Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2423845

Supremacy in Skill? Unveiling the Secrets of Tennis Player “Momentum”

Summary

This study delves into the elusive concept of ”momentum” in the field of tennis, utilizing a comprehensive dataset that includes the 2023 Wimbledon Men’s Singles final, tennis player rankings, and various other data points.**momentum shifts** in matches. Employing an innovative dual-model approach, we integrate a **Back Propagation (BP) neural network** for its robust predictive capabilities and a **Random Forest model** for its insightful feature importance analysis.

Firstly, to obtain the formula for **momentum**, we normalized the data and then employed the **entropy weighting method** and the **analytic hierarchy process** to determine the weights of the primary and secondary indicators selected.

Subsequently, to verify whether the calculated **momentum** could indeed predict the performance of tennis players, we conducted **Pearson correlation analysis** and **Run test**. These tests analyzed the consistency and randomness of our model, yielding favorable anticipated results. This outcome supports the hypothesis that **momentum** plays a significant role in influencing the trend of a tennis match.

Following this, we adopted an innovative dual-model approach, integrating a **Back Propagation (BP)** and a **Random Forest model**, to achieve enhanced analytical effectiveness. The model captures the intricate nonlinear interdependencies prevalent in match dynamics and offers a nuanced understanding of factors influencing **momentum**.Moreover, the model’s interpretability is enhanced through the **Random Forest’s quantification of feature relevance**, offering strategic insights into player performance.

In terms of the model’s generality, we use women’s singles tennis matches as an example to showcase the limitations of the model in different types of competitions. We present improvements made to address these limitations and compare the differences before and after the improvements. Additionally,we provide guidance on how to apply and refine the model in other competitions. Navigating the inevitable trade-off between generality and accuracy, we make informed decisions to achieve a balance between the model’s performance on unknown data **(generality)** and known data **(accuracy)**

The research transcends mere prediction, delving into the potential applications of predictive modeling in sports. It provides a multidimensional analysis of **momentum**, integrating quantitative assessments with strategic insights, thereby offering a holistic framework for understanding and leveraging **momentum** in tennis.

Keywords: AHP;EWN;BP Neural Network;Random Forest;Box Plot;Pearson correlation analysis;Run test

Contents

1 Introduction

1.1 Problem Background

Tennis, a globally esteemed sport, originated in France before permeating into the UK and the USA, subsequently ascending to the status of the second most prominent ball game worldwide. In the 2023 Wimbledon Men's Singles Final, the youthful Spanish player Carlos Alcaraz garnered attention by triumphing over the legendary Novak Djokovic, thereby halting the latter's streak of victories. This match precipitated an in-depth discourse on the concept of "momentum," the invisible force purported to influence the dynamics of the game. Our investigation involved the development of a quantitative model aimed at predicting the trajectory of matches and examining the role of momentum within them. The ability to anticipate shifts in the momentum of tennis matches can aid players and coaches in timely adjustments to their state and competitive strategies, thereby enhancing their performance.

1.2 Restatement of the Problem

This study aims to explore the existence, influence, and measurability of the concept of momentum in tennis matches, with a particular focus on the 2023 Wimbledon Men's Singles Final. Our research will revolve around the following five principal tasks:

• Question 1:

Develop a model that reflects the varying performance levels of players while also encapsulating the advantage of the server.

• Question 2:

Apply the model to actual matches to assess and refute the notion that swings in play and successive triumphs by a single player are mere products of randomness..

• Question 3:

1. Employing data from at least one match, this study aims to develop a model to predict momentum fluctuations within the game. The research focuses on identifying which factors exhibit the strongest correlation with the variability in match outcomes.

2. Highlighting key influencing factors, while providing guidance for the players' upcoming match.

• Question 4:

Conduct an evaluation of the developed model across multiple matches to assess its predictive efficacy and areas for potential enhancement. Furthermore, investigate the extent of the model's generalizability across various types of matches, surfaces, and other similar sports.

• Question 5:

"Propose recommendations to coaches regarding the utilization of 'momentum' effects, emphasizing strategic training of players to address key events in tennis matches."

1.3 Our Work

Figure 1 shows the process of our work:

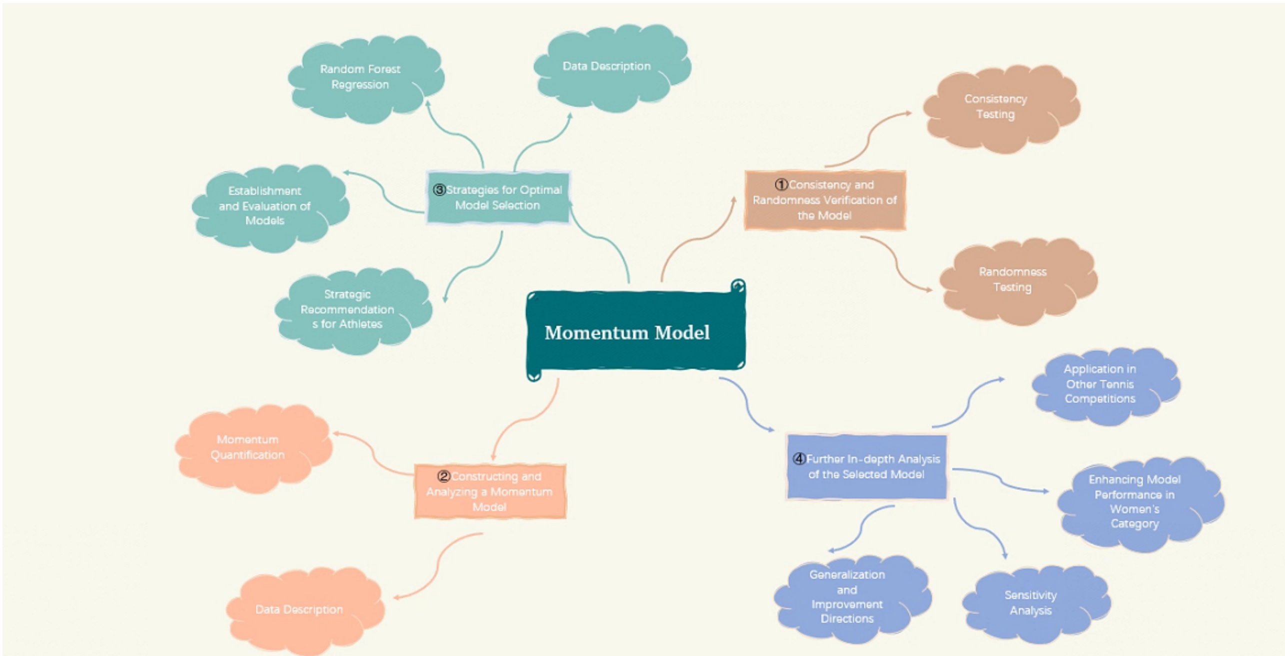


Figure 1: The Process of Out Work

2 Assumptions and Justifications

Considering that real problems always contain many complex factors, we first need to make reasonable assumptions to simplify the model, each followed closely by its corresponding interpretation. All assumptions will be re-emphasized once they are used in the construction of our model.

