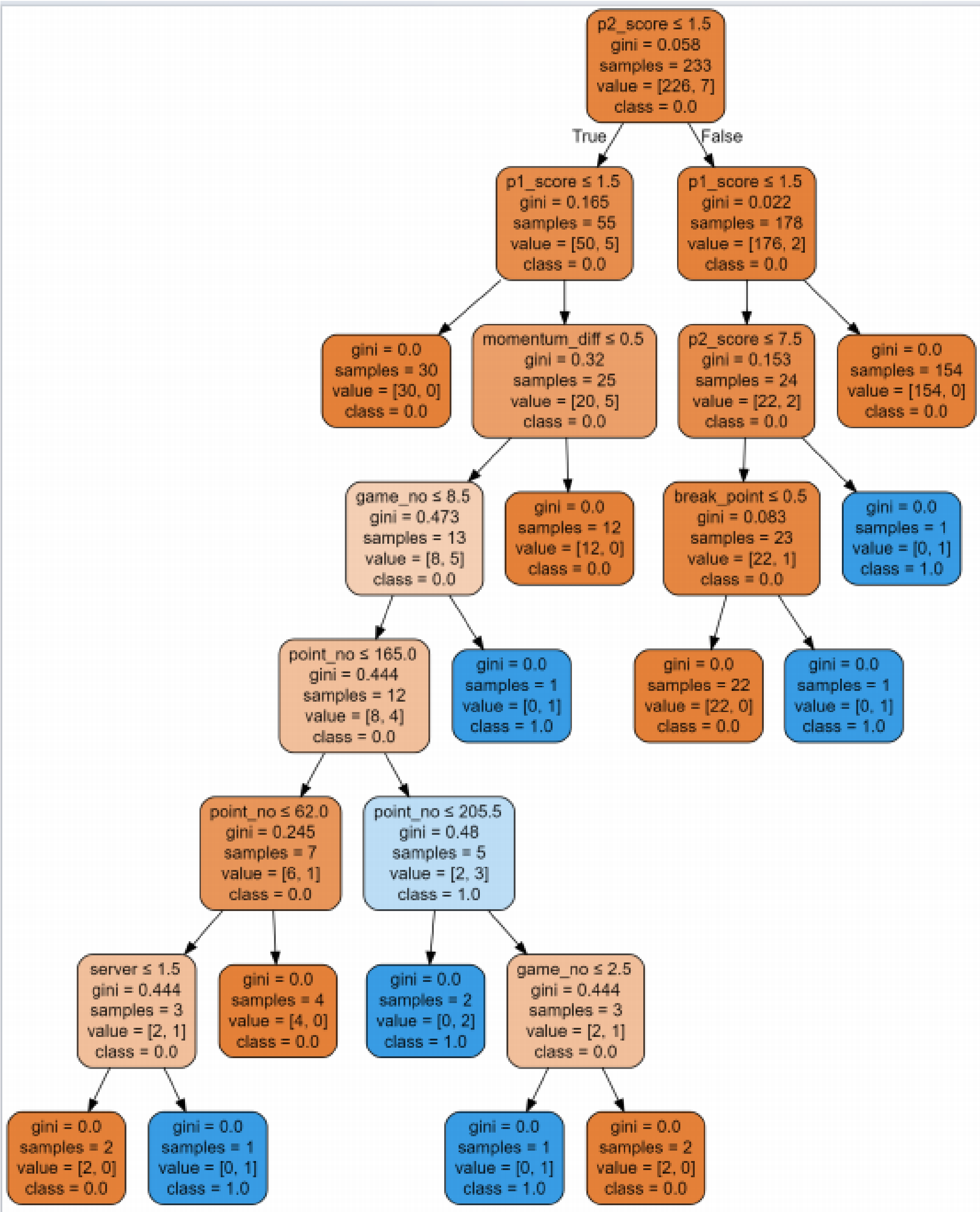


Figure 11: decision tree model

The figure above shows the structure of the decision tree, the internal nodes give the specific segmentation of the branched features, that is, according to a certain segmentation value of a feature.

• Information entropy is used to determine which feature to slice.

• Sample type distribution refers to the number of samples belonging to each classification group in this node. For example, [10,5,5] indicates that there are 10,5 and 5 samples in the three classification groups respectively.

• The classification situation is the classification group into which the samples of the node are uniformly divided, which is determined by the group with the largest sample size.

