Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 7
TPCEE 2024

Study on Momentum Assessment in Tennis Matches Based on Dynamic Weights and Machine Learning

Qinghua Liu[1, *], Boshen Zhu[#, 2], Jiran Wang[#, 2], Tongtong Sun[#, 2]

[1]Department of Management Science and Engineering, Beijing Institute of Petrochemical Technology, Beijing, China, 102617

[2]School of Economics and Management, Beijing Institute of Petrochemical Technology, Beijing, China, 102617

* Corresponding Author Email: shjg066@163.com

#These authors contributed equally

Abstract. In recent years, with the popularization and technological progress of tennis, momentum has been an important indicator of changes in the players' state. Traditional evaluation methods cannot fully reflect their dynamic changes and multi-factor effects, so it is essential to establish a more accurate momentum evaluation model. In this study, seven key indicators, such as serve advantage, breakpoints, etc., were selected to measure momentum fluctuations in matches, and weight was assigned to each indicator to reflect their importance in the overall momentum. Among them, for serving advantage, the dynamic weight Wdy-serve is newly introduced in this study. Moreover, the momentum fluctuation of players in the tiebreaker between Carlos Alcaras and Novak Djokovic is further analyzed, and its effectiveness is verified using the match results. Further, to evaluate index influence on momentum, this study conducted an XGBoost and decision tree algorithm performance, based on analyzing the interpretability displayed shapes. Finally, the XGBoost model with better performance is selected for deep training and iteration in this study. To better understand the influence of each momentum index on the prediction results, this study uses the SHAP model to analyze the correlation of the training set data. Among them, serve advantage and breakpoints are the two most influential indicators, and they contribute the most to the prediction of momentum fluctuations. The dynamic weight Wdy-serve is particularly helpful in capturing the fluctuations in a player's serve performance during a match. The analysis results of this study show that the change of momentum is not only closely related to the technical performance of the players but also significantly related to psychological factors.

Keywords: Momentum evaluation Model; XGBoost Model; Machine Learning; SHAP model; decision tree.

1. Introduction

In recent years, with the popularization of tennis and the continuous progress of technology, how to scientifically evaluate and analyze the performance of players in a match has become one of the important directions of sports science research. MOMENTUM is an important indicator that reflects the change of a player's state during a game [1]. It refers to the progress of an athlete on the psychological and physical level [2], and sometimes even leads to the reversal of the game result. More and more scholars and coaches are concerned. Traditional momentum evaluation methods often rely on static weights and simple statistical data, which cannot fully reflect the dynamic changes in players' states and the influence of multiple factors during the game. Therefore, it is essential to establish a more accurate and dynamic momentum evaluation model.

To deeply research the momentum of the tennis match, this study USES the XGBoost model and decision tree model to predict the momentum change XGBoost improves the decision tree algorithm is a kind of gradient, first put forward by Chen and Guestrin, widely used in various kinds of prediction problem, These include areas such as finance, healthcare, and sports analytics. Both of these models perform well in analyzing and predicting complex data patterns, so they are very suitable for the study of momentum. At the same time, this study also applied the SHAP model to analyze the correlation of momentum indicators in the training data set, to identify the key factors that have a significant impact on the fluctuation of the prediction results.

This work first collects and handles the relevant data, to the momentum of the tennis match, change is proposed based on a key indicator of the momentum evaluation model. Next, this study calculates the momentum difference sequence and calculates the cumulative sum of the momentum difference using the CUSUM detection algorithm to determine the turning point of the momentum change. At the same time, this study used a Run test to evaluate the randomness of momentum change and found that the change of momentum of players was affected by some key factors. Then, this study divided the data set into a training set and test set according to the ratio of 7:3, constructed the XGBoost and decision tree model [3], Different metrics have varying degrees of impact on players' motivation during an innings or longer game. For instance, when there are more match points, it is relatively more important for a player to win a serve than at the beginning of the match. Simultaneously, as the game progresses and wins accumulate, the effect on the athlete's momentum differs.and finally discussed the application prospect of the model.