• **Assumption 1: The statistics we collect from the website are reliable and accurate.**

Justification: The availability of data is a fundamental issue. If the data itself is unreliable or untrue, we cannot make an effective assessment of "momentum". Therefore, continuity and authenticity of the data we obtain is very important. The data we collected mainly came from authoritative websites or open source websites such as `https://www.atptour.com/` and `https://github.com/JeffSackmann/tennis_atp`.

• **Assumption 2: Each player endeavors to achieve victory to the utmost extent of their capabilities.**

Justification: Sportsmanship, a fundamental concept in competitive sports, necessitates that athletes exhibit respect towards their adversaries, uphold the integrity of the game, and approach every match with the utmost seriousness in order to guarantee equitable competition.

•**Assumption 3: There will be no overwhelming disparity in skill levels among players.**

Justification: While skill is undoubtedly a crucial factor in determining the outcome of a match, excessive differences in skill levels can introduce significant interference in the discussion of 'momentum.' Therefore, we allow for relative differences between athletes but not to the extent of being overly disparate.

3 Notations

The primary notations used in this paper are listed in Table 1.

Table 1: Key notations used in this paper

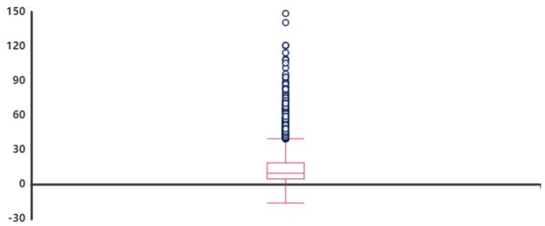
Symbol	Description	Unit
SMPH	Serve Speed	mph
SD	Serve Depth	-
SW	Serve Width	-
STB	Short-term Breaks in five Games	Count
LTB	Long-term Breaks	Count
STH	Short-term Holds in five Games	Count
LTH	Long-term Holds	Count
NC	Net Approaches	Count
NSR	Net Scoring Rate	%
RE	Running Return Rate	%
UEC	Unforced Errors in the Current Game	Count
BTS	Whether Serving	Boolean
SWG	Games Won in Current Set	Count
SGA	Score Lead Progress in the Current Game	Count

4 Task 1: Constructing and Analyzing a Momentum Model

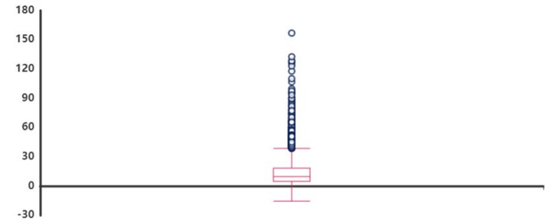
4.1 Data Description

The data utilized in this study encompasses the files provided in Problem C titled 'Wimbledon featured matches.csv.' Additionally, we sought relevant data from two websites, namely `https://www.atptour.com/` and `https://github.com/JeffSackmann/tennis_atp`, to supplement our analysis. But before using it, the data needs to be preprocessed.

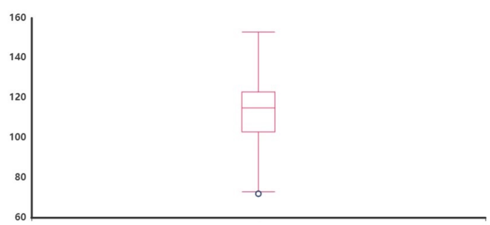
Step1: Data Anomaly Handling: The serve speed, direction, depth of the serving player, and the response strategy of the receiving player undoubtedly have a significant impact on the outcome of the match. After conducting a boxplot analysis, considering the presence of missing values in the entire dataset, such as all serve speeds being marked as NA (2023-wimbledon-1310), traditional interpolation methods are no longer applicable due to the independence of matches. Therefore, we excluded data with missing or anomalous values to ensure the consistency and completeness of the dataset.



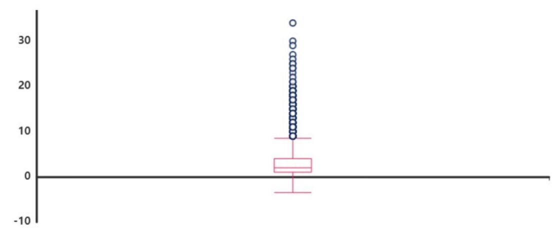
(a) p1_distance_run



(b) p2_distance_run



(c) speed_mph



(d) rally_count

Figure 2: Boxplot Results

Step2: Numerical Transformation of Character Data: In the original dataset, numerical values for parameters such as serve depth, serve width, and return depth were presented in character form. To convert these characters into analyzable numerical data, we conducted a process of numerical transformation. This process involved gathering information to understand the numerical range represented by each character. Once we obtained this information, we employed a method of substituting intermediate values to replace the characters with their corresponding numerical equivalents. The purpose of this step was to ensure that the data could be accurately analyzed and interpreted, laying a reliable foundation for further research.

Step3: Visual Adjustment of Score Markings: To enhance the data's ease of observation and analysis, we made adjustments to the scoring system used in the matches. The original data represented scores using the conventional 15, 30, and 40 system. We replaced these values with more intuitive 1, 2, and 3, making the scoring system clearer and more observable. This adjustment not only simplified the data but also improved its readability, facilitating a deeper understanding of the match progress and outcomes. Through these processing steps, we successfully transformed the original data into a more analyzable and visually comprehensible format, laying the foundation for further research.

4.2 Momentum Quantification

4.2.1 Indicators Determination

Through an analysis of relevant literature, momentum can be primarily categorized into four main aspects: serving conditions, match performance, individual technical performance, and match situation. Therefore, we have identified these four key areas as primary indicators.

Incorporating insights from the literature, we retained 14 of the most representative secondary indicators to construct our model. The specific descriptions and selected indicators are presented below.

- **serving circumstances**

In tennis competitions, the party serving holds a substantial advantage. The speed of the serve critically influences the reaction time of the opponent, while the depth and width of the serve can force alterations in the opponent's court positioning and the angle at which they return the ball. Based on available data, this study utilizes variables such as ”serve speed” (**SMPH**), ”serve depth” (**SD**), and ”serve width” (**SW**) to assess their impact on the outcome of matches. The research specifically quantifies how various serve conditions affect the momentum within the game.

- **competitive performance**

In light of the serving advantage, the ability of the serving side to win this match becomes particularly crucial. We utilize metrics such as ”5-game short-term break points” (**STB**), ”break points in this match” (**LTB**), ”5-game short-term hold points” (**STH**), and ”hold points in this match” (**LTH**) to measure how match performance affects the momentum.

