Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080505
ISSN 2371-8838 Vol. 8 Num. 5

An Assessment and Prediction Model for Momentum in Tennis Based on EWM-TOPSIS and Random Forest Method

Yitian Yin*

School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an, 710072, China
**Corresponding author: ytyin@mail.nwp u.edu.cn*

Keywords: Random Forest, Momentum, Tennis

Abstract: In the realm of sports, the concept of "momentum" encapsulates the mechanism wherein athletes or teams, spurred by favorable factors within a competitive encounter, exhibit enhanced performance, thereby fostering a virtuous cycle of "success begetting success." The current research endeavors to dissect and analyze the momentum exhibited by tennis players, particularly utilizing empirical data stemming from the 2023 Wimbledon Men's Singles Final. The study's primary objective is to quantify this momentum and delve into its potential impact on player performance. This study analyzes momentum in tennis by developing the Player Performance Evaluation Model, based on Entropy Weight Method and TOPSIS evaluation algorithm. The study incorporates factors like winning status, match lead, movement distance, winning shots, and double faults, differentially weighing the winning incentives for servers and receivers and uses an exponential decay accumulation of evaluation indicators, akin to the Momentum algorithm in deep learning. Through binomial testing, the study builds a significant correlation between momentum score and win rate fluctuations and focuses on quantifying momentum and determining its influence on player performance. The Momentum Advantage Prediction Model based on Random Forest instead of LSTM model, predicts the next play's momentum advantage from previous moment data. The model attained accuracy 84.7%.

1. Introduction

Markman and Guenther (2007) presented a theoretical framework to illuminate laypeople's comprehension of psychological momentum[1]. They delineate velocity as having a positive or negative valence, depending on whether it is directed towards (advancing) or away from (retreating) a goal. The mass, on the other hand, is shaped by contextual variables that impart significance, urgency, and weight.

The influence of momentum in sports is profound, spanning both psychological and physical domains. It bolsters team or athlete confidence, motivation, and focus, thereby exerting a psychological edge over opponents. Momentum shifts initiate self-reinforcing cycles that have ramifications for team dynamics and coaching methodologies [2-3]. Ultimately, momentum transcends

the confines of the physical realm of the game, shaping outcomes through its influence on the mental and emotional fabric of competition.

In the context of men's tennis, momentum exerts a substantial influence on match outcomes, particularly through its impact on individual performance. Positive momentum fosters a heightened sense of confidence and aggression, placing pressure on opponents and potentially precipitating errors. Conversely, negative momentum diminishes confidence, thereby affecting decision-making process and executional capabilities [4]. Therefore, effectively capitalizing on or recovering from momentum shifts is paramount for achieving success in men's tennis.

2. Dataset preparation and preprocessing

2.1 Dataset preparation

The dataset (https://github.com/JeffSackmann/tennis_slam_pointbypoint) provides detailed information regarding the settlement of each point in the game, including key metrics such as p1_score (representing player 1's score within the current game), server (indicating the server of the point), and point_victor (identifying the winner of the point). To address instances of missing or incomplete data within the match dataset, imputation techniques are utilized for managing such gaps. Specifically, the imputation process involves replacing missing values in 'speed_mph' (representing the speed of serve in miles per hour) with the mean value, while missing values in 'serve_width' (indicating the direction of serve), 'serve_depth' (representing the depth of serve), and 'return_depth' (describing the depth of return) are imputed using the mode.

2.2 Feature extraction

In the dataset, various features are extracted for each score point, which are believed to play a significant role in momentum incentives for athletes. These features can be divided into two sets for players on both sides of the competition for each score point:

(1) Leading in Sets, Games, and Points: Being ahead in sets, games, or points can have a substantial impact on a player's momentum. This advantage provides a psychological boost of confidence and control, as well as a strategic edge in dictating the pace and exerting pressure on the opponent.

(2) Running Distance: The distance covered during play is crucial as it directly affects a player's momentum. Increased running depletes energy levels, influencing shot precision, reaction time, and mental concentration, serving as a key indicator of physical and mental exertion.

(3) Deciding Point Winner: Winning a deciding point holds significant importance for a player's momentum. Success in this critical moment not only boosts immediate confidence but also influences the player's mindset and overall performance, potentially leading to a shift in momentum.

(4) Ace Balls: Ace serves disrupt the opponent's rhythm and provide a strategic advantage, impacting a player's momentum by boosting morale and unsettling the opponent both psychologically and strategically.

