# Beyond the Hot Hand: Skill, Experience, and Context as Determinants of Elite Badminton Performance

Vaasvi Kuthiala
Strawberry Fields High School, Chandigarh, India

## Abstract:

This study develops a predictive model for elite badminton match outcomes to identify the key performance drivers in the sport. Using a comprehensive dataset of 3,761 men's singles matches from the BWF World Tour (2018-2021), features have been engineered to capture player skill, via custom Elo rating system, experience, recent form and match context. The Elo was then benchmarked against logistic regression and an optimized XGBoost classifier, with evaluations tested on a held-out test set. The XGBoost model achieved superior prediction accuracy of 76.49%, statistically improving upon traditional methods.

Crucially, beyond predictive accuracy, the model's feature importance analysis reveals a definitive hierarchy of factors influencing wins across tournaments and varying levels. Long term player skill and career experience are the primary determinants, substantially outweighing short term influences and changes in form and hot streaks, as well as exceeding contextual factors like tournament level and qualification rounds.

These findings challenge the traditional emphasis on "hot-hand" momentum, providing data-driven evidence that sustained skill and accumulated experience are more critical for victory. The results offer a practical framework for strategic decision-making by coaches, talent scouts, and sports analysts, highlighting the value of machine learning not just for prediction, but for generating actionable insights into athletic performance.

**Keywords:** badminton, sports analytics, machine learning, prediction, XGBoost, Elo rating, feature importance, “hot hands”.

# 1. Introduction

Badminton’s global popularity is undeniable, with nearly 220 million regular worldwide players along with its status as a premier Olympic sport. At elite badminton, the margin between victory and defeat is exceptionally narrow, most often determined solely by a limited critical set of points in high pressure extreme environments. As a result of this complexity, match prediction and player evaluations have been placed in the realm of intuition from experts, experienced coaches and narratives about “momentum” and “current form” pushed by commentators. However, with the rise of sports analytics and machine learning, there is a promising shift from intuition and anecdotal assessment to augmented, objective, data-driven insights.

Applying quantitative models in sports has a rich history, from the initial development of the Elo system in chess (Elo, 1978) to creating sophisticated player tracking models now present across sports like baseball (Lewis, 2003) and basketball (Silver, 2012). Within racket sports, tennis has constantly been the primary focus of countless studies, using techniques ranging from logistic regression (Klaassen & Magnus, 2001), neural networks and ensemble methods to predict match outcomes, most often derived using ranking points and serving statistics. In badminton, on the other hand, the analytical landscape is notably less developed. Although there are some studies that apply basic statistical models, they often fall short due to limited datasets or failure in applying the insights gained to explain the underlying performance drivers.

One of the main controversies in sports analytics is the "hot hands" fallacy (Gilovich, Vallone, & Tversky, 1985), which debates whether a player on a "winning streak" possesses a predictable momentum or is it simply due to a lucky statistical coincidence. This debate, like in other sports, still remains unresolved in the context of elite level badminton, representing a significant gap in research literature.

This paper helps to address this gap by conducting a comprehensive analysis of elite level badminton. Beyond just the primary goal of prediction, this paper aims to answer a more fundamental question, what are the most important factors that determine success in elite badminton? To this end, the overall goal of the paper is twofold: first to develop and compare various predictive models, from an Elo baseline to an advanced machine learning approach, and second, to make use of the interpretability of the models to create a data-based hierarchy and importance of various features that influence performance. This can then be used to test the validity of conventional wisdom like the "hot hand" against quantifiable metrics like long-term skill and accumulated career experience.

By analyzing a comprehensive dataset of 3,761 matches from the BWF World Tour, this research provides a foundational framework for understanding badminton performance. The findings offer actionable insights for coaches, players, and analysts, while contributing to the broader sports analytics literature by validating and refining established theories in a new, dynamic context.

# 2. Materials and Methods

## 2.1 Data Collection and Description

The analysis made use of a comprehensive dataset for elite-level badminton matches, spanning from January 2018 to April 2021, with a total of 3761 matches. All data was obtained from the official BWF website using python with walkover matches excluded from the dataset. For model development, the dataset was partitioned using an 80/20 stratified split, resulting in 3,008 training matches and 753 testing matches, preserving the distribution of match outcomes in both subsets.

Models were trained on detailed statistics with 38 parameters per match. These included player-specific information, such as nationalities and identification, game-specific information such as point-by-point scoring, consecutive points and game point per game, as well as tournament-specific data such as name and level, including HSBC BWF World Tour events (Super 100 to Super 1000 levels), with match rounds ranging from qualification stages to finals. The detail and precision of information allowed detailed analysis for match prediction and robust feature engineering to capture both player skill and in-match dynamics.