- **individual technical proficiency**

Serving and volleying is an offensive tactic in tennis, which presents a considerable level of technical difficulty and poses a significant threat to the opponent. Moreover, as the match progresses, players will expend more energy, potentially leading to unforced errors. Therefore, we utilize metrics such as the frequency of serving and volleying (**NC**), serving and volleying success rate (**NSR**), running efficiency (**RE**), and whether unforced errors occur in this game (**UEC**) to assess the impact of ”individual technical performance” on the momentum of the match.

- **game dynamics**

The fluctuations in the match situation directly impact the psychological state of the players and their tactical decisions. Moreover, the ability to quickly adapt to changes in the match situation is crucial for achieving victory. We use metrics such as whether it is the serving side (**BTS**), the number of games won in the current set (**SWG**), and the points lead in the current game (**SGA**) to measure the impact of the ”match situation” on the momentum of the match.

4.2.2 Weight Calculation

In order to obtain scientifically valid weights for indicators at different levels, we employed a hybrid approach that integrates both subjective and objective methods. This approach leveraged data normalization, the entropy weighting technique, and the Analytic Hierarchy Process (AHP) as part of a comprehensive methodology. The data processing steps were carried out as follows:

Step1. data normalization

To nullify the effects of different scales and units, data normalization is applied, transforming data with various scales and dimensions into dimensionless numerical values. This process facilitates effective comparison and analysis across diverse indicators.

After a detailed examination of the data's characteristics and distribution, appropriate normalization methods are selected to boost the model's stability and prediction accuracy:

- For data not following a uniform or normal distribution, such as 'SMPH' and 'RE', Z-score (Standard Score) normalization is utilized. The Z-score normalization is defined as:

$$z = \frac{(x - \mu)}{\sigma} \tag{1}$$

  where x is the original value, μ is the mean, and σ is the standard deviation of the data.

- For data that is not normally distributed and has clear boundaries, such as 'SD' and 'SW', Min-Max normalization is applied. The Min-Max normalization formula is defined as:

$$x_{\text{norm}} = \frac{(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})} \tag{2}$$

  where x is an original value, x_{min} and x_{max} are the minimum and maximum values of the data, respectively.

These normalization methods are carefully chosen based on the inherent properties of the data to ensure accuracy and reliability in the modeling process.

Step2. Entropy Weight Method

After normalizing the data, the entropy weighting method is employed to determine the weights, following a mathematical framework and rationale outlined as follows:

- **Calculation of Proportions**: For each criterion (indicator), the proportion p_{ij} is first calculated. It represents the value x_{ij} of the ith criterion for the jth evaluation object divided by the total sum of the values of that criterion across all evaluation objects. The mathematical expression is:

$$p_{ij} = \frac{x_{ij}}{\sum_{j=1}^{n} x_{ij}}, \quad i = 1, 2, ..., m; \quad j = 1, 2, ..., n \tag{3}$$

  This step aims to standardize the data, ensuring subsequent computations are conducted on a consistent scale.

- **Calculation of Entropy**: Next, the entropy e_i for each criterion is calculated using the proportions. Entropy is a measure of uncertainty or the amount of information. In this context, it is used to measure the uniformity of the data distribution of a criterion. A higher entropy value for a criterion implies a more uniform distribution, indicating lower utility in differentiating between evaluation objects. The formula for calculating entropy is:

$$e_i = -\frac{1}{\ln(n)} \sum_{j=1}^{n} p_{ij} \ln(p_{ij}), \quad i = 1, 2, ..., m \tag{4}$$

Here, $\ln$ denotes the natural logarithm.

- **Determination of Redundancy (Degree of Diversification)**: Based on the entropy, the redundancy d_i for each criterion is further determined. Redundancy reflects the effectiveness of a criterion in differentiating between evaluation objects. The lower the redundancy of a criterion, the more important it is in the evaluation process. The formula for redundancy is:

$$d_i = 1 - e_i, \quad i = 1, 2, ..., m \tag{5}$$

- **Determination of Weights**: Finally, the weight w_i for each criterion is determined based on its redundancy. The weight of a criterion is the ratio of its redundancy to the sum of redundancies of all criteria. The mathematical expression is:

$$w_i = \frac{d_i}{\sum_{i=1}^{m} d_i}, \quad i = 1, 2, ..., m \tag{6}$$

This step ensures that the sum of weights for all criteria is 1 and that the weight distribution reflects the relative importance of each criterion in the entire evaluation system.

Through the above process, the entropy weighting method allocates weights to each criterion in an evaluation model in an objective and mathematically grounded manner, providing a solid foundation for comprehensive evaluation.

Step3. Analytic Hierarchy Process

After establishing the allocation of weights for each primary indicator within the secondary indicators through the entropy weighting method, we proceeded to determine the weight proportions of the primary indicators using the Analytic Hierarchy Process (AHP) in accordance with the following steps:

- **Construction of Pairwise Comparison Matrix** A: Assemble matrix A with elements a_{ij}, where $i, j = 1, 2, \ldots, n$. Each element a_{ij} quantifies the relative importance of criterion i in comparison to criterion j. The matrix is structured as follows:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix} \tag{7}$$

The matrix is reciprocal, where $a_{ij} = \frac{1}{a_{ji}}$, ensuring consistency in the comparison.

- **Eigenvalue and Eigenvector Calculation**: Compute the eigenvalues λ and eigenvectors v of matrix A. Identify the maximum eigenvalue λ_{max} and its corresponding eigenvector v_{max}, which will be used to derive the priorities of the criteria.

- **Weight Derivation**: Normalize the principal eigenvector v_{max} to obtain the priority weights W of the criteria. The normalization is performed as:

$$W = \frac{v_{\text{max}}}{\sum_{i=1}^{n} v_{\text{max},i}} \tag{8}$$

  This step ensures that the sum of all weights is 1, providing a proportional representation of each criterion's importance.

- **Consistency Index Calculation (CI)**: Evaluate the consistency of the pairwise comparisons by calculating the Consistency Index (CI) using the formula:

$$CI = \frac{\lambda_{\text{max}} - n}{n - 1} \tag{9}$$

  CI assesses the probability that the matrix judgments were randomly generated, with lower values indicating higher consistency.

- **Determination of Consistency Ratio (CR)**: The Consistency Ratio (CR) is calculated to provide a quantifiable measure of the pairwise comparison matrix's consistency. The CR is derived by comparing the Consistency Index (CI) with the appropriate Random Index (RI) value, which is obtained from a standardized table corresponding to the matrix's dimensionality. The CR computation is articulated as:

$$CR = \frac{CI}{RI} \tag{10}$$

  In this context, a CR value below the threshold of 0.1 is indicative of a satisfactory consistency level within the pairwise comparison matrix. For this specific evaluation, the CR value was calculated to be:

$$CR = 0.01147537 \tag{11}$$

  This value substantiates the matrix's high level of consistency and reliability in reflecting the true priorities among the compared elements.

