WEP
Warwick
Evans
Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Investigating and Forecasting Momentum Shift Effects on Strategy Development in Sports Competitions

Wenqian Sun *, Heng Hua

School of Land Science and Technology, China University of Geosciences, Beijing, China, 100083

* Corresponding author: sunwenqian2003@gmail.com

Abstract. Momentum shifts in competitive sports have a significant impact on game outcomes and strategy development, and despite their widely recognized importance, there is a relative lack of relevant systematic research. In this regard, this study quantified the relationship between momentum shifts and player success, using the metrics analysis method to analyze their correlation and stochasticity. The relevant metrics mainly comprised Spearman rank correlation analysis, and P-value test calculations, which yielded a strong correlation of 0.84 and a P-value less than 0.05. Furthermore, to predict and visualize possible shifts, our study used the LightGBM algorithm to build a predictive model, where we considered shifts to represent changes in strengths and weaknesses between players i.e. moments of change. Afterward, the researchers did an error test for our model, which had a low mean squared error of 0.0019. Finally, this research performed a features importance analysis to identify the most relevant factors such as GPCI, and SGD, and based on the results four ways were suggested for players to deal with new opponents.

Keywords: Momentum-Shifts prediction; Correlation metrics; Light Gradient Boosting Machine algorithm; Error tests; Features importance analysis.

1. Introduction

Momentum is widely recognized as one of the key factors influencing the outcome of a match, especially in tennis matches, which is highly dependent on skill and strategy. Traditional methods of analyzing tennis matches often rely on basic statistics such as points won, games won and sets won. While these data are informative, they fail to capture the subtle changes in momentum during a match, which may profoundly affect a player's mental state and hence his performance [1].

To solve this problem, this paper utilized the Light Gradient Boosting Machine algorithm. The LightGBM algorithm, developed by researchers from Microsoft Research Asia in 2017, constitutes an efficient gradient-boosting framework tailored for tackling machine learning challenges associated with large-scale datasets and high-dimensional feature spaces [2]. The algorithm demonstrates robust applicability across various domains including medicine [3], biology [4], economics [5], and sports [6, 7], thus showcasing its versatility and broad utility in scientific research and practical applications.

In this work, data was driven from www.mcm.com. This paper initially established momentum swing indicators, comprising four metrics. Using the LightGBM algorithm, a swing prediction model was developed and swings were visualized graphically. Based on the results of feature importance analysis in LightGBM and relevant analyses, strategies, and recommendations were provided for athletes to prepare for competitions.

2. The role of momentum in Tennis Matches

2.1. The Establishment of Metrics

This research considered shifts to be swings in strengths and weaknesses between players, and since our study has previously used momentum to quantify player performance during a match, shifts were defined as the point at which momentum is equal between players, i.e., the "moment of transition"[8].

A player's success is closely related to the probability of winning each point for each player at any given time, so this research used P_{win} as one of the criteria for judging a player's success.

Shifts are quantified via Equation (1):

$$Sg_{player} = M_{player1} - M_{player2} \tag{1}$$

Where Sg_{player} represents a change in strengths and weaknesses between players, $M_{player1}$ represents the momentum of player1 and $M_{player1}$ represents the momentum of player2.

Game-Level Psychology Change Index (GPCI): $GPCI_{player} = \sum(E_{i,player} - E_{i-1,player}) * \prod(game_i = 1)$, measuring psychology state changes per game.

Set-Level Psychology Change Index (SPCI): $SPCI_{player} = \sum(E_{i,player} - E_{i-1,player}) * \prod(set_i = 1)$, like GPCI but at a set level.

Success (*Ss*) is then Equation (2):

$$Ss_{player} = 0.10 * GPCI_{player} + 0.15 * SPCI_{player} + 0.75 * P_{winplayer} \tag{2}$$

2.2. Correlation analysis

Firstly, the results showed the visualization and characteristics of the above two indicators Sg and Ss below in Figure 1. They followed a normal distribution approximately.