(5) Double Faults: Committing a double fault not only results in losing a point but also forfeits the server's advantage, causing a notable psychological setback and disrupting the flow of the game.

(6) Unforced Errors: Unforced errors can disrupt a player's mental state, leading to frustration, loss of confidence, and a negative shift in momentum during a match.

(7) Winning a Game on Opponent's Serve: Winning a game while the opponent is serving can significantly impact a player's momentum both psychologically and strategically by boosting confidence, focus, and motivation.

(8) Victory Incentive: This feature is set according to the outcome of winning scenarios in a match

and is labeled for further analysis. For the serving player winning a match, this value is set to $\frac{1}{p}$; for the receiving player winning a match, it is set to $\frac{1}{1-p}$; and for non-winning scenarios, it is set to 0 (where p represents the prior probability of the serving player winning, empirically obtained from each data point, with a value of 0.6731).

The above calculated features are then labeled a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 in turn. The value of the athlete's feature i at point number t is denoted as $a_i(t)$.

2.3 Further Processing

Consider the impact of an athlete's historical athletic accomplishments on their present performance, noting that recent events hold greater sway while past achievements hold diminishing relevance. To quantify this diminishing influence, an exponential decay model is employed to illustrate how the effect of an event decreases exponentially over time. The formula is as follows:

$$S_i(\mathrm{t}) = \lambda \cdot a_i(\mathrm{t}) + (1-\lambda) \cdot S_i(\mathrm{t}-1) \quad (1)$$

Given that $0 < \lambda < 1$, S_i represents the cumulative result of a_i after undergoing exponential decay. In our evaluation model, the model substitute S_i for a_i as the eigenvalue used. The decay constant accurately signifies the rate at which the influence of prior events wanes within the specific framework of a tennis match. A higher value of λ implies a swifter decay of influence, indicating that the impact of an event diminishes more rapidly. Through an analysis of historical match data, the value of λ = 0.3 was derived empirically, aligning closely with the observed patterns of momentum fluctuations and successes in tennis.

Drawing inspiration from the Momentum-based Stochastic Gradient Descent (SGD) algorithm in deep learning, our methodology exhibits a conceptual similarity to the dynamics observed in tennis matches. In deep learning, momentum SGD involves utilizing prior gradient directions to enhance learning optimization. This concept is comparable to the notion of momentum in tennis, where a player's current game is influenced by preceding actions. Just as momentum in SGD facilitates smoother learning processes and hastens convergence, in tennis, players rely on recent successes or failures to shape their subsequent performances. This analogy underscores an intriguing convergence between computational optimization techniques and sports psychology, illustrating how principles from diverse domains can metaphorically align with each other [5].

3. Random Forest - LSTM - based on Momentum Advantage Prediction Model

3.1 The establishment of logistic model

At a specific moment t, assuming we have $X_1, X_2, \ldots X_{10}$ representing ten variables in a tennis match. Leading in set count (X_1), leading in game count (X_2), and leading in point count (X_3) contribute to partial momentum through linear addition while the remaining seven variables ($X_4, X_5, \ldots X_{10}$) are calculated recursively according to $S(t)=\lambda \cdot a(t)+(1-\lambda)\cdot S(t-1)$. $S(t-1)$, and then added together. Assuming λ is a given decay factor, $S(t-1)$ is the momentum value at the previous moment, and $a(t)$ is the current value of the variable.

The momentum calculation formula at moment t is given as:

$$Momentum(t) = X_1(t) + X_2(t) + X_3(t) + \sum_{i=4}^{10} (\lambda \cdot X_I + (1-\lambda) \cdot S_i(t-1)) \tag{2}$$

Here, $X_1(t), X_2(t), X_3(t)$ are observed values at moment t, and $S_i(t-1)$ is the momentum of X_i at the previous moment.