4.2.3 Quantitative Momentum Outcomes

In our quantitative analysis of athlete performance, we have employed a combined approach of the Entropy Method and Analytic Hierarchy Process (AHP), meticulously integrating both subjective and objective elements to ensure a comprehensive and scientific assessment. This multidimensional and multilevel quantitative analysis effectively encapsulates and quantifies the athlete's comprehensive performance in competitions. The two-level weight table presented below elucidates the distribution of weights among primary and secondary indicators, highlighting the relative importance of each indicator in the assessment of athlete performance.

Herein, we present the two-level weight Table 2 derived from the combined application of the Entropy Method and the Analytic Hierarchy Process (AHP). This table meticulously displays the distribution of weights among various assessment indicators in our model, offering a clear perspective for an in-depth understanding of the model's construction and evaluative logic.

Table 2: Two-Level Weight Table

Primary Indicator	Primary Weight (%)	Secondary Indicator	Secondary Weight (%)
Serving Technique	46.73	SMPH	5.97
		SD	30.88
		SW	63.15
Match Performance	27.72	STB	56.28
		LTB	15.07
		STH	23.12
		LTH	5.53
Personal Technique Performance	16.01	NC	31.97
		NSR	12.51
		RE	54.85
		UEC	0.67
Match Situation	9.54	BTS	65.49
		SWG	29.55
		SGA	4.96

This two-level weight table intricately reveals the significance of each assessment indicator in the analysis of quantitative momentum outcomes. At the primary indicator level, 'Serving Technique' commands the highest proportion, underscoring its pivotal role in competitions; at the secondary indicator level, 'SW' (Serving Win Rate) and 'STB' (Short-Term Performance) hold the highest weights in their respective categories, further accentuating the centrality of short-term performance and serving phases in the assessment of an athlete's performance. This meticulous distribution of weights not only enhances the precision of the model but also provides a multidimensional perspective for a profound understanding of athlete performance.

4.2.4 Visual Representation Of Momentum

Ultimately, leveraging our meticulously developed momentum calculation formula, we embarked on a comprehensive and detailed analysis, meticulously comparing the momentum exhibited by the two contenders in the prestigious 2023 Wimbledon Men's Singles final. This profound analysis entailed an in-depth examination of the players' performance dynamics, their strategic adjustments, and their psychological endurance throughout the intensely fought match.

The insights were visually encapsulated in a meticulously crafted to Figure 3, which served as a testament to the robustness and predictive accuracy of our model. The figure vividly illustrates the intricate interplay of the players' momentum, juxtaposed against the backdrop of the match's pivotal moments, game-changing breaks, and critical service holds.

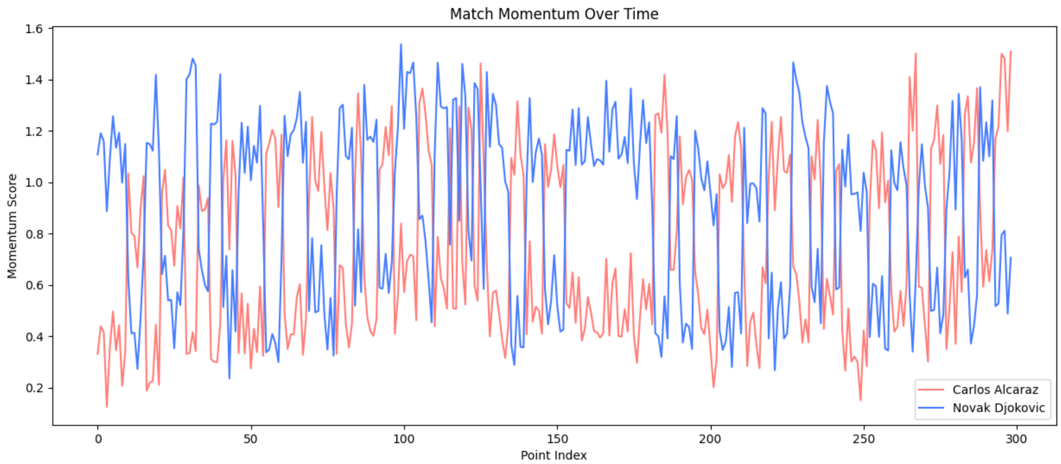


Figure 3: Match Momentum Over Time

It's noteworthy that the momentum trends of both players didn't just mirror each other in isolation but were also in strong alignment with the ebb and flow of the match's overall course.

This alignment underscores the nuanced interdependence between a player's moment-to-moment psychological state and their physical performance on the court. It also highlights the utility of our model in capturing these complex dynamics, offering an innovative lens to understand and predict the outcomes of high-stakes tennis matches with greater precision. The triumph of Carlos Alcaraz, as foretold by the model, wasn't just a testament to his skill and tenacity but also a validation of the sophisticated analytical framework that underscored our study.

5 Task 2: Consistency and Randomness Verification of the Model

5.1 Consistency Testing

5.1.1 Pearson Correlation Analysis

In our analytical exploration of the relationship between athletes' scores and momentum, we employed the Pearson correlation coefficient test. The Pearson correlation coefficient test is a parametric statistical method aimed at quantitatively assessing the strength of a linear relationship between two continuous variables. This technique systematically quantifies the degree of association between the variables, providing a quantifiable statistical underpinning for our research. The principal procedures involved in the Pearson correlation coefficient test are as follows:

- **Calculation of Mean Values**: Calculate the mean values $\bar{x}$ and $\bar{y}$ for datasets X and Y, respectively. The means are computed as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{12}$$

- **Computation of Covariance and Standard Deviations**: Determine the covariance between X and Y, along with the standard deviations of each dataset. Covariance is given by:

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{13}$$

  The standard deviations are calculated as:

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \sigma_Y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{14}$$

- **Pearson Correlation Coefficient Calculation**: Compute the Pearson correlation coefficient r using the formula:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \tag{15}$$

  This coefficient r measures the linear relationship between X and Y, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation).

- **Statistical Significance Testing**: Evaluate the statistical significance of the correlation coefficient r by calculating the t-statistic:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{16}$$

  Compare this t-value against critical values from the t-distribution table for a chosen confidence level to determine the significance of the correlation.

Table 3: Results of the Run Test for Momentum Values

	momentum	**SGA**
momentum	1 (0.000***)	0.379 (0.000***)
SGA	0.379 (0.000***)	1 (0.000***)

Note: ***, **, and * respectively denote significance at the 1%, 5%, and 10% levels.

As indicated in the Table 3, the correlation coefficient between 'momentum' and 'SGA' stands at 0.379, with an accompanying p-value of less than 0.001. This denotes a statistically significant correlation at the 1% level of significance. The inference drawn from this is that the likelihood of such a correlation occurring by chance is exceedingly low, thereby allowing us to conclude with a high degree of confidence that there exists a substantial linear relationship between the two variables.