By comparing the probability of success at moments of high swing with the probability of success at moments of low swing, this study could assess the impact of changes in swing on the outcome of the match. The Spearman's rank correlation coefficient (r_s) is calculated using the formula in Equation (3):

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{3}$$

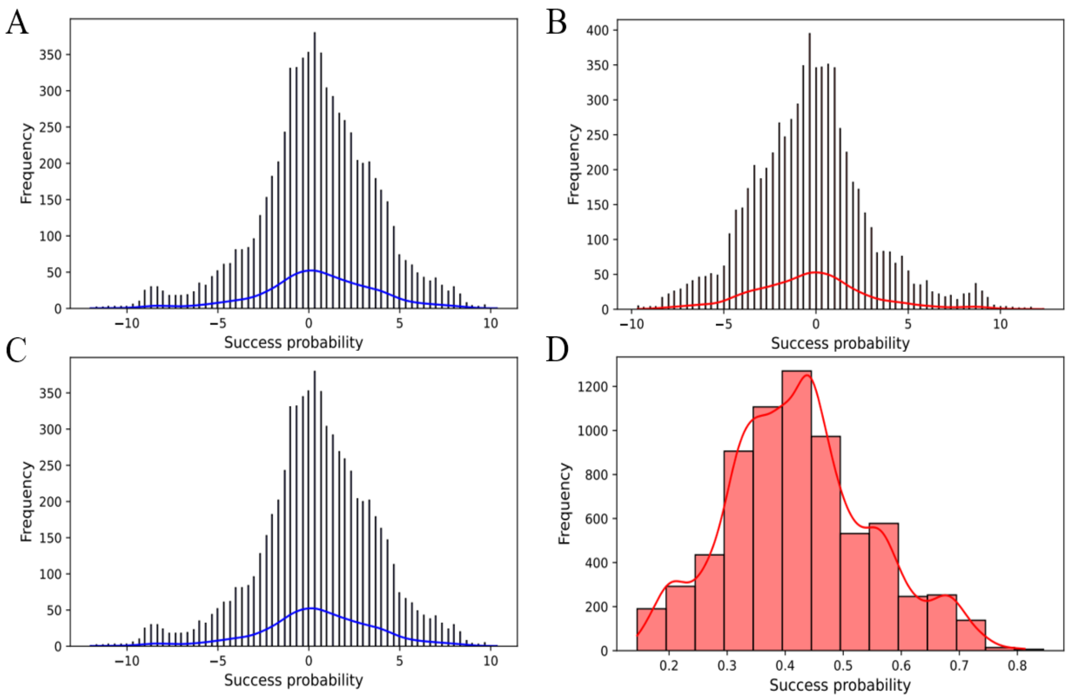

Figure 1. The distribution plots for shift and success

Where d_i is the difference between the ranks of corresponding values x_i and y_i, $and\ n$ is the number of observations.

The analysis of Figure 2 revealed a significant Spearman correlation of 0.84 between swing and the probability of success, with a p-value less than 0.05. This strong correlation suggested that swing changes have a substantial impact on the outcome of the match.

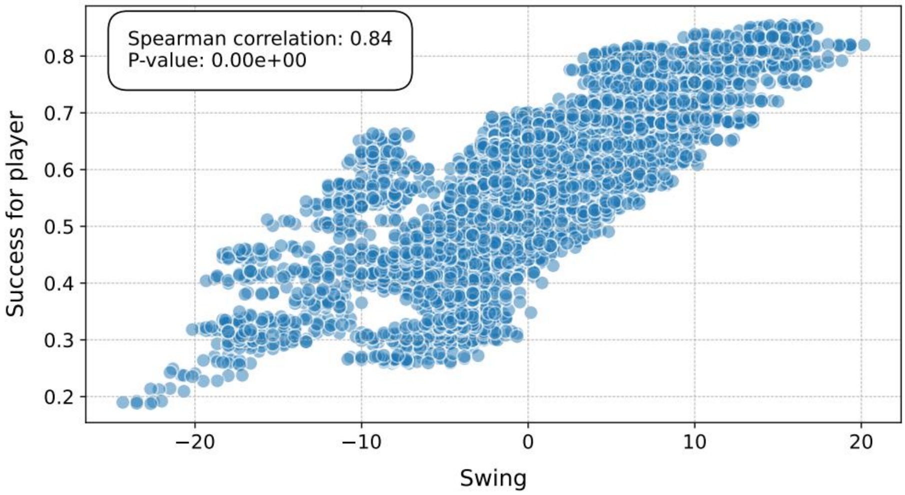

Figure 2. Correlation between Swing and Success.

2.3. Randomization Analysis

In our investigation into the randomness of momentum shifts, a median test was used to assess the distributions of momentum values segmented by the median. This test was designed to determine if two samples come from populations with the same median.