To calculate $S_i(t)$, we use the recursive relationship:

$$S_i(t) = \lambda \cdot X_i(t) + (1-\lambda) \cdot S_i(t-1) \tag{3}$$

Starting from $t=1$ up to the current moment t. If observations begin at $t=0$, $S_i(0)$ can be an initial value. To derive the formula for total momentum, combining linear addition and Exponentially Weighted Moving Average (EWMA):

$$Momentum(t) = a_1 X_1 + a_2 X_2(t) + a_3 X_3(t) + \sum_{i=4}^{10} a_i \{ [1 - (1-\lambda)^t \cdot X_I + (1-\lambda)^t \cdot S_i(0)] \} \tag{4}$$

Where $S_i(0)$ is the initial momentum value, λ is the decay factor. Simplified total momentum formula:

$$Momentum(t) = a_1 X_1 + a_2 X_2(t) + a_3 X_3(t) + 7a_i \{ [1 - (1-\lambda)^t \cdot X_I + (1-\lambda)^t \cdot S_i(0)] \} \tag{5}$$

This formula quantitatively analyzes and compares momentum at different time points, considering the cumulative effects over time.

3.2 Optimized model

By differentiating the dynamic performance scores of the two players in real time, we obtain the psychological momentum of each moment on the field. The larger the absolute value of momentum is, the player's momentum is more dominant at this moment [6-7]. By correlating momentum with points scored, it can be concluded that: when momentum is favourable, next 28 points performance is better than the original level, and the probability of scoring points consecutively increases significantly. The model answers the coach's question: when players of similar ability compete on the same match, it is easier for the player with superior momentum to trigger swings in play and runs of success by one player.

The momentum literature comprises two distinct forms of momentum: strategic momentum and psychological momentum. Although both theories predict that success breeds success, the rationale behind these two theories is different, Strategic momentum arises from the different relative positions of competing agents in a dynamic contest, which leads to asymmetric future expected prizes. In contrast, psychological momentum suggests that a precipitating event triggers a performance increase due to changes in the perception of the agents.

The momentum literature comprises two distinct forms of momentum: strategic momentum and psychological momentum. Although both theories predict that success breeds success, the rationale behind these two theories is different. Strategic momentum arises from the different relative positions of competing agents in a dynamic contest, which leads to asymmetric future expected prizes. In contrast, psychological momentum suggests that a precipitating event triggers a performance increase due to changes in the perception of the agents.

The paper believe that strategic momentum is closely related to a player's level of performance. In

contrast. Empirically, psychological momentum has been predominantly examined in sports setting. After visualizing the performance scores of each player, we can directly observe that when the difference in performance between the two players is too significant, the break performances (swing in game, runs of success) are still dominated by their ability. In order to reduce the error and to highlight the weight of momentum in triggering swing in game and runs of success, we initially built the model.

The paper use betting odds to estimate players' chances of winning a tennis match, as these odds include information like home advantage and injuries. The paper grouped players with similar odds and winning probabilities into seven categories for analysis. The paper focuses on understanding the role of psychological momentum in tennis [8]. The paper analyzed specific matches, looking at how players' performance varied point by point and how runs of success affected the game. The paper chose one player in each match to study their momentum during the game.

For this, it calculated the likelihood of this player winning the next 28 points in a match, considering their current momentum. Normally, with players of equal ability each has a 50% chance of winning a point. The paper compared this player's actual win rate over the next 28 points with the expected 50%. If their win rate was higher, it indicated positive momentum, and if lower, it suggested negative momentum.

The model chose to analyze 28 points because a smaller number would give unreliable results, while a larger number would show less variation. By dividing the total number of points where momentum was a factor by the total number of points played, the paper found that momentum was a significant factor in 60.48% of the points.

The model also improved our model by comparing the actual win rate per point instead of the assumed 50% rate. This also showed that momentum significantly affects a player's performance. The model examined the first 30 points of each game, as fewer points would make results unstable. This analysis also supported the idea that momentum plays a significant role in a player's performance in tennis, beyond their skill level.

Firstly, the model pre-process the key data and the momentum of the players. Using One-Hot Encoding, the category of untouchable shot, direction of serve discrete features are converted into a numeric form that can be handled by machine learning algorithms. This method expands the values of discrete features into a new binary feature, where each feature has only two possible values: 0 or 1. For features with n different values, One-Hot Encoding generates n new binary features, each corresponding to one of the original feature values. The data such as player 1's distance ran during point, speed of serve, where data with different scales or ranges are converted to a uniform standard range. Normalisation eliminates the differences in scale between different features in tennis match, making the model more stable and accurate. Considering that the momentum of a player at each moment should be related to the previous batting performance, we first introduce Long Short Term Memory networks (LSTMs), which process the previous performance characteristics to estimate the momentum of the next moment, thus achieving the prediction of the dynamic state of the player's play on the field. Combined with the evaluation model in the previous section, the momentum shift of a player can reflect a dramatic change in the points, thus, swings in the match [9].