5.1.2 Visualization of Results

The quantified momentum data were visualized in conjunction with the score differentials between competitors and their opponents, specifically focusing on the final match data. This visualization culminated in the creation of Figure 4, which encapsulates the data-driven narrative of the contest.

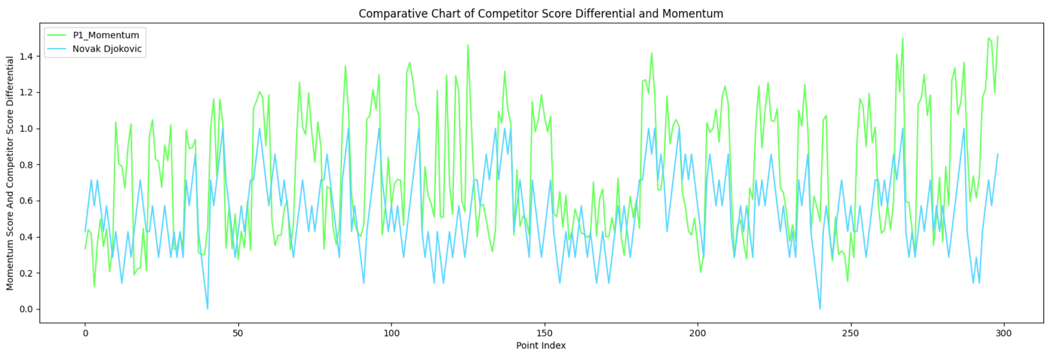


Figure 4: Comparative Chart of Competitor Score Differential and Momentum

The visualized results reveal a trend characterized by a marked degree of similarity, suggesting that fluctuations in momentum are not random occurrences. Instead, there exists a significantly high correlation between the changes in momentum and the scoring dynamics of the players. This indicates that momentum shifts may be systematically associated with the performance outcomes within the context of the

5.2 Randomness Testing

In our investigation of the randomness characteristics within momentum data, we employed the Run Test, also known as the Wald-Wolfowitz Test. The Run Test is a non-parametric statistical method designed to assess whether a sequence of data points exhibits randomness. This method objectively determines the randomness within the data through a series of well-defined steps, providing a quantifiable foundation for the analysis. The specific steps are as follows:

- **Binary Sequence Construction**: Form a binary sequence S from the series of momentum values by comparing each observation x_i to a benchmark criterion, such as the median. The sequence is defined as:
$$S = [s_1, s_2, \ldots, s_n] \tag{17}$$
where $s_i = 1$ if x_i is above or equal to the benchmark and $s_i = 0$ if x_i is below the benchmark.

- **Run Counting**: Identify and count the total number of runs R within the binary sequence, with a run being a consecutive sequence of identical elements.

- **Expected Runs Calculation**: Determine the expected number of runs $E[R]$ as follows:
$$E[R] = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{18}$$

where n_1 and n_2 are the counts of 1s and 0s in S, respectively.

- **Variance of Runs Calculation**: Compute the variance of the number of runs $Var[R]$ using:

$$Var[R] = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \tag{19}$$

- **Test Statistic Computation (Z)**: Calculate the test statistic Z which is the normalized measure of the deviation from the expected number of runs:

$$Z = \frac{R - E[R]}{\sqrt{Var[R]}} \tag{20}$$

- **Decision Rule**: Compare the calculated Z value with the critical values of the standard normal distribution. If Z falls outside the acceptance region (e.g., for $\alpha = 0.05$, the two-tailed critical values are approximately ± 1.96), reject the null hypothesis that the sequence is random. If Z falls within the acceptance region, there is not enough evidence to conclude that the sequence is non-random.

- **Result Interpretation**: Conclusions based on Z value and critical values are as follows:
  - If Z is within the acceptance region, the null hypothesis of randomness cannot be rejected.
  - If Z is outside the acceptance region, reject the null hypothesis, suggesting that the sequence may follow a non-random pattern or trend.

- **Two-Tailed Test P-value**: Calculate the p-value which indicates the probability of observing the calculated Z value, or a more extreme, under the assumption that the null hypothesis (the sequence is random) is true. The p-value is calculated using the standard normal distribution as:

$$p = 2(1 - \Phi(|Z|)) \tag{21}$$

where Φ is the cumulative distribution function of the standard normal distribution.

As delineated in Table 4, the results procured from the Wald-Wolfowitz Test elucidate the empirical findings of our statistical inquiry.

Table 4: Results of the Run Test for Momentum Values

Variable	**Observations**	**Z**	**P**
p1_momentum	5463	-38.576	0.000***

Note: ***, **, and * respectively denote significance at the 1%, 5%, and 10% levels.

With a calculated p-value of $p < 0.01$, we have strong evidence against the null hypothesis at a 1% significance level. This indicates that the sequence of momentum values is not random, thus demonstrating the presence of a significant pattern or trend in the data.

6 Task 3:Predictive Momentum Model

6.1 Strategies for Optimal Model Selection

In the process of selecting a predictive model, an evaluation methodology was established based on the cumulative sum of the absolute values of predictive errors. Specifically, the process involved splitting the data into training and validation sets, training the model on the training set, and then assessing its accuracy using the validation set,for each candidate model, we calculated the absolute value of the differences between the predicted and actual values and summed these to yield a total error amount. This assessment metric provided us with a quantified means to gauge the accuracy of the model's predictions.Utilizing this method, we assessed four different predictive models and generated a comparative graph Figure 5 that visually displayed the total cumulative error for each model.

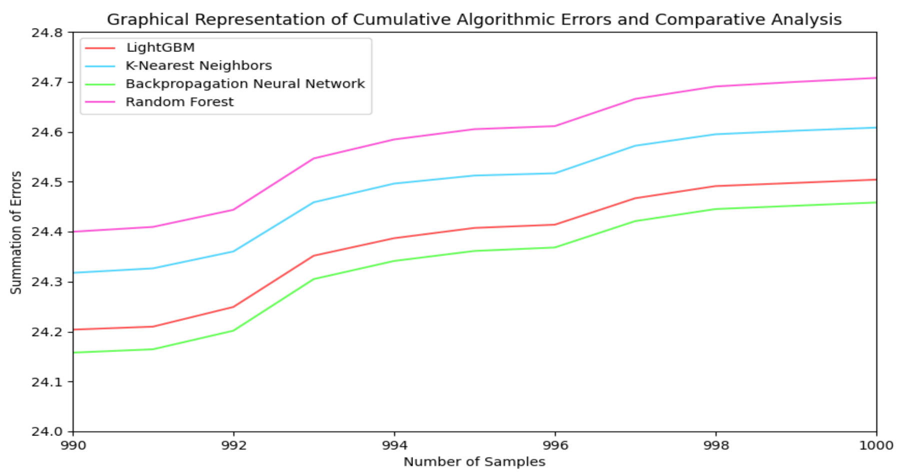


Figure 5: Comparative Chart of Competitor Score Differential and Momentum

After thorough analysis and comparison of the total errors, we selected the best-performing Back Propagation (BP) neural network model. The BP neural network was chosen for its ability to capture complex nonlinear relationships and its flexibility in adjusting to various types of data via its hidden layers and backpropagation algorithm. The architecture of the model, including the number of input layers, the number of hidden layers and their neurons, and the output layer size corresponding to the predicted outcome, was determined based on various criteria influencing the match's momentum.