To calculate the P-value in the context of our median test, our study used the following formula, where the P-value represented the probability of observing test results at least as extreme as those observed, under the null hypothesis in Equation (4):

$$P - value = P(T \geq t | H_0\ is\ true) \tag{4}$$

Here, T represents the test statistic calculated from our data, and t is the observed value of the test statistic. The null hypothesis (H_0) posits that there is no difference in the median values between the two groups being compared.

The statistical outcome of the test yielded a statistical value of 6935.9993, with a P-value less than 0.05, indicating a significant deviation from randomness. The observed median momentum value was approximately 0.667.

2.4. Evaluation Results

First of all, the visualization and analysis suggested that while stochastic elements were inherent in match play, consistent momentum gains were indicative of a player's superior performance at certain stages of the match. This observation contradicted the notion that success runs were entirely random, providing a data-driven refutation of the coach's initial skepticism.

The correlation coefficient indicated a direct relationship whereas swing increases, the probability of success also tended to increase. This finding supported the hypothesis that swing played a crucial role in determining the match's outcome, underlining the importance of momentum in sports performance.

The pronounced difference in the distribution of momentum values, as evidenced by the contingency table and the negligible P-value, strongly suggested a non-random pattern within the dataset. This pattern underscored the influence of momentum shifts on the outcomes of interest, highlighting the potential for predictive analysis based on momentum trends.

3. Unveiling the Dynamics of Momentum Swings

3.1. The Establishment of Model

In this section, the Light model's predictions were analyzed to identify patterns that could be indicative of momentum swings within matches. LightGBM (Light Gradient Boosting Machine) is a framework for implementing the GBDT algorithm, which supports efficient parallel training and has the advantages of faster training speed, lower memory consumption, better accuracy, distributed support for rapid processing of massive data, etc. At the same time, it implements the optimization of the Histogram-based decision tree algorithm.

The core algorithmic flow of the model is shown below in Figure 3.

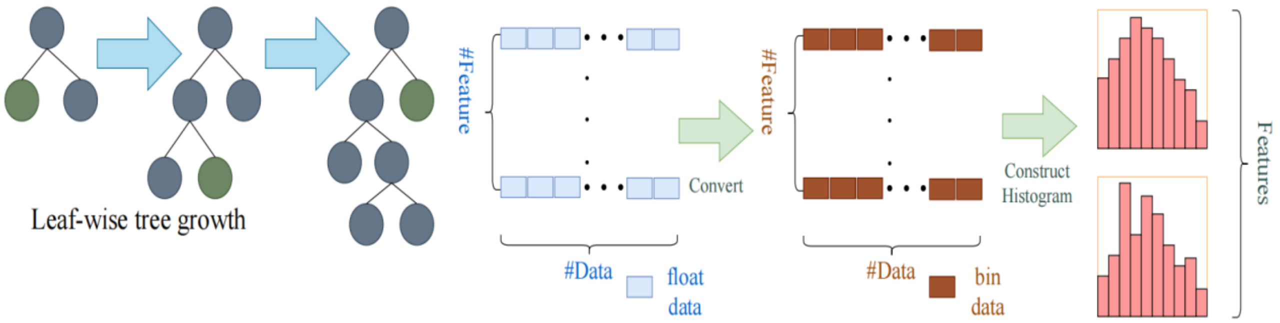

Figure 3. Data segmentation to form leaf nodes and histogram construction.

3.2. The Results of Tennis Match Swing Prediction

The research trained the LightGBM model using the Wimbledon 2023 Gentlemen's matches data and then our study plotted the changes favoring one player to the other based on the training results. The match this study predicted was 2023 Wimbledon-1302, where Alexander Zverev was player 1 and Matteo Berrettini was player 2. The prediction for the match is shown below in Figure 4:

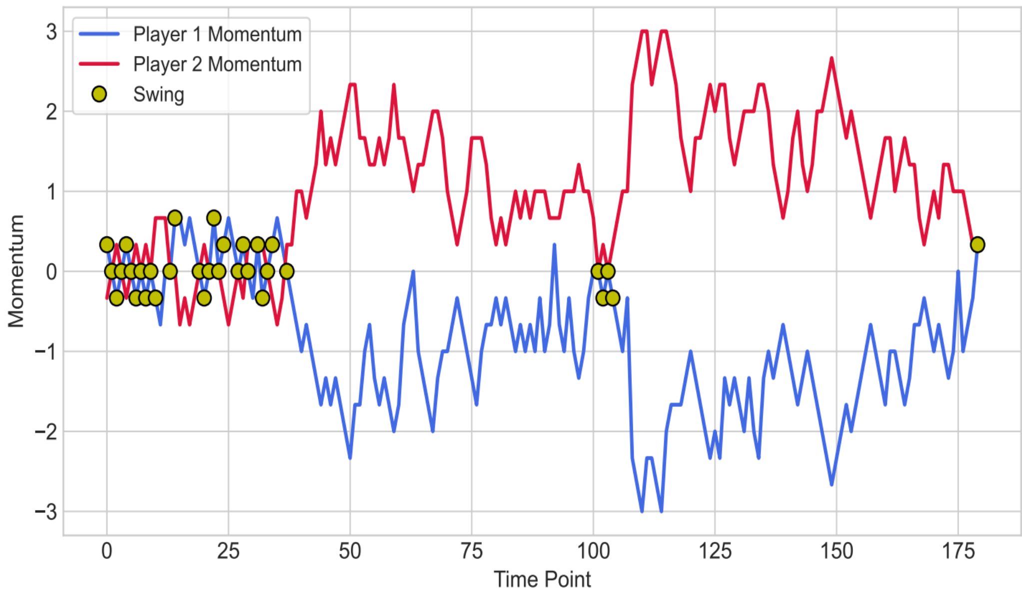

Figure 4. Match momentum and swings point over time for Alexander Zverev (Player 1) and Matteo Berrettini (Player 2)

Readers can clearly see the presence of the swing turning point in the diagram, and it can be probably seen that the early part of the match was a bit of a struggle for both players, and when the total number of points came in at 30 (i.e., the point in the match), Matteo Berrettin seized the momentum change (swing), maintained his momentum advantage, and suppressed his opponent, Alexander Zverev, into a state of disadvantage, and ultimately won the match. In this match Matteo Berrettini easily overcame Alexander Zverev, winning with a 3-0 result in the set score of the whole match. Readers can also observe this in the momentum swings, where player 2's momentum was stronger than player 1's for almost the whole match.

The performance of the prediction model was evaluated using the Mean Squared Error (MSE), which is mathematically represented as Equation (5) [9]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

where n is the number of data points, y_i is the actual value for the i^{th} data point, $\hat{y}_i$ is the predicted value for the i^{th} data point.

The MSE for player 1 is 0.00193, MSE for player 2 is 0.00194, these two values are very close to each other, indicating that the model predicts both players' win rates with similar and high accuracy.

3.3. Features Importance Analysis

Understanding the flow of a tennis match and predicting swings in momentum is of paramount importance to coaches and players. The LightGBM model was employed to analyze various factors that could influence shifts in the game, potentially favoring one player over the other.

The feature importance (*FI*) in LightGBM is calculated as follows in Equation (6):

$$FI = \frac{\sum gain\ of\ feature\ f\ across\ all\ splits}{total\ gain\ of\ all\ features} \tag{6}$$

Based on the feature importance output from the model, as depicted in Figure 5, it can draw several conclusions.

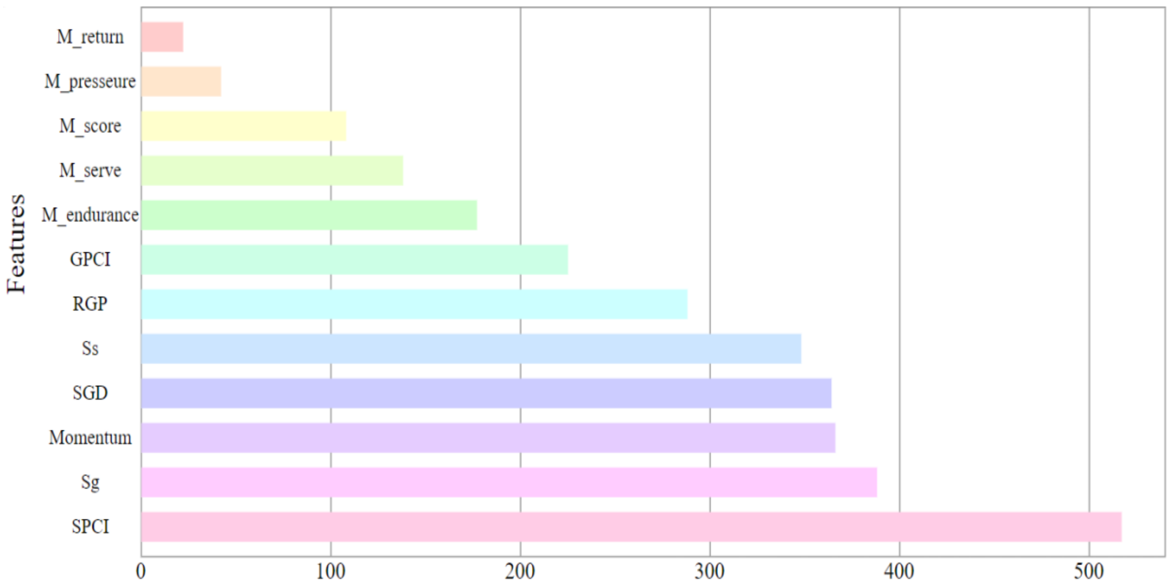

Figure 5. Features importance for the player.