Layers are arranged sequentially as follows: The initial layer is a single LSTM unit employing a hyperbolic tangent (tanh) activation function. Then, another LSTM layers introduced, with its activation function not specified. Subsequently, another dropout layer is added to maintain effectiveness against overfitting. The architecture concludes with a fully connected (dense) layer using a sigmoid activation function[10], which is apt for binary classification tasks. The optimization process utilizes the Adam optimizer recognized for its efficiency in reducing loss, particularly beneficial in intricate model sand extensive datasets.

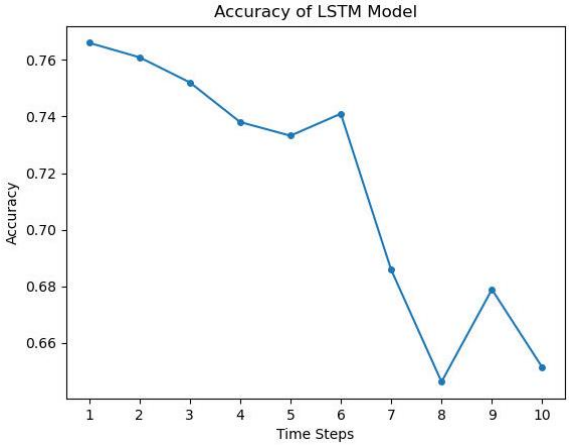

Figure 1: Accuracy of LSTM model

The model set the step size to 10 for the prediction of the fluctuation of the field in the next moment, and the accuracy is 65.13%. Next, the model shorten the step size gradually, and finally find that the model has the highest prediction accuracy of 76.66% when the step length is 1. The model has generated graphical representations of data illustrating the relationship between varying step lengths and the predictive accuracy of Long Short-Term Memory (LSTM) models [11-12]. It is evident that as the step length decreases, the predictive accuracy of the LSTM model increases.

In observations, as it shown in Fig.1, it was discerned that utilizing a reduced number of temporal steps actually yielded a heightened accuracy rate. Consequently, this led to the inference that the momentum shifts within a tennis match do not exhibit significant temporal correlations. This phenomenon can be conceptualized as each point in a tennis match representing a Markov process, where the momentum's alteration is predominantly influenced by the preceding state rather than by any extended historical sequence [10]. Hence, the paper used the Random Forest Model to predict the swing in game.

In the Random Forest Classifier, the feature importances attribute encapsulates the significance of each feature in the context of the prediction task. The computation of this significance is predicated on the assessment of how each feature contributes to the enhancement of the model's accuracy during the decision tree construction phase. Specifically, the calculation of feature importances adheres to the following methodology:

Subsequently, another dropout layer is added to maintain effectiveness against overfitting. The architecture concludes with a fully connected (dense) layer using a sigmoid activation function, which is apt for binary classification tasks. The optimization process utilizes the Adam optimizer recognized for its efficiency in reducing loss, particularly beneficial in intricate model sand extensive datasets.

For the cumulative importance, within each tree of the random forest, the aggregate sum by which each feature decreases impurity is computed. Subsequently, these values are normalized to ensure that the sum of importances across all features equals one [13]. Given that a random forest is comprised of multiple trees, the final feature importance is derived by averaging the importance scores of the same feature across all trees.

3.3 Model testment

In our study, the paper applied a dynamic momentum model to the 2023 Wimbledon Gentlemen's final, focusing on the intriguing match-up between the 20-year-old Spanish rising star Carlos Alcaraz and the seasoned 36-year-old Novak Diokovic. Our objective was to quantify the shifts in momentum within the match and to estimate the swings in game. The model's results are visually represented in

the accompanying Fig. 2.