2. Data Source and Preliminary Analysis

The data of this study come from the website as shown in Table 1:

Table 1. Websites used in this paper

Database Names	Database Websites
Google Scholar	https://scholar.google.com
Comap	https://www.comap.com
Sofascore	https://www.sofascore.com

For the given data set, the Wimbledon_featured_matches data table was found to have some missing data. Supplement of missing values: Given the existence of some missing data on service speed, this study uses patterns as the filling of missing values. If the missing value here represents a service error or out of bounds, 0 is used in this study, while forehand and backhand are quantified with 1 and 2 labels, respectively. At the same time, the motion indicators (such as service direction, service depth, etc.) are quantified by labels. The direction is a position noun, so this study quantifies its labels, B, BC, W, BW, and C are quantified by 1, 2, 3, 4, and 5 respectively.

3. Dynamic Weighting Approach to Momentum Analysis

3.1. Model establishment

Referring to the previous definition of momentum in sports [4], the concept of momentum used in this paper includes psychological effects and physical effects. Among them, psychological improvement refers to the positive change in the athlete's cognition, and physiological improvement refers to the positive change in the behavior. At the same time, through the previous study [5][6], it was reported that the level of reliability test (k= 0.64) when distinguishing forced errors from unforced errors was lower than that when considering all errors simultaneously (k= 0.89). So, at the moment, there is no difference between forced and unforced errors. In addition, in the study of Moss et al [7], it was found that in professional tennis, the dominance of the serve and the rotation of the player's Wdy-serve make long sequence scores of the same outcome less likely to be observed. Therefore, this paper will not conduct an in-depth study of the changes in the long-term momentum of men's singles tennis matches, and only briefly analyze and evaluate the changes in the long-term momentum of players when necessary. This paper decided to select the following indicators to build the momentum evaluation model of athletes (Table 2):

Table 2. Athletes Momentum Assessment model-related indicators

Symbol	Description	The direction of influence
$S_{serve,\ i}$	Serve advantage	+
$S_{sets,\ i}$	Number of winning sets (i = 1,2)	+
$G_{games,\ i}$	The number of winning games	+
$P_{points,\ i}$	Difference advantage	+ / -
$B_{break,\ i}$	Breakpoint	+
$U_{unf,\ i}$	Unforced errors	-
$S_{win,\ i}$	Winning points advantage	+

3.2. Indicator description

3.2.1. Serve advantage（Sserve,i）

In the given data set, we counted the proportion of the final victory / serve times and the final victory/catch times of all the players in the 7284 matches. From Figure1 (A) we found that the proportion of the serving side to win the game (68.6 %) is significantly higher than the proportion of the catcher to win the game (50.3 %), that is, the probability of the serving side to win the points/innings is much higher, so we take whether the player is the serving side as one of the positive evaluation indicators.

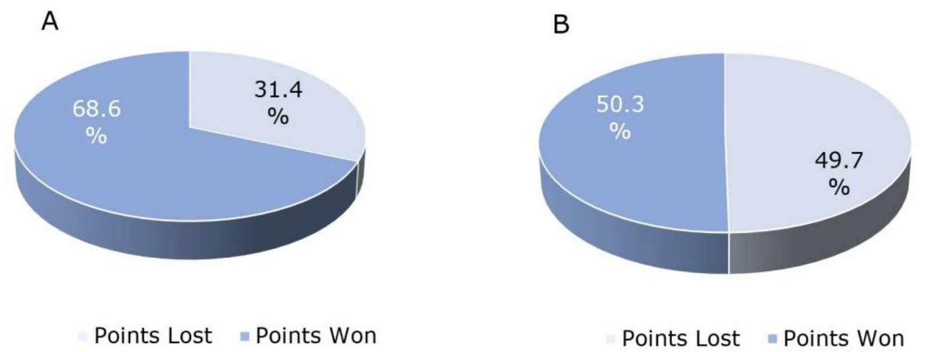

Figure 1. (A)Serve points ;(B)return points

We use Wserve as the weight of the serve advantage for the victory of the game. The initial weight coefficient of Wserve is 1.1, but considering that the serve advantage has different effects on the momentum of athletes in different periods of a game, this paper innovatively sets Wserve as a dynamic weight that changes with the game, that is, Wserve adds a dynamic weight value Wdy-serve that can change with the game process, so the final serve advantage is reflected in the player momentum change model as follows :

$$S'_{serve,i} = S_{serve,i} \times W_{serve} \times W_{dy-serve} \tag{1}$$

3.2.2. he number of winning games（Ggames, i）

This paper argues that the more games won by athletes in a game, the more advantages they have in the game because the more games won means the stronger the ability of the athletes, and the stronger the belief of the final victory, that is, the greater the momentum score of the athletes, so the number of games won by the athlete in the game is also one of the positive evaluation indicators. At the same time, we also set the corresponding weight Wgames for this index.