Owing to the supervised learning characteristic of the Back Propagation (BP) neural network algorithm, it is adept at recognizing complex nonlinear relationships between inputs (independent variables) and outputs (dependent variables). However, unlike regression models, it does not explicitly calculate or quantify the relationship between each independent variable and the dependent variable in a transparent manner. To further quantify and understand the significance and contribution of each independent variable towards the dependent variable, we employed the Random Forest model. This model provides a comprehensive view of feature importance by measuring the contribution of each feature to the model's accuracy, thereby helping to identify which features are most influential in predicting the outcome. This method not only enhances our understanding of the underlying dynamics within the data but also augments the BP neural network by offering an additional layer of interpretability regarding the relationships between variables. By integrating insights gained from the Random Forest's feature importance with the predictive prowess of the BP neural network, a more robust and informed decision-making process is achieved in predicting shifts in match momentum.

6.2 Back Propagation neural network model:Volatility Prediction

6.2.1 The establishment of the BP model

In light of the modest volume of data available, a neural network architecture of lower complexity was elected. The activation function implemented was 'identity', providing a direct transmission of the input signal; the 'L-BFGS' algorithm, known for its efficacy, was selected as the optimization method. A fixed learning rate of 0.1 was established to modulate the incrementation of weight adjustments. To mitigate the risk of overfitting, an L2 regularization coefficient was set at a value of 1. The model underwent a sequence of 1000 iterations during training, with the first hidden layer comprising 100 neurons. The configuration of these parameters plays a pivotal role in determining the training efficiency and the resultant efficacy of the model.

6.2.2 The evaluation of the BP model

. The evaluation results on the test set are illustrated in Table 5.

Table 5: BP Model Evaluation Metrics

	MSE	RMSE	MAE	MAPE	R^2
Training Set	0	0.003	0.003	0.501	1
Test Set	0	0.003	0.003	0.466	1

Model Descriptions:

- MSE (Mean Squared Error): The average of the squares of the errors—the average squared difference between the estimated values and the actual value.
- RMSE (Root Mean Squared Error): The square root of MSE, providing a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better.
- MAE (Mean Absolute Error): A measure of errors between paired observations expressing the same phenomenon.
- MAPE (Mean Absolute Percentage Error): A measure that expresses the relative error between the actual value and the prediction as a percentage.

- R^2 (Coefficient of Determination): Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

Based on the provided evaluation metrics, the model demonstrates exemplary predictive performance on both the training and testing datasets, with negligible error. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) values have reached zero, while the Root Mean Squared Error (RMSE) remains at a minimal level. Moreover, the model's Mean Absolute Percentage Error (MAPE) on the test set is slightly lower than on the training set, indicating a commendable generalization capability to unseen data. The coefficient of determination (R^2) value of 1 signifies that the model has achieved a perfect fit to the data. These results indicate that the model excels in both predictive accuracy and interpretability, affirming its reliability as a predictive tool within the domain of analysis.

6.2.3 The prognostication rendered by the BP model

We utilized the acquired model to forecast the momentum of both players in the 2023 Wimbledon Men's Singles final, resulting in Figure 6, which closely aligns with the actual values.

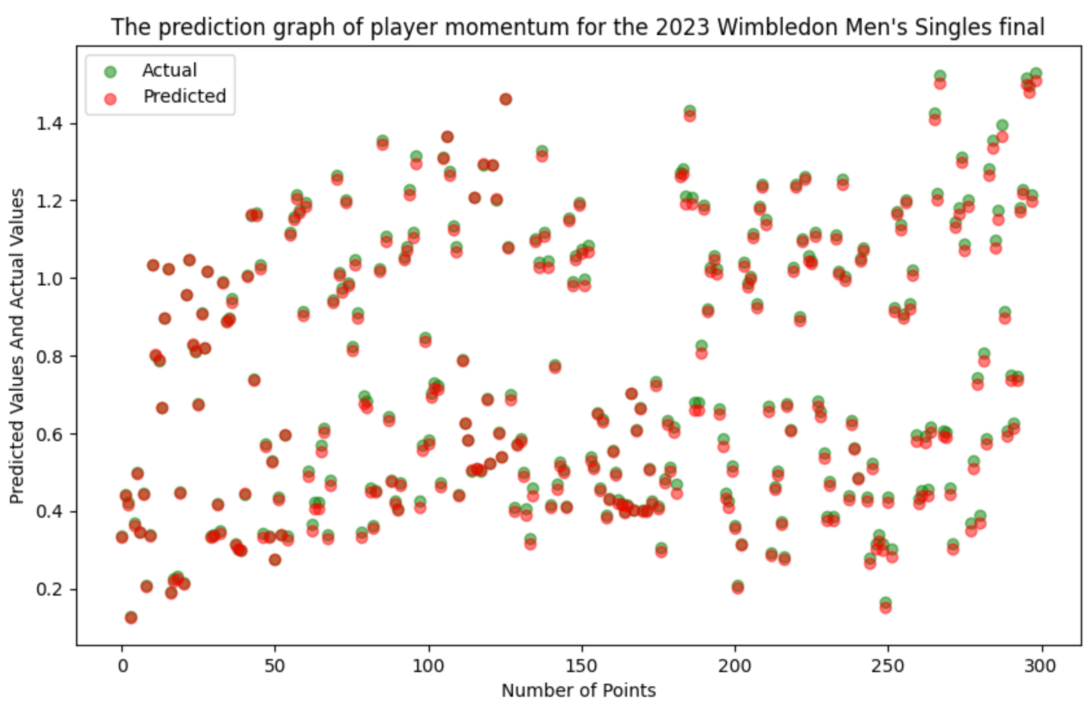


Figure 6: The prediction graph of player momentum for the 2023 Wimbledon Men's Singles final.

6.3 Random Forest Regression: Evaluating Feature Relevance

6.3.1 Establishment of a Random Forest Model

In our model configuration, we specified several key parameters for the random forest algorithm. Specifically, we chose not to perform cross-validation or shuffle the data during training. The criterion used for node splitting was mean squared error (MSE), and we constructed an ensemble of 100 decision trees. Each individual tree was constrained to have a maximum depth of 10 and was allowed to have a maximum of 50 leaf nodes. Additionally, our model employed bootstrapped sampling (with replacement) and did not conduct out-of-bag testing. These parameter settings hold significant influence over the model's performance and behavior, providing the fundamental framework for both training and prediction tasks.