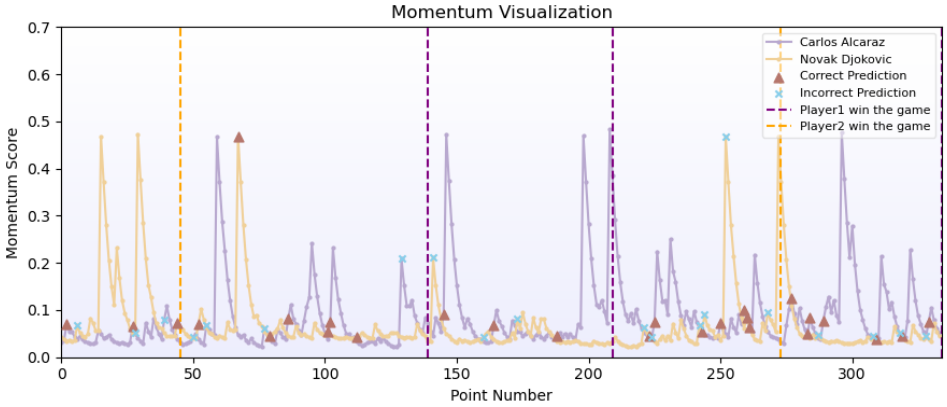

Figure 2: Match Flow for 2023-wimbledon-1701

This high-stakes match unfolded over five sets. The horizontal axis of the graph denotes the 334 points played during the entire match, while the vertical axis indicates the fluctuating momentum levels of the players. Vertical lines within the graph signify the winner of each set. Initially, in the first set, both players alternated scoring in the initial games, maintaining a similar momentum level. However, Djokovic gradually established a significant momentum advantage over Alcaraz, sustaining a higher level throughout the set and concluding it with a decisive 6-1 lead. This outcome might have been a strategic maneuver by Alcaraz, who quickly concluded the first set in 50balls, potentially using the inter-set break to recalibrate his strategy for the subsequent sets.

In the second set, the graph reveals a gradual increase in Djokovic's momentum extending the match's duration. After winning two consecutive games, the younger player reached his peak momentum in this set, followed shortly by a strong response from Djokovic, who also peaked in momentum, winning two games to even the score. Despite this, Alcaraz maintained a consistently higher momentum for the rest of the set, displaying two significant peaks, indicative of a swing in the game, and ultimately securing the set in a tie-breaker 7-6.

For the third set, Alcaraz capitalized on his momentum, leading initially and maintaining a slight edge over Djokovic. Following a peak in Djokovic's momentum and a subsequent game win, his momentum did not sustain at a high level. ConverselyAlcaraz's momentum remained elevated towards the end of the set, culminating in a 6-1 victory.

In the decisive final set, Alcaraz initially led in momentum, swiftly winning three games. However, Djokovic rapidly increased his momentum, consistently catching up in the score, marked by two significant peaks corresponding to his winning games. Despite alternating scoring later in the match, Alcaraz consistently held a higher momentum position, eventually leading the young athlete to seize control and secure a 6-4 victory. The Fig. 3 visualized the relative momentum during this match.

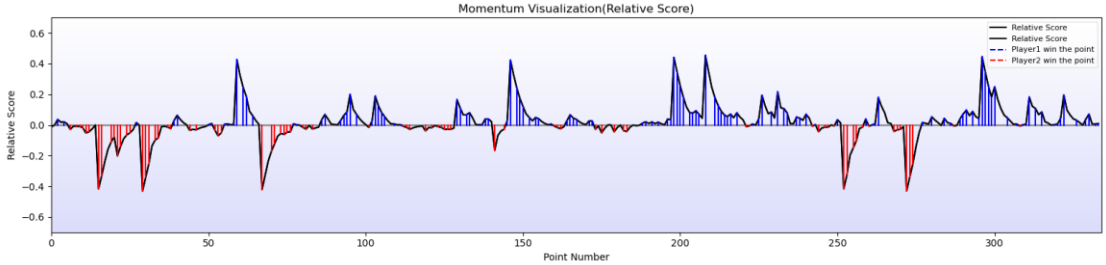

Figure 3: Relative Momentum Visualization of 2023-wimbledon-170l

By calculating the difference in dynamic momentum between the two athletes, the paper could

determine which player had the momentum advantage at any given moment. Projecting this dynamic score onto the player with the momentum advantage and corresponding scoring, our model accurately predicted swings in the game with a high accuracy rate of 84%.

4. Results and analysis

The importance of a feature essentially reflects its average efficacy in reducing impurity across multiple decision trees. Features with higher importance play a more pivotal role in the decision-making process of the predictive model. In study, we employed a Random Forest model to perform predictive analysis. Specifically, we utilized data from 22 matches as the training dataset, while the remaining 9 matches constituted the testing dataset. The paper objective was to employ the information from the previous time step to forecast which entity would possess a greater momentum in the subsequent time step. The results demonstrated a predictive accuracy of 84.6%.