$$G'_{games,i} = G_{games,i} \times W_{games} \tag{2}$$

3.2.3. Number of winning sets（Ssets,i）

Similarly, we also use the number of rounds won by the athletes in the game as one of the important positive evaluation indicators. We also set up the corresponding weight Wsets.

$$S'_{sets,i} = S_{sets,i} \times W_{sets} \tag{3}$$

3.2.4. Difference advantage（Ppoints,i）

In this paper, we take the score difference between the two players as the evaluation index of the momentum of the athletes in the short-term competition.

$$P'_{points,i} = P_{points,i} \times W_{points} \tag{4}$$

3.2.5. Breakpoint（Ssets,i）

In a game, winning the breakpoint usually marks a change in the pace of the game. At the same time, a broken serve can also have a great exciting effect on itself to a certain extent and hit the opponent.

$$B'_{break,i} = B_{break,i} \times W_{break} \tag{5}$$

3.2.6. Unforced errors（Uunf,i）

When there is an unforced error, the probability will hurt the momentum of the athletes. As Figure1 (B) it can be seen that the greater the number of mistakes, the worse the stability and pressure management ability of the players during the game, which ultimately leads to lower scores. Therefore, we take the unforced error of athletes as an important negative indicator.

$$U'_{unf,i} = U_{unf,i} \times W_{unf} \tag{6}$$

3.2.7. Winning points advantage（Swin,i）

Whether it is at the beginning of the game or a certain point during the game, winning points have a significant positive impact on athletes. Additionally, the number of winning points also reflects the players' technical advantages and aggressiveness.

$$S'_{win,i} = S_{win,i} \times W_{win} \tag{7}$$

Whether it is at the beginning of the competition or some point during the competition, winning points has a significant positive effect on the athletes. In addition, the number of winning points also reflects the technical advantage and aggression of the players.

3.3. The basis of dynamic weight calculation

Different metrics have different effects on a player's motivation in a match, for example, winning the serve is relatively more important for a player when there are more points in the match. At the same time, the impact on a player's momentum varies as the game progresses and wins accumulate.In the following, the dynamic weight of the preemptive advantage Sserve, i is explained and illustrated in detail in this study. Based on the previous description, Wdy-serve is used as the dynamic weight of Sserve, i. At the same time, this study also carefully learned the rules of tennis. Finally, a score of 40 is taken as the peak value to measure the state of the player in a game, which is the default value when the player's serve, i is affected by his current score. When the score reaches 40 points, the player's serve advantage is the largest. At the same time, considering that the two sets of players to the score of 40:40 will be the first to fight for the right to serve, this match will determine the final serve of the match who serves first. Based on the above analysis, the Wdy-serve calculation is based on the following:

$$W_{dy-serve} = 1 + \left(\frac{P1_{score}}{40 \times \left(1 + 0.2 \times \left(AD_{count} - 1 \right)\right)} \right) \tag{8}$$

3.4. Integrated Momentum Evaluation Model

Through the selection and processing of the above indicators, this study establishes a model that can quantify the momentum change of players in the game. The model not only takes into account the traditional score and outcome but also includes key indicators such as breakpoints and unforced errors, which calculate momentum more.

$$\begin{aligned} M_i = {} & S_{serve,i} \times W_{serve} \times W_{dy-serve} + G_{games,i} \times W_{games} + S_{sets,i} \times W_{sets} + P_{points,i} \times W_{points} \\ & + B_{break,i} \times W_{break} + U_{unf,i} \times W_{unf} + S_{win,i} \times W_{win} \end{aligned} \tag{9}$$

Through the selection and treatment of the above indicators, this study developed a model that can quantify the momentum change of the players in the game. The model not only considers the traditional scores and wins and losses but also includes key indicators such as breakpoints and unforced errors, which calculates momentum more comprehensively and accurately. On this basis, this study can be al Callas and Djokovic's final momentum analysis model of the third inning, as shown in Figure 2.