**Table 1: Dataset Summary Statistics**

| **Category** | **Metric** | **Value** | **Notes** |
|---|---|---|---|
| **Scope** | Total Matches | 3,761 | Men's Singles only |
| | Time Period | Jan 2018 - Apr 2021 | 3+ years |
| | Focus Discipline | Men's singles | Controlled analysis |
| | Unique Players | 611 | |
| **Participants** | Represented Nationalities | 69 | Global representation |
| | Player experience range | 136 matches | |
| | Average matches per player | 6.02 | Meaningful career spans |
| **Competition Level** | BWF Tour Super 100 | 1,338 (35.6%) | Development circuit |
| | HSBC World Tour Super 300 | 1,053 (28.0%) | Mid-elite |

| | | | |
|---|---|---|---|
| | HSBC World Tour Super 500 | 680 (18.1%) | Upper-elite |
| | HSBC World Tour Super 750 | 340 (9.0%) | High elite |
| | HSBC World Tour Super 1000 | 305 (8.1%) | Premier events |
| | HSBC World Tour Finals | 45 (1.2%) | Season finale |
| **Match Dynamics** | 3-Set Matches | 1313 (34.9%) | High competitiveness |
| | Qualification Matches | 771 (20.5%) | Early-round analysis |
| | Avg. Total Points/Match | 83.5 | Match length consistency |

## 2.2 Feature Engineering

To successfully analyse matches, features were engineered across three categories: player skill assessment, performance trends, and match context indicators.

### 2.2.1 Elo Rating Implementation

A custom Elo rating system was implemented so that player skill could be quantified dynamically based on match performance. The system was initialized with a baseline rating of 1500, following the convention established in its original application for chess (Elo, 1978).

The K-factor, which controls how much ratings change after each match, was set to 32. This value is standard for individual sports where game-to-game volatility is expected, as it allows for meaningful skill updates without excessive fluctuation (Lasek et al., 2013). Based on every match performance, Elo was updated using the standard Elo algorithm:

$$E_{expected} = \frac{1}{1+10^{(R_{opponent}-R_{player})/400}}$$

$$R_{new} = R_{old} + K \times (S_{actual} - E_{expected})$$

where $S_{actual}$ = 1 for win and $S_{actual}$ = 0 for loss.

Players with no prior match history in the dataset were assigned the baseline Elo rating of 1500, treating them as average players until their performance data suggests otherwise, which is a standard and conservative handling of the 'cold start' problem in rating systems.

For implementation, a Google Apps Script was created and run for all 3761 matches, generating an Elo rating for all players, with values ranging between 1346.52 and 1998.85. The Elo system was selected for its proven effectiveness in individual sports and ability to dynamically capture skill evolution over time.

### 2.2.2 Performance and Experience Metrics

**Recent Performance:** Matches were processed chronologically to calculate the rolling 10-match win percentages. For each match, each player’s win rate in the previous 10 encounters was calculated. This required maintaining a window of match outcomes for all 611 players, which updated dynamically throughout the dataset. Calculating the win percentage over each player's last 10 matches captured short-term form and momentum effects, which was further used in analysis when implementing the logical regression and XGBoost.

**Career Experience:** Total career matches were computed by counting each player's frequency across the entire dataset, providing a simple but effective proxy for tournament exposure.

### 2.2.3 Tournament Level & Competition Stage

**Tournament Level:** To quantify tournament level, numeric mapping from BWF classification was used, Super 1000 = 1000, Super 300 = 300 etc.

**Competition Stage:** Qualification matches were distinguished from main draws to test whether player performance differed across qualification matches to investigate into the conventional wisdom that qualification matches are more unpredictable as a result of higher expectations and pressure.

## 2.3 Modeling Approaches

The goal of my modeling was to predict the probability of a given player winning a match, for which, I used three models, increasing in complexity and thus picking up on more nuanced and varied relationships.

**Table 2: Feature Sets Used in Predictive Models**

| **Model** | **Features Used** |
|---|---|
| Elo Prediction | Player 1 Elo Rating,<br>Player 2 Elo Rating |
| Logistic Regression | Player 1 Elo Rating,<br>Player 2 Elo Rating,<br>Tournament Level,<br>Qualification Match |
| XGBoost | Player 1 Elo Rating,<br>Player 2 Elo Rating,<br>Player 1 Total Matches,<br>Player 2 Total Matches,<br>Player 1 Recent Win %,<br>Player 2 Recent Win %, |

| | |
|---|---|
| | Tournament Level,<br>Qualification Match |

### 2.3.1 Baseline Model: Elo Rating Predictions

The baseline for analysing the matches was the Elo rating system, where predictions for match outcomes were based on a simple rule such that the player with higher pre-match Elo was predicted to win. This model closely resembles the standard practice in sports rating systems and predictive analytics, and using solely Elo allowed me to create a competitive baseline, which could then be a benchmark for evaluating more statistically complex models and ML methods.