The computational analysis of our model yielded conclusive results, indicating the factors that exhibit varying degrees of predictive capability concerning the swing in a match. These factors are ranked in descending order of importance as follows: player 2's momentum, player 1's momentum, whether the player won a game this point whether the player won a set this point, player 1's execution of an untouchable winning shot, player 2's score within the current game, and the speed of serve.

Additionally, the directional aspects of the serve, including direction of serve (Center), depth of serve (Close to Line or Not Close to Line), direction of serve (Body/Center), direction of serve (Body), and direction of serve (Wide), were observed to exert an influence on the prediction of the swing in the match, albeit to a lesser extent. The Table. 1 showed the correlation coefficient of various variable indicators.

Table 1: Correlation coefffcient of Various Variable Indicators

Factors indicate Swing in Game	Correlation Coefficient
player 2's momentum	0.336543
player 1's momentum	0.328546
whether the player won a game this point	0.019559
whether the player won a set this point	0.017516
player 2's score within current game	0.013628
speed of serve	0.012739
direction of serve	0.012246
Center depth of serve	0.011931
depth of serve: Close To Line or Not Close To Line	0.011366
direction of serve: Center	0.011206
direction of serve: Body	0.011141
direction of serve: Wide	0.011133

In addition to assessing feature importance through the observation of each feature's contribution to the accuracy of the model in the construction of decision trees within the Random Forest, as discussed previously, we also employed Partial Dependence Plots (PDPs) for a visual sensitivity analysis. As illustrated in the Figure. 4, PDPs were generated for features with higher importance, including the player's momentum score from the previous moment, recent wins in games or points, and recent successful shots and serves. The Y-axis represents the model's average predicted response, namely the probability of a class, while the X-axis indicates the values of individual features.