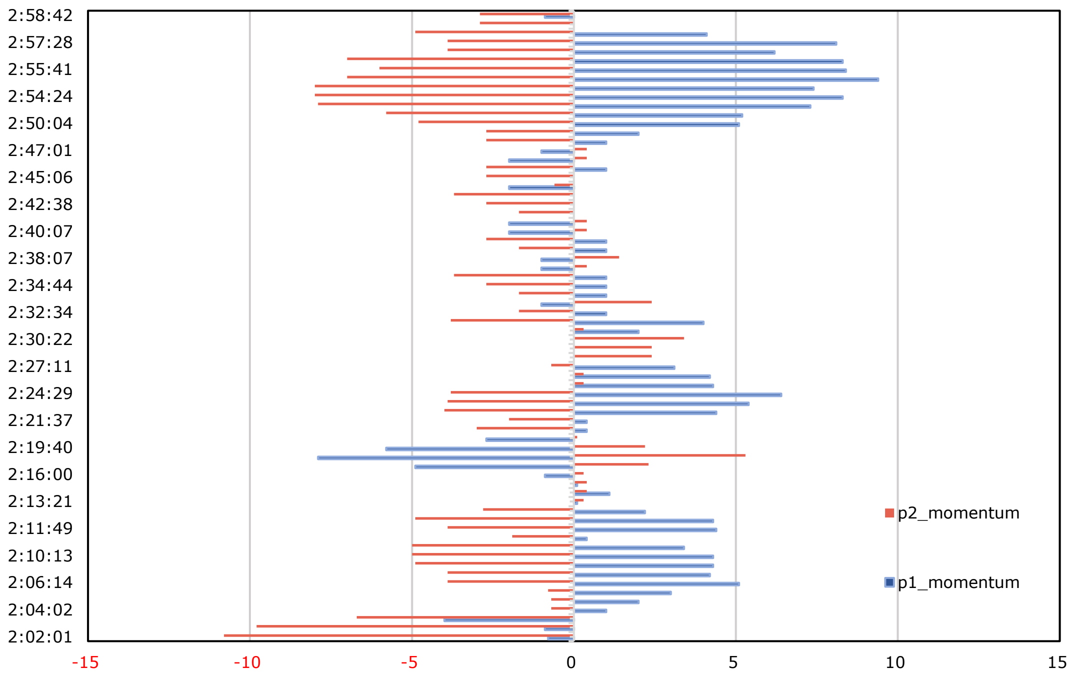

Figure 2. The momentum analysis results of the third set of the finals

With the known match information, this study can show that Carlos Alcaras easily defeated Novak Djokovic 6-1 in the third game of the deciding set. Through the above analysis chart of momentum evaluation, this study can also see that the momentum of Carlos Alcaras is rising or even far ahead of his opponent in most of this game. It can be verified that the model can indeed predict the momentum of the tennis player in the match. The following is a specific analysis of the performance of the athletes during the whole game using the model.

3.5. Analyze the degree of performance over a given period

Based on the momentum evaluation model tested above, the data of the entire final was selected for momentum analysis in this study to show the performance of the players at a specific time. From the specific data, due to the influence of various factors, the momentum of players on both sides has positive and negative values, that is, they will not always be in a dominant position. This fluctuation may be related to changes in their physical condition, mental state, or game strategy. In this study, a region map was also drawn to visualize the data (Figure 3).