Ultimately, we obtained Figure 7 through random forest analysis.

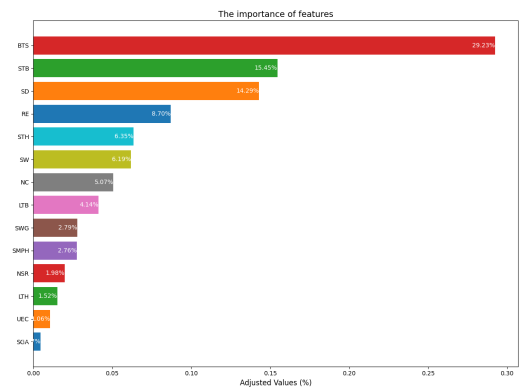


Figure 7: The prediction graph of player momentum for the 2023 Wimbledon Men's Singles final.

"We observed that a player's momentum is most strongly correlated with BTS (whether the player is the server). This suggests that the advantage conferred by serving during the game has a significant impact on a player's momentum."

6.3.2 Strategic Recommendations for Athletes

Based on the analysis conducted using our model, we offer the following recommendations for players when facing various opponents:

- **Importance of Serving:** Given the model's indication of serving being paramount in influencing momentum, players should focus on enhancing the quality of their serve, including speed, accuracy, and variability. A strong and unpredictable serve can help establish an early advantage, setting a positive tone for the remainder of the match.

- **Holding Serve Strategy:** Maintaining one's service game is crucial for sustaining and amplifying momentum. Players should remain composed during their service games, leveraging their strengths, whether it be powerful baseline strokes or deft net play, to suppress the opponent.

- **Breaking Serve Strategy:** Seeking to break the opponent's serve is equally vital. Players should intensify their offense during the opponent's service games, exploiting weaknesses such as unstable second serves or poor baseline defense. Psychological tactics, like displaying confidence and aggression, can also help disrupt the opponent's momentum.

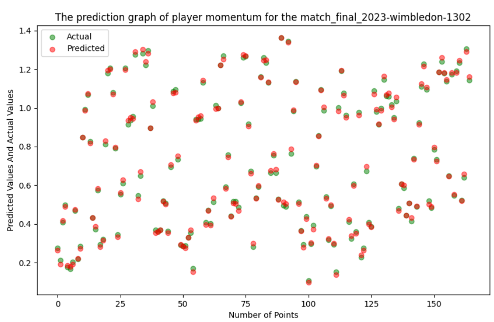
- **Adaptation to Various Opponents:** When facing different types of opponents, players need to adjust their strategies accordingly. For instance, against baseline players, varying ball trajectories and pace can disrupt their rhythm; against net players, employing low and deep shots is advisable.

7 Task 4:Further In-depth Analysis of the Selected Model

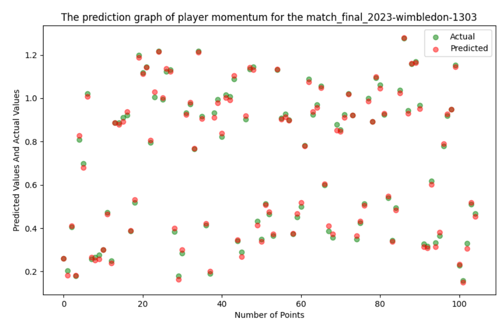
To further analyze the predictive performance and the generalization capability of the Backpropagation (BP) model, we will adhere to the following steps:

7.1 Application of the Model in Other Tennis Competitions

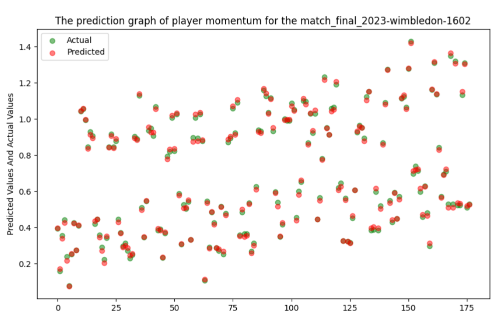
We applied the developed model to other tennis competitions, and a subset of the results is graphically presented in Figure 8.



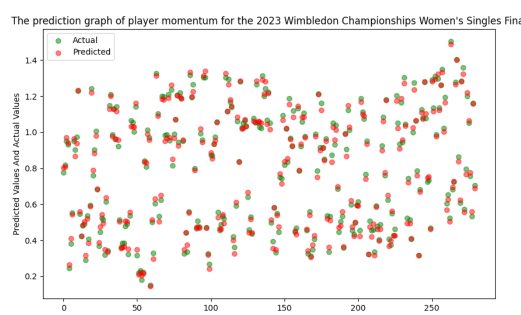
(a) 2023-wimbledon-1302



(b) 2023-wimbledon-1303



(c) 2023-wimbledon-1602



(d) 2023-wimbledon-women's singles final

Figure 8: Comparison of Predicted and Actual Player Values

Upon observing the results, it is apparent that the model exhibits remarkably high accuracy in men's tennis matches. However, in the women's category, there is a slight shortfall in accuracy.

7.2 Enhancing the Model's Performance in Women's Matches

In order to simulate the differences in skill levels among players and the disparities between men and women in serving advantages, we introduced two new factors: the ranking difference and gender coefficient. The ranking difference is defined as the difference in rankings between two players.

$$Rank_{i,j} = Rank_i - Rank_j \tag{22}$$

Here, i and j represent the player indices in the dataset.

The refined model was then validated using a fresh set of data to ensure its robustness and accuracy. This step also included cross-validation techniques to prevent overfitting and to assure the model's performance across different data subsets.

Table 6: The performance of the model before and after improvement.

Before Improvement

	precision	recall	f1-score	support
0	0.88	0.96	0.92	461
1	0.78	0.53	0.63	218
accuracy			0.87	679
weighted avg	0.86	0.87	0.86	679

After Improvement

	precision	recall	f1-score	support
0	0.90	0.96	0.93	461
1	0.83	0.57	0.66	218
accuracy			0.89	679
weighted avg	0.88	0.89	0.89	679

Performance metrics such as accuracy, precision, recall, and F1 score were recalculated after model optimization to measure and confirm the improvements in the predictive performance of the neural network.

7.3 Sensitivity Analysis

Sensitivity analysis aims to assess the impact of variations in independent variables on specific dependent variable outcomes, maintaining a predefined set of assumptions. To determine the influence of each input variable on the network's predictions, a sensitivity analysis was performed. This helped in identifying the most significant features and allowed for the refinement of the input data set to include only the most influential variables. In our model, certain elements such as the techniques for serving and receiving in sports are designated as constants. Likewise, the judgment matrix within the Analytic Hierarchy Process (AHP) model is also treated as a constant entity. However, recognizing the rarity of strictly constant parameters in real-world scenarios, we introduced a random perturbation term, denoted as ε_i , to our dataset. This addition aims to simulate the variations in outcomes that could arise from such a phenomenon. We assume that this random perturbation fluctuates within a range of 5$\%$ of the original data and parameter values.

$$x'_{i,j} = x_{i,j} \cdot \varepsilon_i \qquad \varepsilon_i \sim U(0.95, 1.05) \tag{23}$$

The data and parameters, after the integration of this random disturbance, are then redefined accordingly. Following this modification, the data underwent processes such as normalization. It was observed that only a percentage (3.67%) of the results experienced any change. This minimal variation implies that the outcomes were not significantly influenced by the constants, thereby suggesting that our model exhibits a relative degree of stability.