6.Visualization of confusion matrix heat map

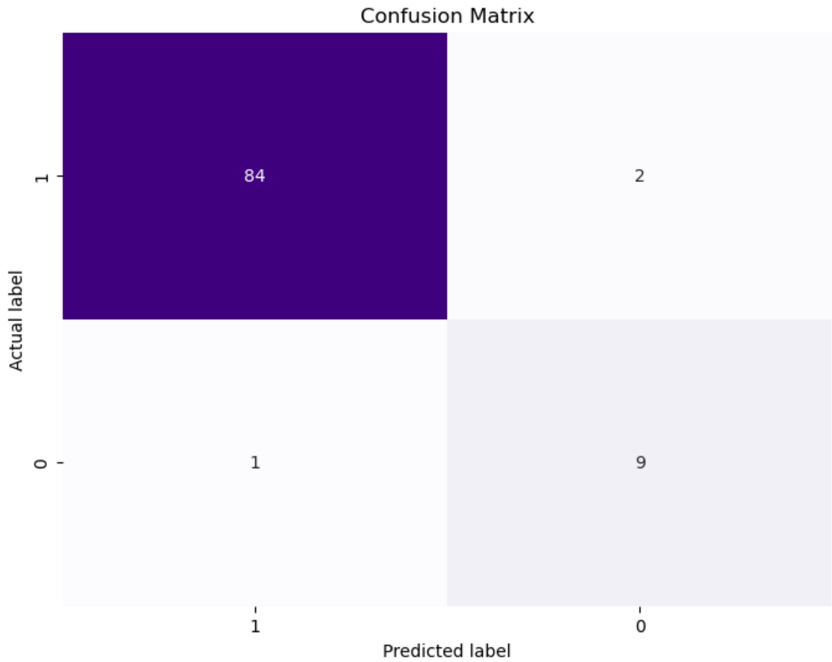


Figure 12: confusion matrix heat map

The above figure shows the confusion matrix in the form of a heat map.

Table 8: model evaluation result

	Accuracy	Recall	Precision	F1
Training Set	1	1	1	1
Test Set	0.978	0.976	0.988	0.982

The above table shows the classification evaluation indexes of training set and test set, and measures the classification effect of decision tree on training and test data through quantitative indexes.

●Accuracy: Predict the proportion of correct samples in the total sample.

●Recall rate: The proportion of predicted positive samples in results that are actually positive samples.

●Precision: The proportion of predicted positive samples that are actually positive.

●F1: The harmonic average of accuracy rate and recall rate, accuracy rate and recall rate are mutually affected.

5.3 Match Suggestion

The decision tree classification model was used to predict turning points and achieved 97.8% accuracy, which shows that the model is very effective at capturing patterns in the match data, and we can use the model to understand and predict the key factors of momentum shifts, and accordingly provide players with the following preparation strategies:

● Analyze the Correct Game Pattern

The decision tree model is used to analyze the key moments and triggers of ”momentum” shifts in past matches of opponents. For example, if the model finds that an opponent tends to lose ”momentum” after a particular score, then a strategy can be developed for this during the competition.

• Focu on Key Points

By analyzing which scoring points are most likely to trigger a ”momentum” shift, players can focus on those points to ensure they play their best when it matters.

• Optimize Serving and Receiving Strategies

The model shows that serving or receiving is one of the most important factors that lead to the ”momentum” change, so players should pay special attention to these links in training.

• Adjust the Pace of the Game

According to the model’s findings, players may need to adjust their pace during a match, such as speeding up the pace of the game when the momentum is shifting in their favor, or slowing down when the momentum is not in their favor, to disrupt the opponent’s rhythm.

• Develop a Specific Competition Plan

Based on the ”momentum” transition points predicted by the model, players can develop detailed competition plans, such as maintaining a high level of concentration before and after the predicted turning point or using specific tactics at these moments.

• Train Simulation

By simulating the ”momentum” changes that might occur in a race during training, players can better cope with these situations in a real race.

6 Task4 :Model Evaluation

6.1 Model Prediction

We applied the constructed decision tree model to the two tennis matches numbered 1403 and 1503 to predict the turning point in the match according to the index we extracted, and compared it with the actual turning point we defined. The result is shown in the figure below.