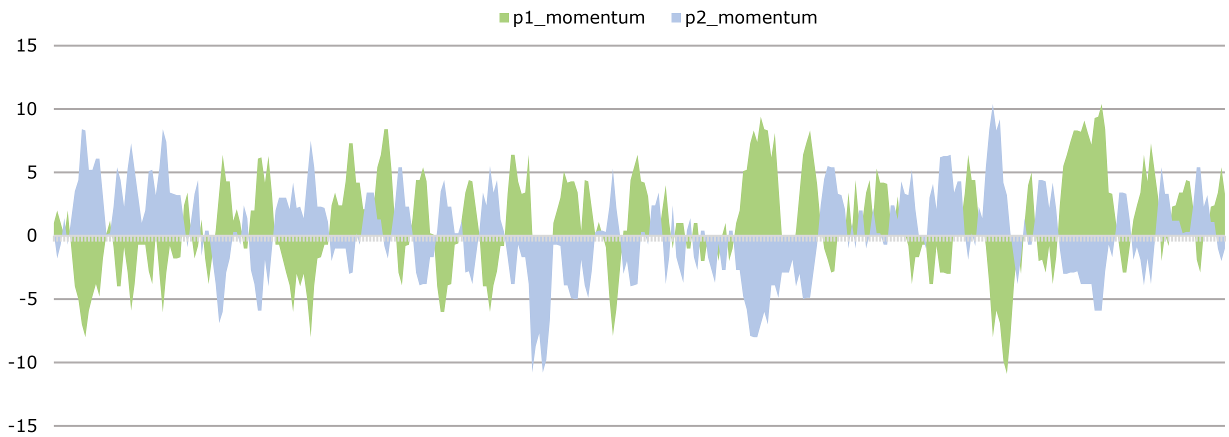

Figure 3. The momentum analysis results of the finals

Overall, the momentum of the two players fluctuated throughout the game, and there were many intersections. This means that the lead has changed hands many times, which is a common characteristic of tense, competitive games. At some characteristic periods, the momentum changes from negative to positive or positive to negative with large fluctuations, which means that the game may have a key turning point. For example, the momentum of Carlos Alcaraz increased sharply in certain periods, indicating that he stood out in those periods. Conversely, a sharp drop in momentum could be a mistake or a game-winner played by the opponent. Given that momentum is calculated taking into account statistics such as points, serves, catches, etc., this study can also infer match-specific situations from these fluctuations, such as which player is better at serving or has an advantage on the longboard. For example, in terms of individual metrics, champion Carlos Alcaraz has 66 wins, while Novak Djokovic has only 32.

4. Impact of Momentum Indicators on Prediction Results

Furthermore, to help athletes identify the core factors of momentum and help coaches provide better advice to cope with the fast-paced game, this study intends to use the ML model to evaluate the influence of factors on momentum. Specifically, this study uses the XGBoost machine learning algorithm and decision tree regression algorithm to train and iterate on the data from the previous competition (2023 Wimbledon). Common steps The two algorithms can be roughly divided into the following steps: (1) The dataset is divided into a training set and a test set according to the ratio of 7:3, and the momentum indicators are trained and predicted on the currently known data. (2) According to the prediction results obtained by the two algorithms, the accuracy of the test set data is tested, and the advantages and disadvantages of the two models are evaluated by MSE, RMSE, MAE, MAPE, R2 and other indicators to determine the optimal algorithm and conduct subsequent interpretability analysis.

4.1. Application of XGBoost model in momentum prediction

Extreme Gradient Boosting (XGBoost) is an efficient gradient-boosting decision tree algorithm [8]. As a forward addition model, its core is to use the ensemble idea UBoosting idea, using multiple trees to make decisions together and using the results of each tree. The difference between the target value and the prediction results of all previous trees is accumulated to obtain the final result, to improve the effect of the whole model. At the same time, the advantages of decision trees (such as efficiency, accuracy, flexibility, and interpretability) are played, which is the reason why it is selected as the research algorithm in this study. Due to the complexity of the XGBoost model, here this section does not discuss the related optimization of other algorithm-derived models [9] but constructs the basic model architecture for tennis momentum index prediction.

4.2. Comparative analysis of prediction models

Based on the above model, factors such as winning the ball, scoring, and serving of P1 players were taken as parameters of each leaf node and training tree in this study. In this study, the XGBoost regression algorithm model and decision tree model are used to obtain the distribution of the training set and the predicted distribution points.

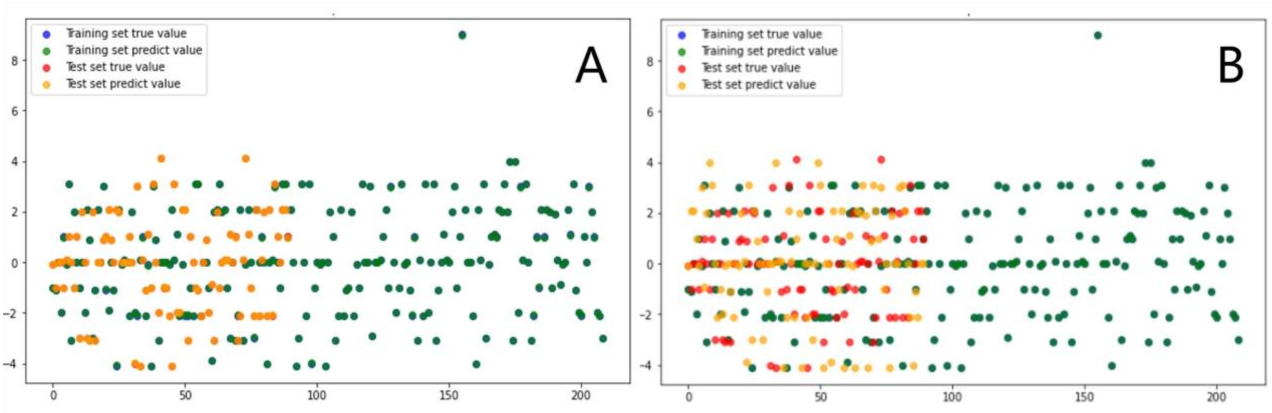

Figure 4. Prediction Results Comparison of XGBoost (A) and Decision Tree (B) Models