The most significant indicators, as identified by the model, include:

GPCI and SPCI: These two features have the highest importance scores, indicating that the psychological state changes during the game and between sets can be precursors to swings in match flows.

Serve Game Dominance (SGD) and Serve Game Dominance (RGP): These psychological factors also rank high in importance, suggesting that a player's ability to control their service games and effectively return serves are pivotal factors for gaining and losing momentum within a match.

Momentum: This composite metric, which includes performance indicators such as point-by-point success and match progress, has a significant importance score as well.

Less influential but still notable factors include service and return effectiveness (M_{serve}, M_{return}), scoring dynamics (M_{score}), and pressure handling ($M_{pressure}$).

Surprisingly, physical endurance ($M_{endurance}$) appears to be less of a deciding factor according to the model, which may suggest that the technical and psychological aspects are more critical in the context of the analyzed matches.

The insights from this analysis could guide coaches in emphasizing specific training aspects, such as improving service games and returns and developing strategies to manage psychological states. Additionally, these findings underscored the importance of in-match performance and suggested that while physical conditioning is essential, the ability to dominate in serve and return games could be more decisive in changing the flow of matches.

In summary, the analysis points to technical prowess in serving and returning, psychological momentum control, and cumulative performance momentum as the most critical factors in predicting match swings. These insights can be instrumental for coaches and players to identify potential turning points in a match and adjust their strategies accordingly.

3.4. Strategic Preparations Considering Momentum Swings

The unpredictability of match outcomes in tennis is compounded when considering the nuanced impact of momentum swings experienced in past matches. Understanding these fluctuations is pivotal for players and coaches alike as they prepare strategies for upcoming matches against new opponents. The advice for a player entering a match is thus multifaceted and must consider the dynamic interplay of past momentum swings [10].

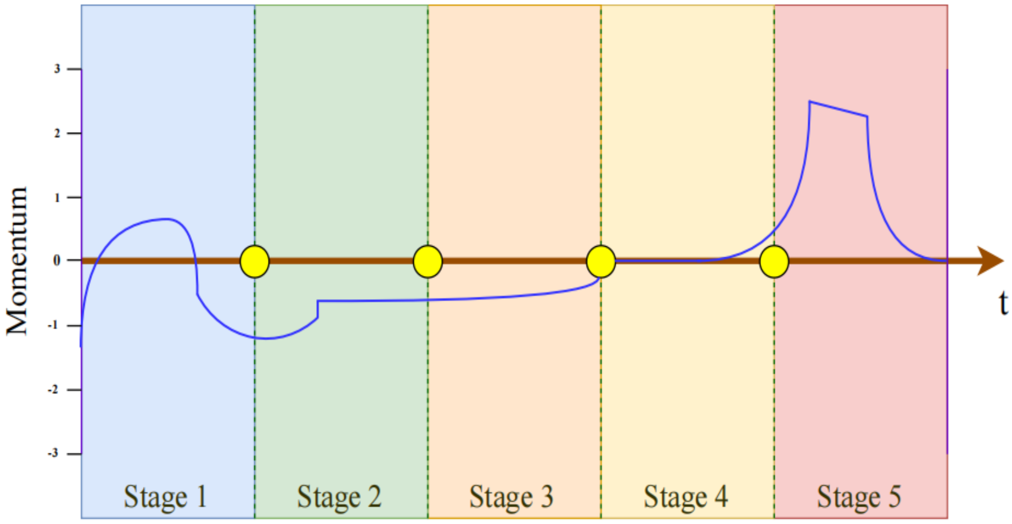

Figure 6. Illustration of the five stages of momentum in a tennis match, highlighting critical points where momentum shifts occurred.

Complete Dominance by the Opponent: In this phase, the opponent significantly dominates the play, overwhelmingly surpassing our side. If this momentum continues, it could lead to a lopsided match outcome.