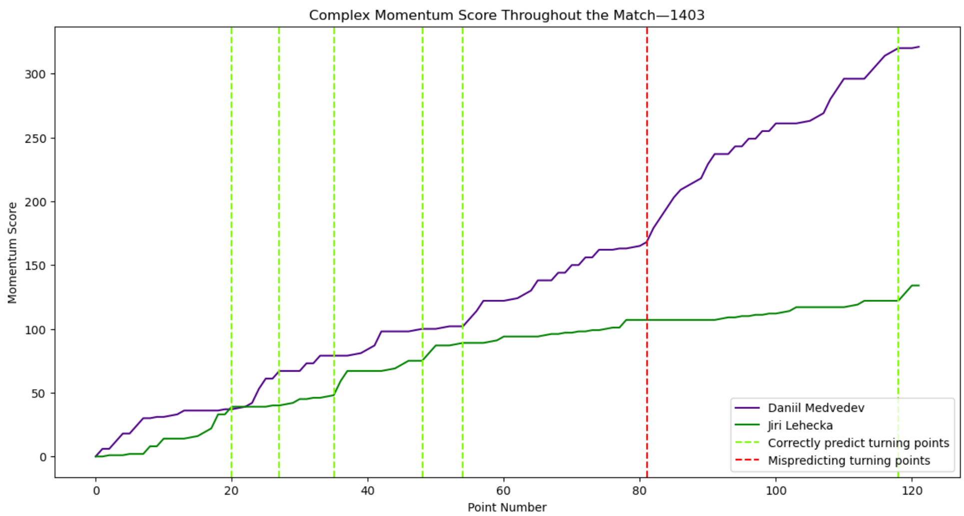


Figure 13: 1407 match prediction results

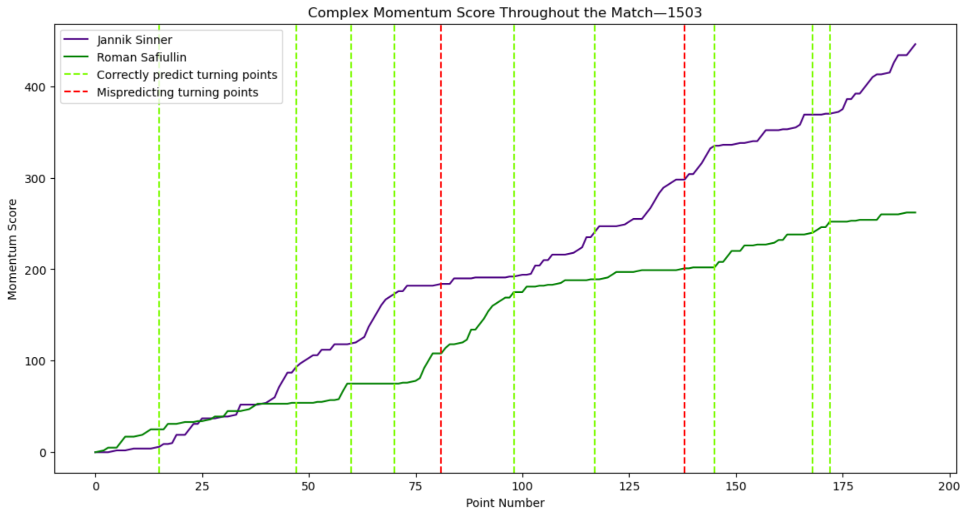


Figure 14: 1503 match prediction results

The results clearly show that our model is very good at capturing turning points accurately, so our model did a good job of capturing the moment when the turning point is coming up during the tennis tournament 2023Wimbledon. In addition, we also found that the turning point is about to be accompanied by small or large fluctuations in ”momentum”, which also shows that the ”momentum” of the players is closely related to the change of the situation.

Although the model is quite satisfactory for capturing upcoming turning points, this paper identifies factors that may need to be included in the model in the future:

1. Player Fatigue and Physical Fitness
2. Playing Conditions (Weather), Court Surface, Ball Speed, etc
3. Psychological Factors
4. Tactical Changes[2]
5. The Importance of the Stage
6. Player Interaction

6.2 Model Generalization

In order to verify the generalization ability of the model established in this paper on other competitions, this paper selects the 2020 Tokyo Olympic Games men’s table tennis singles final as an example to apply the model established in this paper. (The data source is the game score data collected by watching the game video and using manual data)

We applied the momentum model established in this paper to table tennis match to visualize the momentum change process of both players in the match. At the same time, we used the decision tree model established earlier in this paper to identify and capture the time nodes about to change advantages in the match and mark them, and compared them with the turning point defined by us.