As shown in Figure. 4, the red and blue points are not lost but are covered by the predicted values. The blue and green dots represent the true and predicted values of the training set, respectively. According to observation, the distance between the blue point and the green point is very small, which means that the difference between the predicted value and the true value is not too large. At the same time, the data of the test set is also very satisfactory. However, compared with the results of the XGBoost model, the prediction results of the decision tree model [10] (Figure 4) are not as good as the former in distribution. Therefore, this study selects the XGBoost model with better performance to make more in-depth predictions and perform training iterations on more data and uses the XGBoost training set data to use the SHAP model to analyze and explain the correlation of momentum index data [11].

4.3. SHAP models analyze the impact of key indicators

Having identified the optimal algorithm, this study underscores the pivotal contribution of the XGBoost model in predicting the momentum index of tennis matches. Furthermore, the study integrates the SHAP (SHapley Additive exPlanations) model to provide a deeper understanding of feature contributions, thereby enhancing the interpretability and reliability of the momentum index predictions.SHAP is a way to explain the predictions of machine learning models [12]. Based on the Shapley value in game theory, each index of the model is assigned a value to represent the contribution of the feature to the prediction result. The core idea of the SHAP value is to decompose the prediction of the model into the marginal contributions of each metric[13]. It provides a quantitative way to assess the importance of metrics and can explain how models generate predictions based on combinations of input metrics [14], the results are shown in Figure 5.