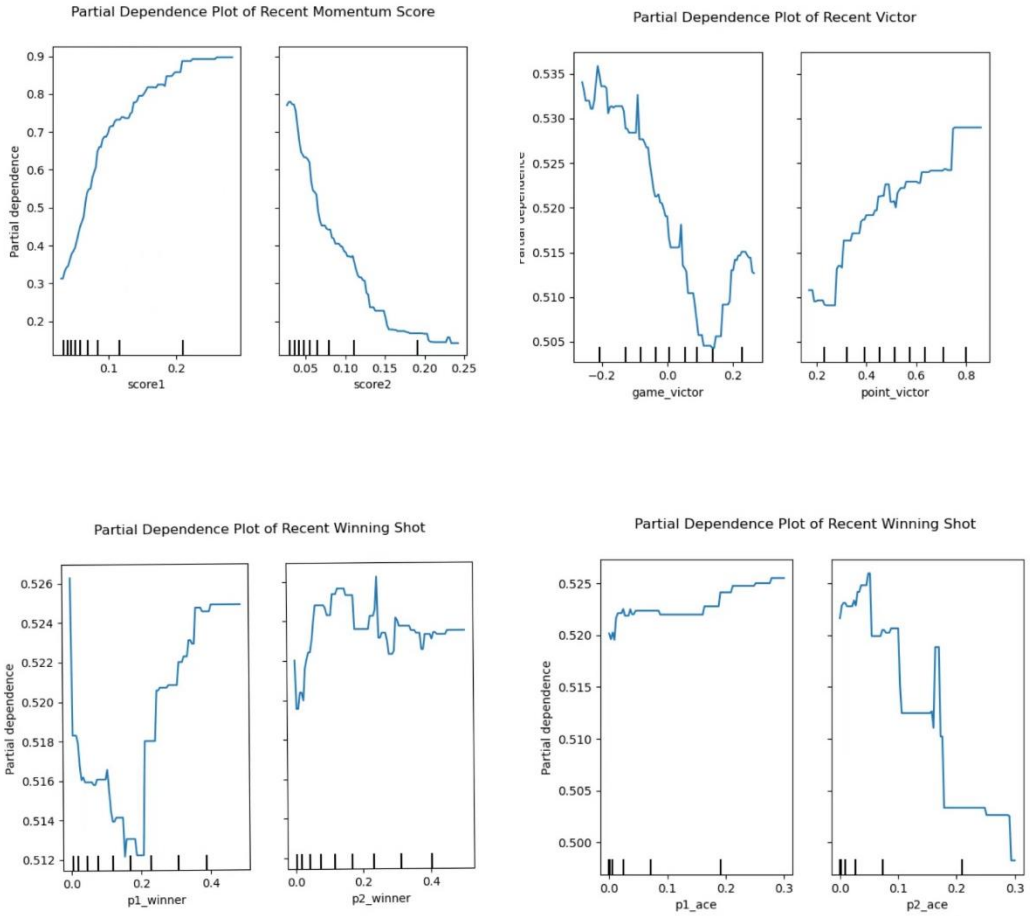

Figure 4: Sensitivity analysis results

These plots demonstrate how the model's average prediction response varies as the value of a particular feature changes within its range. This analysis conffrms a strong correlation between the a forementioned features and the predicted outcomes.

5. Conclusions

This paper conducts an analytical study on the momentum of tennis players based on the match data from the 2023 Wimbledon Men's Singles top 32 final. It quantifies the momentum score of a player at a certain scoring point. A Player Performance Evaluation Model was developed, taking into account various factors such as the player's win conditions, degree of lead in the match, distance covered, winning shots made, and double faults, as evaluative indicators. These scores were defined as the players' momentum scores, which were then utilized as a standard for visualizing the match conditions. The paper concludes by examining whether the advantage of momentum scores is linked to an increase in players' win rates. If a player's momentum score exceeds that of the opponent and is followed by an increased win rate in subsequent points, momentum is considered to be effective. In our study, the paper employed a Random Forest model on The Momentum Advantage Prediction Model to perform predictive analysis. Our model employs the information from the previous time step to forecast which entity would possess a greater momentum in the subsequent time step. Also, we applied a dynamic momentum model to the 2023 Wimbledon Gentlemen's final, quantifying the shifts in momentum within the match and to estimate the swings in game. The results demonstrated a predictive accuracy of 84.6%.

References

[1] Walker, M., Wooders, J, Amir, R. Equilibrium play in matches: Binary Markov games. Games and Economic Behavior, 2011, 71(2), 487-502.
[2] Meier, P., Flepp, R., Ruedisser, M., Franck, E. Separating psychological momentum from strategic momentum: Evidence from men's professional tennis. Journal of Economic Psychology, 2020, 78, Article 102269.
[3] Depken, C. A., Gandar, J. M., Shapiro, D. A. Set-level strategic and psychological momentum in best-of-three-set professional tennis matches. Journal of Sports Economics, 2022, 23(5), 598-623.
[4] Mago, S. D., Sheremeta, R. M., Yates, A. Best-of-three contest experiments: Strategic versus psychological momentum, International Journal of Industrial Organization, 2013, 31(3), 287-296.
[5] Romain Gauriot, Lionel Page, Does Success Breed Success? A Quasi-Experiment on Strategic Momentum in Dynamic Contests, The Economic Journal, Volume 129, Issue 624, November 2019, Pages 3107-3136.
[6] Den Hartigh RJR, Gernigon C. Time-out! How psychological momentum builds up and breaks down in table tennis. J Sports Sci. 2018 Dec; 36(23):2732-2737.
[7] Strumbelj, E. On determining probability forecasts from betting odds. International Journal of Forecasting, 2014, 30(4), 934-943.
[8] Chen, H.; An, Y. -c. Green Residential Building Design Scheme Optimization Based on the Orthogonal Experiment EWM-TOPSIS. Buildings 2024, 14, 452.
[9] Lin L, Wei S, Shuyu C, et al. A Dynamic Adaptive and Resource-Allocated Selection Method Based on TOPSIS and VIKOR in Federated Learning [J]. Neural Processing Letters, 2024, 56(2)
[10] Chen, T., Guestrin, C., He, X., & Garcia, E. XGBoost: Extreme Gradient Boosting with Random Forests. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(1), 342-356.
[11] Zhang, Y., & Chen, H. An Improved Random Forest Model for Credit Scoring. Expert Systems with Applications, 2022, 196, 116654.
[12] Wang, H., & Li, G. Random Forest Regression with Optimized Parameters for Stock Price Prediction. Journal of Computational and Theoretical Nanoscience, 2022, 19(5), 2154-2161.
[13] Liu, J., Zhang, R., & Wang, L. Random Forest-Based Classification of Remotely Sensed Images: A Review and Prospect. Remote Sensing, 2023, 15(10), 2345-2378.