### 2.3.2 Traditional Statistical Model: Logistic Regression

To build a stronger model for predictions, the next model applied was logistic regression as a traditional statistical model. The model used the sigmoid function:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 + x_4$$

Where $p$ represents the probability of Player 1 winning, and $x_1$through $x_4$ correspond to the four engineered features used to train the model, (is_qualification, p1_elo_rating, p2_elo_rating, tournament_level). The model was implemented using scikit-learn 1.6.1 and default parameters (C = 1.0, max_iter = 100, random_state = 42) and no feature scaling.

Feature scaling was not applied for the logistic regression model. Although scaling can be beneficial for gradient-based solvers, the main goal for this model was interpretability. By using the unscaled features, the resulting coefficients directly represent the log-odds change per unit of the original feature (e.g., per one-point increase in Elo rating). This made the model more transparent and also made the outputs more directly actionable. Through this, logistic regression helped to create a baseline that was transparent and interpretable for comparison to the simpler Elo and more complex XGBoost.

### 2.3.3 Advanced Machine Learning: XGBoost

For the final advanced stage in modelling, I employed the XGBoost (v3.0.5) model, a gradient boosting framework which is well known for its performance on tabular data. This model was trained using 8 parameters (p1_elo_rating, p2_elo_rating, p1_total_matches_played, p2_total_matches_played, is_qualification, p1_recent_win_pct, p2_recent_win_pct, tournament_level), double of the parameters used for logistic regression. Hyperparameter tuning for the XGBoost model was performed using BayesSearchCV from the scikit-optimize library over 50 iterations.

The search optimized the 'binary:logistic' objective function, exploring the following parameter space: n_estimators (50-300), max_depth (3-10), learning_rate (0.01-0.3), subsample (0.6-1.0), and colsample_bytree (0.6-1.0). The final optimized model used 217 estimators, max_depth = 3, learning_rate = 0.055, subsample = 0.6, and colsample_bytree=0.6. The relatively shallow tree depth suggests that the model captured meaningful interactions of features without overfitting to any statistical noise in the training data.

## 2.4 Validation Framework

For validation, I employed a 80/20 train-test stratified split. This stratification preserves the distribution of match outcomes in both subsets, ensuring a representative sample for both model training and evaluation. Model performance was evaluated mainly using prediction accuracy as a primary metric, along with a 5 fold cross-validation to assess robustness and generalisation capacity, which reported a mean accuracy ± standard deviation across folds.

# 3. Results

This section presents the evaluation of the performance of all three predictive models, the Elo-rating, logistic regression, and tuned XGBoost, on the held-out test set of 753 matches. The results demonstrate and show a clear hierarchy and ranking in predictive capability. The machine learning model, XGBoost, as expected, shows a clear superiority over the logistic regression and Elo rating baseline.

## 3.1 Overall Predictive Performance

All three predictive models significantly outperformed the naive starting point of random guessing (50.30% accuracy), which demonstrates that the engineered features are capable of capturing meaningful signals for predictive analysis. As shown in Table 3, a clear performance hierarchy emerged, with model complexity correlating with accuracy.

**Table 3: Model Comparison**

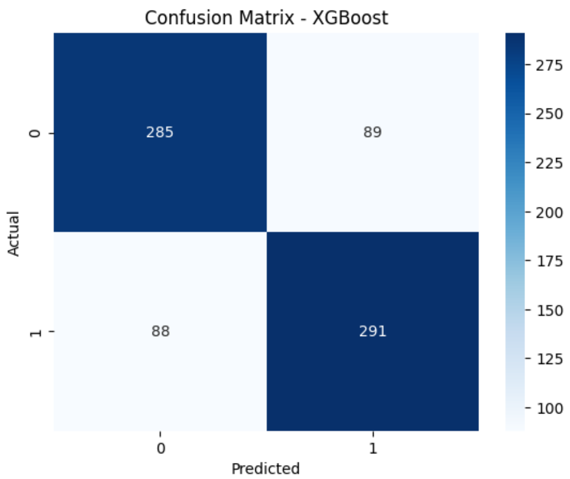
| **Model** | **Accuracy** | **Improvement vs Baseline** | **Key characteristic** |
|---|---|---|---|
| Random Guessing | 50.3% | - | Theoretical minimum |
| Elo Rating System | 70.19% | +19.89 | Player skill based |
| Logistic Regression | 72.11% | +21.81 | Statistical modeling |
| XGBoost (Optimized) | 76.49% | +26.19 | Ensemble machine learning |