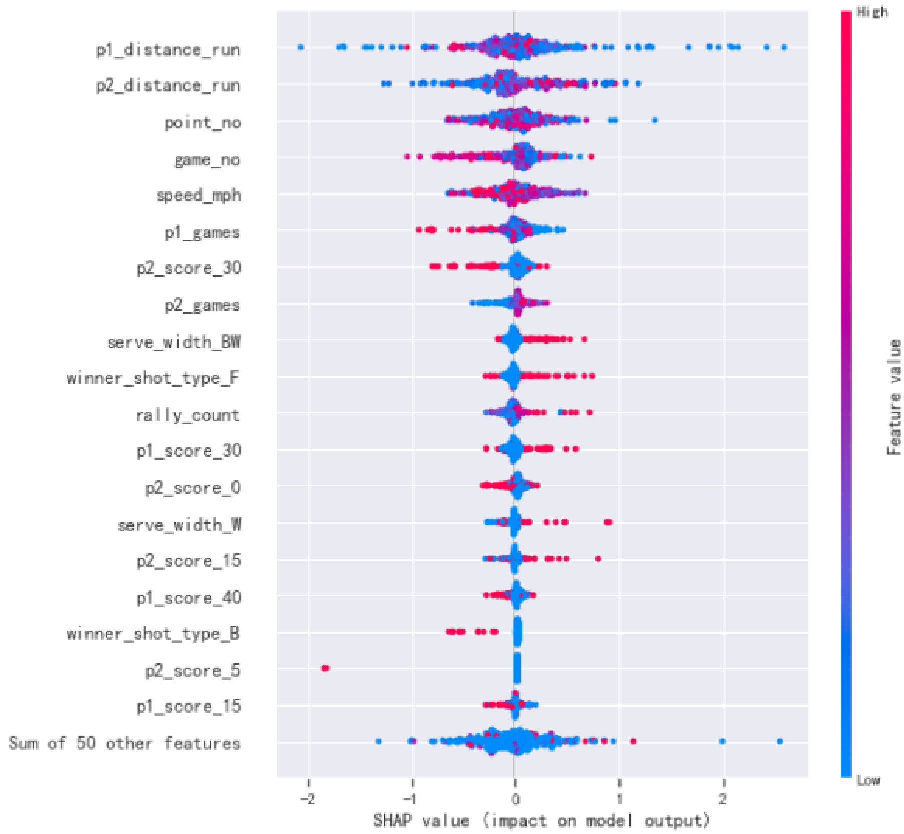

Figure 5. Summary Plot of features

Overall, the impact of negative indicators is greater than the impact of positive indicators. At the same time, from the specific values, p1_distance_run and p2_distance_run have the largest impact on the momentum fluctuation, which indicates that when the distance between the two athletes is large, in addition, to the sum of the other 50 features, it can be seen that most of the remaining indicators hurt the prediction of momentum fluctuation. This suggests that athletes are affected by several other disadvantages, as well as cumbersome little indicators of momentum fluctuations (such as sudden drops). This could be bad news for athletes, who are hard to predict. Subsequently, this study also carried out quantitative statistics on which indicators had a greater impact on the prediction results and drew a data visualization diagram (Figure 6). From Figure 6, it is clear from this study that, The eight metrics pl_distance run, p2_distance_run, p1_score, service width, point_no, game_no, p2_score, speed_mph have a strong correlation with the fluctuation of the prediction results.

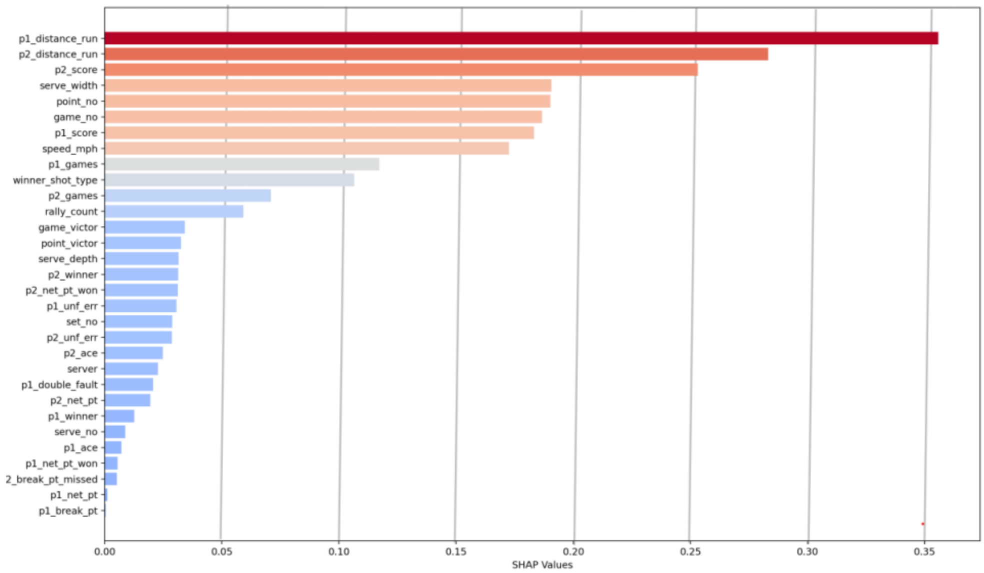

Figure 6. SHAP values with color gradient effect

4.4. Advise the players to enter the new competition

Based on the above analysis, this study takes the strategy of Carlos Alcaras against Novak Djokovic in the new match as an example[15]: 1. Develop the ability to score winning points: Winning points are the key to Alcaras' potential, so you'll want to focus on his attacking play and use his strength and speed as much as possible. At the same time, it is also important to maintain a good service style, such as forehand serve. 2. Reduce running distance in the game: It is clear from this study that running distance has a significant negative impact on motivation, so it is necessary to develop methods to deal with long passes to maintain energy 3. Minimize turnovers: Turnovers (such as unforced errors) have a great negative impact on momentum, so Alcaras needs to control high-stakes shots in practice and remain calm at critical moments in the game. 4. Flexible strategy: Adjust the strategy flexibly according to the progress of the game and the state of the opponent. This includes real-time assessment of the opponent's weaknesses and his state in the game, as well as the coach guiding the adjustment of tactics[16].