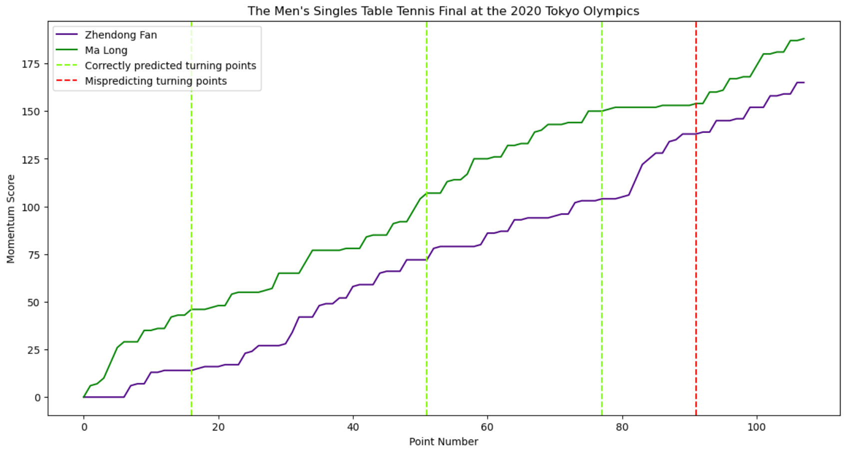


Figure 15: Tokyo 2020 Olympic Games men's singles final

From the figure, we can find that our model performed reasonably well in the selected table tennis match, but did not identify the last turning point in the match. We consider that it may be because of the obvious difference between the competition system of tennis and table tennis. Especially for players with equal strength, there is a significant difference from tennis matches, in table tennis, the conversion of service rights is more frequent, the server has a greater probability of victory, when the service rights are converted, the original advantage of the server in the previous round will be more easily broken.

In terms of different games in the same sport:

1.The Women Player and the Men Player

• Physical Intensity Differences: Men's games tend to emphasize strength and endurance, which can lead to different game rhythms and scoring patterns. For example, the men's games may have more ACES, while the women's gameS may have more frequent return fights from the baseline, which will obviously cause the "momentum" to change in a way[3].

• Length of Matches: Men's Grand Slams are usually best-of-five and women's best-of-three, and this structural difference can affect game strategy and physical management, which can affect trends in "momentum."

2.Different Tournaments

• Level of Competition: Grand Slams have a higher level of competition and pressure compared to ATP/WTA Tour or smaller tournaments, which can cause players to perform differently.

• Audience and Environment: Different tournaments and audience sizes, venue environments and media attention may have an impact on the mental state and performance of team players.

3.Surface Variation

• Style of Play: Different court surfaces have different requirements for the bounce and speed of the ball, requiring players to adjust the style of play, such as more use of top spin on clay, and use low bounce for fast volleys on grass. This leads to the fact that the "momentum" shifts of the game on different surfaces can be driven by different factors.

In terms of cross-motion analysis:

• Differences in Rules: Different sports (such as table tennis and tennis) have very different scoring rules, game lengths, and permissible tactics, and differences in these basic principles mean that the model is adapted to the rules of the particular sport.

• Game Dynamics: The concept of game dynamics and momentum shifts can vary from sport to sport. The pace of the game is much faster than tennis, momentum shifts more frequently and for a variety of reasons.

• Skill Requirements: Different sports have different requirements for the skills and physical attributes of the players. Byeongbyeongball is highly dependent on reaction speed and short distance movement, while tennis puts more emphasis on strength, endurance and long distance coverage.

Conclusion:

Our model takes into account a common feature of different matches, that is, the difference in ”momentum” caused by the change in the score during the match, which is almost all games have a psychological change. However, in some different competitions and cross-sport competitions of the same sport, our model lacks some factors to consider.

For different matches in the same sport, the model may need to be adjusted for different genders, tournament levels, and court-table characteristics and dynamics to ensure the accuracy and applicability of its predictions.

For cross-sport generality, while some basic statistical and machine learning methods may be common across multiple sports, the models themselves often need to be specified and adapted for each sport’s unique rules, skill requirements, and match dynamics.

7 Strengths and Weaknesses

- Strengths:

• Predictive Accuracy: High accuracy in predicting turning points, validated through specific matches with a precision rate.

- Weaknessess:

• Lacks Specific Factors: Does not fully account for unique aspects of different sports and matches.

• Data Dependency: The accuracy and reliability of predictions are heavily dependent on the quality and comprehensiveness of the input data.