The tuned XGBoost model emerged as the best-performing predictor, achieving a test accuracy of 76.49%. This represents a statistically significant improvement of 4.38 percentage points over the logistic regression model ($p < 0.0001$, McNemar's Test) and a 6.30-point improvement over the simple Elo baseline. The model's strong performance was further validated and boosted by its average accuracy of 73.92%(± 2.30%) during tuning, demonstrating its robust generalisability and low variance.

The gap in performance between the optimized XGBoost and Elo baseline as well as logistic regression show statistical significance, although they may seem numerically low. This has helped to demonstrate how machine learning models are adept at capturing non-linear relationships and the complex and intricate variance and relationships between features that could easily bypass rule-based systems like the Elo and statistical models like logistic regression.

## 3.2 Model Calibration and Detailed Classification

Apart from raw accuracy and precision, the XGBoost model also demonstrated well-calibrated predictions as evidenced by a low Log Loss of 0.485. The confusion matrix and classification report provide a more nuanced view of its performance.




*Note: 0 = Player 1 Loss, 1 = Player 1 Win*

**Figure 1. Confusion matrix for the tuned XGBoost model on the test set.**

The model showed equivalent performance across the two classes, with nearly identical precision and recall for wins as well as losses:

**Precision:** 77% for predicting wins, 76% for predicting losses

**Recall:** 77% for wins, 76% for losses

**F1-Score:** 0.77 for both classes

This balance indicates that the model does not exhibit a significant bias towards either outcome, which is a crucial characteristic for a reliable forecasting tool.

## 3.3 Determinants of Match Outcomes: Feature Importance Analysis

The relative importance of features of the tuned XGBoost model helps to create a data-driven ranking of the factors that truly influence match outcomes in elite level badminton. This in-depth understanding allows statistical analysis to move beyond mere predictions, to actually create a quantitative understanding of the sport's dynamics.