5. Conclusions

This study first presents an evaluation model for quantifying momentum changes during tennis matches. By analyzing seven key indicators, such as serve advantage and breakpoints, and assigning dynamic weights to them, this study successfully captures the momentum fluctuations of players during matches. In this study, the XGBoost model and decision tree model were used to predict player momentum, and it was found that the XGBoost model significantly outperformed the decision tree model in prediction accuracy and explanatory power on training and test sets. In particular, the XGBoost model performs well in multiple performance metrics such as R^2, MSE, RMSE, and MAE, which can explain most of the variability of the training data. This study further compares the performance of different prediction models[17]. The XGBoost model shows stronger generalization ability and prediction accuracy by its ability to capture complex data patterns. In contrast, there is a larger gap between the predicted results and the actual values of the decision tree model. Through these analyses, this study not only verifies the superiority of the XGBoost model but also provides a valuable reference for momentum analysis in other sports events in the future. This study proved that the designed framework and method in the analysis of momentum change in effectiveness, especially in the tennis match. By using the XGBoost model to predict the momentum fluctuations of players[18], this study demonstrates the superior ability of the model to capture complex data patterns. This analysis method not only helps to understand and predict the performance changes of athletes in different sports, but also provides solutions with practical application potential for the field of sports analysis, and further reveals the key factors of dynamic changes in sports competitions.

Acknowledgement

Funding for the Education and Teaching Reform and Research Project at Beijing Institute of Petrochemical Technology,Project approval Number: YB202304001"

References

[1] Briki W. Rethinking the relationship between momentum and sports performance: Toward an integrative perspective [J]. Psychology of Sport and Exercise, 2017, 30: 38-44.

[2] Dietl H M N C. Momentum in tennis: controlling the match[J]. International Journal of Sport Psychology, 2017, 48(4).

[3] Jayapradha J, Abdulsahib G M, Khalaf O I, Prakash M, Uddin M, Abdelhaq M, Alsaqour R. Cluster-based anonymity model and algorithm for 1:1 dataset with a single sensitive attribute using machine learning technique[J]. Egyptian Informatics Journal, 2024, 100485.

[4] Crespo M, Gallego M R ,Filipcic A .Determining the tactical and technical level of competitive tennis players using a competency model: a systematic review[J].Frontiers in Sports and Active Living,2024,61406846-1406846.

[5] Bret J J ,Gorce P .Kinematics characteristics of key point of interest during tennis serve among tennis players: a systematic review and meta-analysis[J].Frontiers in Sports and Active Living,2024,61432030-1432030.

[6] Song Y ,Lin H,Wang B, et al.Factors and Measures Affecting Players' Psychology in Tennis[J].Journal of Social Science Humanities and Literature,2024,7(3):

[7] Sampaio T ,Oliveira P J ,Marinho A D , et al.Applications of Machine Learning to Optimize Tennis Performance: A Systematic Review[J].Applied Sciences,2024,14(13):5517-5517.

[8] Gao X, Yu J, Hu Z, Hu Y. Feature Sorting Algorithm Based on XGBoost and MIC Combination Model[J]. International Journal of Advanced Network, Monitoring and Controls, 2021, 4: 79-88.

[9] Zhang H, Qiu D, Wu R, Deng Y, Ji D, Li T. Novel framework for image attribute annotation with gene selection XGBoost algorithm and relative attribute model[J]. Applied Soft Computing Journal, 2019: 57-79.

[10] Lai J, Liu C, Lin B, Miao D. An improved decision tree algorithm based on hierarchical neighborhood dependence[J]. Soft Computing, 2024, 1-11.

[11] Wang Z, Pang J, Tao P, Ji Z, Chen J, Meng L, Ming D. Locomotion transition prediction at Anticipatory Locomotor Adjustment phase with SHAP feature selection[J]. Biomedical Signal Processing and Control, 2024, 106105.

[12] Ergün S. Explaining XGBoost predictions with SHAP value: a comprehensive guide to interpreting decision tree-based models[J]. New Trends in Computer Sciences, 2023, 1(1): 19-31.

[13] Moss B, O'Donoghue P. Momentum in US Open men's singles tennis[J]. International Journal of Performance Analysis in Sport, 2015, 15(3): 884-896.

[14] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 785-794.

[15] Li Z ,Yin Y .A Quantitative Assessment Model for Momentum in Tennis Based on Exponential Decay and EWM-TOPSIS Method[J].Academic Journal of Computing & Information Science,2024,7(5):

[16] Zhu X ,Wang S ,Wu Z .Prediction of the Swings in Plays Based on Momentum of Tennis Matches[J].Frontiers in Sport Research,2024,6(3):

[17] Tang Y, Ma X, Wang S. Interpretable XGBoost-SHAP model predicts the nanoparticles delivery and reveals its interaction with tumor genomic profiles[J]. bioRxiv, 2022.

[18] Mitchell R, Frank E, Holmes G. GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles[J]. PeerJ Computer Science, 2022, e880.