Opponent's Momentum Ascending: During this period, the opponent's momentum surpasses our own, resulting in a gradually widening score gap.

Momentum Shift Between Players: This stage marks a gradual shift in momentum between the players, with our player gaining momentum and the opponent's momentum declining, thereby narrowing the score difference.

Equilibrium in Momentum: At this moment, the momentum between the players is equal, but subsequently, our player's momentum exceeds that of the opponent.

Our Ascendant Momentum: Our player's momentum exceeds the opponent's, not only equalizing but surpassing the opponent's score and progressively increasing the lead.

Analysis of past momentum: Players should thoroughly analyze their past matches to identify the stages where they gained or lost momentum. This analysis can reveal patterns in a player's game where they are most likely to experience a surge or dip in performance.

Analysis of past momentum: Players should thoroughly analyze their past matches to identify the stages where they gained or lost momentum. This analysis can reveal patterns in a player's game where they are most likely to experience a surge or dip in performance.

In-match strategies: Strategies should be flexible enough to allow players to capitalize on momentum when it is in their favor, and to stabilize the game when momentum swings against them. Quick adjustments, such as changes in serve placement or rally tactics, can be instrumental in shifting the match momentum.

Inter-match adaptation: Players must be prepared to adapt their strategies between matches based on the observed momentum patterns of their opponents. This requires a keen understanding of the opponent's psychological and physical condition, as well as their in-match behavior and tactics.

The graph depicted in Figure 6 illustrates the trajectory of momentum swings throughout the stages of a tennis match. It is evident that a player's ability to navigate these swings effectively can influence the outcome of the match significantly. Therefore, the insights gleaned from past momentum trends should inform the development of robust game plans tailored to mitigate adversities and enhance the exploitation of favorable phases during match play.

4. Conclusion

In this study, Various indicators of momentum fluctuation were established, including GPCI, SPCI, Sg_{player} Ss_{player}, etc. Then correlation analysis and stochasticity analysis of match success and swings were done, and it was found that both of them approximated the characteristics of normal distribution, and the Spearman's correlation coefficient $r_s = 0.84$, with the value of p-value less than 0.05, which indicated that the two of them had high correlation, meant that changes of swings had a significant impact on the match results and the probability of success tends to increase when swing increases.

What's more, the results of the swing prediction model based on the LightGBM algorithm show that GPCI, SECI, SGD, RGP, and other related indicators have a great impact on the performance characteristics of players in tennis matches, and the mean square error of the model prediction is 0.00194, which showed that the model had a high degree of accuracy. Combined with research results, this study provided four preparation strategies for tennis players before the match: Analysis of past momentum, Analysis of past momentum, In-match strategies, and Inter-match adaptation.

This paper provides a research idea and framework in the field of sports science & sports psychology. Combined with the performance and psychological changes of athletes in each game, using LIGHTGBM and other related algorithms to build a model, proving the feasibility of the method of predicting the match outcome by using the player's momentum swings, which provides a certain idea for athletes to prepare for the game.

References

[1] Iso-Ahola, S.E. and W. J. Blanchard, Psychological momentum and competitive sport performance: A field study. Perceptual and Motor Skills, 1986. 62 (3): p. 763 - 768.

[2] Ke, G., et al., Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 2017. 30.

[3] Wang, Y. and T. Wang, Application of improved LightGBM model in blood glucose prediction. Applied Sciences, 2020. 10 (9): p. 3227.

[4] Chen, C., et al., LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and Intelligent Laboratory Systems, 2019. 191: p. 54 -64.

[5] Sun, X., M. Liu and Z. Sima, A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters, 2020. 32: p. 101084.

[6] Razali, M.N., et al., Football matches outcomes prediction based on gradient boosting algorithms and football rating system. Human Factors in Software and Systems Engineering, 2022. 61: p. 57.

[7] Vu, Q.H., et al. A combination of general and specific models to predict victories in video games. in 2021 IEEE International Conference on Big Data (Big Data). 2021: IEEE.

[8] Palut, Y. and P. Zanone, A dynamical analysis of tennis: Concepts and data. Journal of sports sciences, 2005. 23 (10): p. 1021 - 1032.

[9] Tang, C., N. Luktarhan and Y. Zhao, an efficient intrusion detection method based on LightGBM and autoencoder. Symmetry, 2020. 12 (9): p. 1458.

[10] Briki, W., et al., How psychological momentum changes in athletes during a sport competition. Psychology of Sport and Exercise, 2013. 14 (3): p. 389 - 396.