By comparing the model performance before and after improvement, we observe an overall enhancement in evaluation metrics. However, due to variations in rules and other factors across different competitions, the model's performance may decrease. Nevertheless, this can be addressed by adding relevant variables and simultaneously implementing measures to prevent overfitting, thereby improving the model's overall performance.

8 Model Evaluation and Further Discussion

8.1 Strengths

- **Robust Predictive Performance**: The Back Propagation (BP) neural network model demonstrated exemplary predictive accuracy, as evidenced by near-zero error metrics (MSE, RMSE, MAE, MAPE) and a perfect R^2 value on both training and testing datasets. This indicates a high level of precision in the model's ability to forecast outcomes based on the input data.

- **Feature Importance Analysis**: Integration of the Random Forest model provided an additional layer of interpretability by quantifying the significance and contribution of each independent variable towards the dependent variable. This not only enhanced the understanding of the underlying dynamics within the data but also complemented the BP neural network model by offering a nuanced perspective on feature relevance.

- **Effective Handling of Nonlinearity**: The BP neural network's architecture and the backpropagation algorithm proved effective in capturing complex nonlinear relationships within the data, a crucial aspect given the intricate nature of the variables influencing match momentum in sports.

- **Visual Representations**: The use of meticulously crafted figures (e.g., Figure 1, Figure 4) to visualize the momentum trends and the predictions offered by the models provided an intuitive and compelling narrative of the data, aiding in the interpretation and presentation of the results.

8.2 Weaknesses

- **Model Complexity**: While the BP neural network captures complex relationships, its inherent complexity may lead to challenges in interpretation. Unlike more straightforward models, neural networks act as a 'black box,' making it difficult to understand the exact nature of the relationship between input and output variables.

- **Dependency on Data Quality and Quantity**: The performance of the BP model, as with any machine learning model, heavily relies on the quality and volume of the data fed into it.

In scenarios where data is scarce, noisy, or non-representative, the model's performance and reliability might be compromised.

- **Overfitting Risk**: Despite precautions such as L2 regularization, neural networks are prone to overfitting, especially when dealing with complex models and large numbers of features. Ensuring that the model generalizes well to unseen data remains a critical challenge.

8.3 Further Discussion

- **Model Generalizability**: Delve into how the model might perform when applied to different datasets or in different domains. Discuss the potential need for retraining or fine-tuning when adapting the model to other sports or datasets with different characteristics.
- **Future Enhancements**: Propose potential improvements to the model, such as experimenting with different neural network architectures (e.g., Convolutional Neural Networks for spatial data or LSTM for time-series data) or incorporating additional data sources to enrich the model's input (e.g., player psychological states or external environmental factors).
- **Ethical Considerations**: Address any ethical implications of predictive modeling in sports, especially regarding data privacy, the potential for misuse of predictive insights, and ensuring the fairness and transparency of the model.

9 Memorandum

Dear Coach,

We are a professional data analysis team from MCM. The concept of ”momentum” in tennis has long been a subject of debate, and that's why we embarked on an in-depth exploration and analysis of the impact and mechanisms of momentum in tennis matches.

In our research, we focused on analyzing detailed data from the 2023 Wimbledon Men's Singles final. Through this data, we formulated a calculation formula for momentum. Firstly, we analyzed the feature values involved in the momentum calculation formula. We employed a combination of subjective and objective methods, using entropy weighting and the Analytic Hierarchy Process (AHP) to determine the parameters of each feature value. Subsequently, we conducted both Pearson and run tests to assess the consistency and randomness of the momentum calculation formula. The results demonstrated that indeed the concept of momentum plays a significant role in predicting match outcomes.

Furthermore, we applied our model to predict momentum changes in other matches. While the performance may not reach the same level as in the training dataset, it does provide a rough prediction of momentum trends in the matches.

Based on our research findings, we would like to propose the following suggestions:

1. Strengthen the players' training in serving and receiving skills, as this directly impacts the hold and break rates, which are key in controlling momentum.

2. Teach players how to make strategic adjustments and adapt on the spot during matches, including timeouts and changes in rhythm, to influence the flow of momentum.

3. Enhance psychological and emotional management training to help players better understand and control momentum shifts during matches.

Additionally, we recommend that you and the players regularly analyze past match footage and data. This practical experience will deepen your understanding and application of momentum control. Understanding how to leverage momentum changes at critical moments can provide a significant competitive advantage for the players in future matches.

Furthermore, we advise paying attention to those inconspicuous but crucial details in matches, such as opponents' weaknesses and changes in match conditions. These factors may subtly influence the direction of momentum.

We believe that with these strategies and preparations, your players will be able to utilize momentum more effectively in future matches, ultimately enhancing their competitiveness.

Wishing you and your players even greater success in future matches!

Sincerely,

MCM Team #2423845

References

[1] Ku lakowski, Konrad, et al. ”Towards secure judgments aggregation in AHP.” arXiv preprint arXiv:2303.15099 (2023).

[2] Saifullah, Shoffan. ”Tourism's trend Ranking on Social Media Data Using Fuzzy-AHP vs. AHP.” arXiv preprint arXiv:2209.14503 (2022).

[3] Mendrofa, G. Aileen, Gatot Fatwanto Hertono, and Bevina Desjwiandara Handari. ”Ensemble Learning Model on Artificial Neural Network-Backpropagation (ANN-BP) Architecture for Coal Pillar Stability Classification.” arXiv preprint arXiv:2303.16524 (2023).

[4] Dorador, Albert. ”Improving the Accuracy and Interpretability of Random Forests via Forest Pruning.” arXiv preprint arXiv:2401.05535 (2024).

[5] MacBride, Cara, Vinny Davies, and Duncan Lee. ”Conditional autoregressive models fused with random forests to improve small-area spatial prediction.” arXiv preprint arXiv:2312.12106 (2023).

[6] Jin, Wen, et al. ”The improvements of BP neural network learning algorithm.” WCC 2000-ICSP 2000. 2000 5th international conference on signal processing proceedings. 16th world computer congress 2000. Vol. 3. IEEE, 2000.

[7] Li, Jing, et al. ”Brief introduction of back propagation (BP) neural network algorithm and its improvement.” Advances in Computer Science and Information Engineering: Volume 2. Springer Berlin Heidelberg, 2012.