References

[1] Mecheri, S., Rioult, F., Mantel, B., Kauffmann, F. and Benguigui, N. (2016), The serve impact in tennis: First large-scale study of big Hawk-Eye data. Statistical Analysis and Data Mining: The ASA Data Science Journal, 9: 310-325. https://doi.org/10.1002/sam.11316

[2] RYixiong Cui, Miguel-Ángel Gómez, Bruno Gonçalves, Hongyou Liu & Jaime Sampaio (2017) Effects of experience and relative quality in tennis match performance during four Grand Slams, International Journal of Performance Analysis in Sport, 17:5, 783-801, DOI: 10.1080/24748668.2017.1399325

[3] Machar Reid, Stuart Morgan & David Whiteside (2016) Matchplay characteristics of Grand Slam tennis: implications for training and conditioning, Journal of Sports Sciences, 34:19, 1791-1798, DOI: 10.1080/02640414.2016.1139161

Suggestion Letter

Dear Coach,

I hope this letter finds you well. Following our ongoing discussions on the role of "momentum" in tennis, we have conducted a thorough analysis using ***Random Forest, Random Test, Replacement Test*** and ***Decision Tree*** models to delve deeper into the concept of momentum, and we wish to share some insights and recommendations.

Our mathematical model analysis has led to the following conclusions regarding the role of momentum:

1.Reflects match pace and control: Momentum indicators, by quantifying the correlation between scoring points, reflect the pace of the match and a player's trend of controlling the game, thus capturing key turning points.

2.Highly correlated with match outcomes: Statistical tests have shown that momentum is highly correlated with winning matches, suggesting that players with higher momentum are usually more likely to win, underscoring its significant impact on outcomes.

3.Reflects psychological states: Changes in momentum reflect shifts in a player's psychological state, where scoring streaks increase momentum and losing streaks decrease it, thus it can be used to gauge psychological conditions.

4.Provides strategic guidance: Analyzing key factors of momentum changes can offer strategic advice, such as focusing on critical scoring points, adjusting serve and return strategies, and pacing adjustments based on momentum shifts.

5.Indicates match intensity: The momentum change curve can reflect the fluctuation in match pace, thus indicating the intensity of the competition. And we use the extracted significant features by ***Decision Tree*** to predict the turning point, which can be a performance that can indicate match intensity. And our model shows 0.978 accuracy rate, 0.976 recall rate, 0.988 precision rate and 0.982 F1 rate.

Based on these findings, we propose the following strategies to better leverage momentum:

1.Advanced Psychological Adjustment: Integrate cutting-edge sports psychology practices, such as cognitive-behavioral methods and stress inoculation training, to provide athletes with mental resilience under competitive pressure, transforming potential momentum into dominant opportunities.

2.Customized Scenario Training: Include a broader range of match scenarios in training, emphasizing resilience and adaptability. This involves practicing comebacks under various conditions (e.g., different courts, weather, and psychological states) to cultivate a diverse and relentless mindset.

3.Strategic Gameplay Focus: Develop a keen awareness of tennis strategic elements, encouraging players to understand and manipulate match rhythm and tempo for gaining an edge. This includes mastering the art of score building, energy management, and opponent analysis to create and capitalize on momentum shifts.

4.Enhanced Physical Training Programs: Implement state-of-the-art physical training programs, focusing on improving speed, endurance, and recovery capabilities. Use the latest in sports science to optimize training, ensuring our athletes possess the physical advantages needed to maintain and switch momentum in the toughest matches.

5.Tactical Adaptability Training: Cultivate players' innovative tactical thinking, enabling them to adjust their playing style in real-time according to the flow of the match. This involves preparing them to effectively switch tactics, whether adopting a more aggressive approach or falling back on defensive strategies, to disrupt the opponent's momentum.

Comprehensive Preparation for Major Match Events:

Intensified Adversity Training: Expand adversity training, including building psychological resilience, to ensure players are prepared both physically and mentally to face and overcome setbacks without losing momentum.

In-depth Opponent Analysis: Utilize advanced analytics and video analysis tools for a deeper understanding of opponents' strategies and tendencies. This allows for a more strategic approach to matches, devising tailor-made game plans to neutralize opponents' strengths and exploit their weaknesses.

Integrated Recovery Protocols: Develop comprehensive recovery protocols that combine physical and psychological recovery techniques. This includes adopting cutting-edge physical recovery methods, such as cryotherapy and hyperbaric oxygen therapy, and stress reduction techniques, ensuring players are in peak physical and mental condition for every match.

By adopting these advanced strategies, we not only optimize our approach to leveraging momentum but also significantly enhance our players' readiness for the unpredictable dynamics of competitive tennis. I look forward to further discussions on these strategies and exploring how we can integrate them into our training and match preparation processes.

Yours sincerely,
MCM Team Number: 2424505