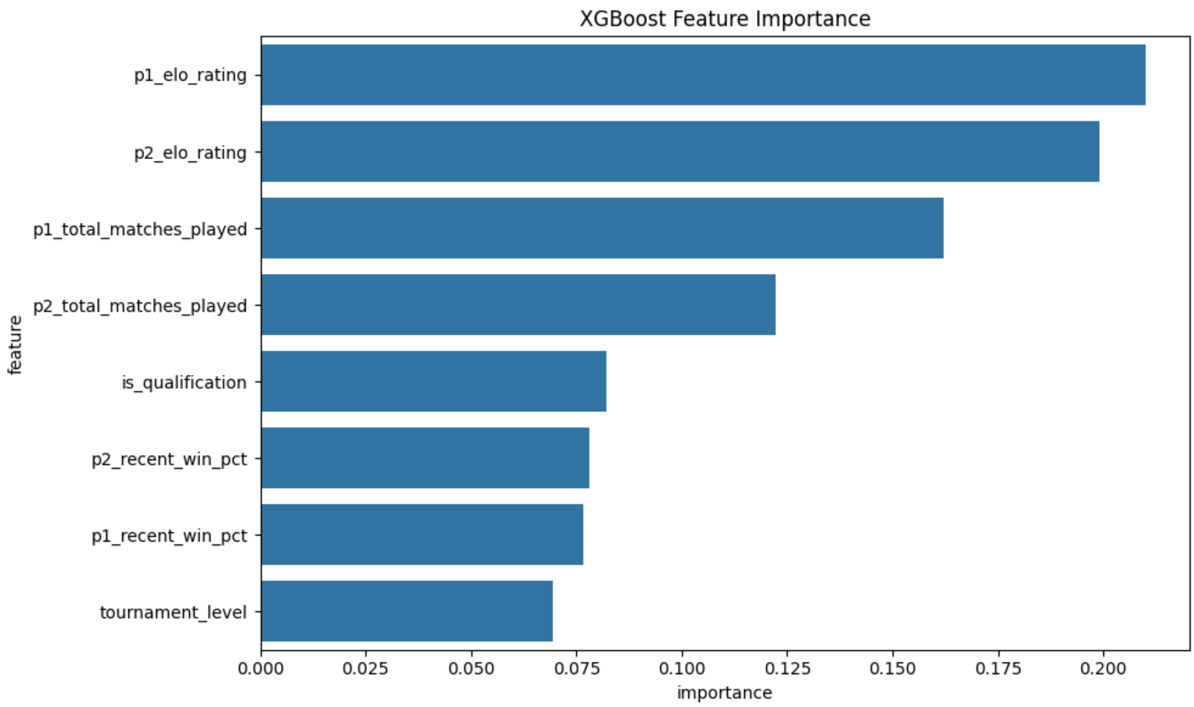



Figure 2. Relative feature importance from the tuned XGBoost model

**Table 4: XGBoost Feature Importance Rankings**

| **Feature** | **Importance** | **Category** | **Business Interpretation** |
|---|---|---|---|
| Player 1 Elo Rating | 20.99% | Skill | Overall player quality |
| Player 2 Elo Rating | 19.91% | Skill | Opponent strength |
| Player 1 Total Matches | 16.21% | Experience | Career development |
| Player 2 Total Matches | 12.23% | Experience | Opponent experience |
| Qualification Match | 8.23% | Context | Pressure environment |

| | | | |
|---|---|---|---|
| Player 2 Recent Win % | 7.82% | Form | Opponent momentum |
| Player 1 Recent Win % | 7.67% | Form | Current performance |
| Tournament Level | 6.94% | Context | Competition quality |

### 3.3.1 The Primacy of Established Skill (Elo Ratings)

The combined importance of Player 1 and Player 2’s Elo ratings, a total of 40.9% in importance, establish that long-term skill is the single most critical factor in match predictions. This further helps to consolidate the Elo rating system’s effectiveness in acting as a measure of a player’s intrinsic skill in badminton. This symmetry in the importance of the skill of not only Player 1 but also Player 2, indicates that the match, at its core, is a contest of relative skill levels. The model also gave an interesting insight that although a higher Elo doesn’t always mean victory, it does establish a powerful probability.

### 3.3.2 The Critical Role of Career Experience

Surprisingly, the combined feature of both players’ total matches played (combined 28.4%) emerged as the second most influential factor, significantly outweighing the impact of short-term performance and form in terms of Recent Win Percentage.

This finding questions the conventional wisdom that “hot hands” and recent momentum are the primary drivers of success and players with a “Winning Streak” are more likely to continue winning. Instead, my research reveals that long term experience, gained from hundreds of high-pressure environments, by adjusting to varying playing styles, players gain a significant tactical advantage that helps to separate raw talent from accumulated experience.

### 3.3.3 The Contextual Over the Transient: Match Context vs. Recent Form

One of the most interesting findings was the clear hierarchy of context and recent form. The fact that the indicator for qualification matches(8.2%) proved to be a more important feature for prediction, compared to the recent win percentage can be interpreted as:

1. **Pressure Environment:** Qualification matches generally carry high-stakes and thus cause a lot of pressure on players, where an opportunity to enter the main draw will lead to elevated performance from underdogs, as well as heightened pressure on long-term favourites, increasing volatility.
2. **Data Artifact:** "Recent Win Percentage" is possibly a nosier metric that solely captures streaks that may have originated due to weaker opponents or other transient factors, while "qualification" signals a reliable and specific match context
3. **Tournament Level's Influence:** Tournament level, although it is the least important feature, it still contributes meaningfully in helping identify that players perform consistently and relative to their intrinsic skill and experience, however, the prestige of an event adds a layer of contextual meaning, influencing decisions.

### 3.3.4 The Asymmetry of Features

The analysis also revealed subtle asymmetries, for example, Player 1's experience (16.2%) is valued more highly than Player 2's experience (12.2%). This makes sense, since the match outcome from a player's perspective does logically depend more on their experience and skill than their opponent's, hinting at a psychological advantage inherent in how the data is structured. Similarly, Player 2's recent form (7.8%) is marginally more important than Player 1's (7.7%), which could suggest that the opponent's momentum is a slightly more important feature for prediction compared to the player's own recent winning streak. These asymmetries warrant further investigation in future research.

# 4. Discussion

## 4.1 Strategic Implications and Applications

The primary contribution of this research isn't predictive accuracy, but creating a data-driven decision making framework for competitive badminton globally. The analysis of the importance of features using the XGBoost model has helped to develop an unambiguous system to determine what actually guides elite badminton and at the same time challenging long held beliefs and assumptions.

### 4.1.1 A Paradigm Shift in Talent Scouting and Development

The model reveals that long-term skill (Elo) and career experience (total matches) are over four times more important than short-term form (recent win percentage). From this, we can gain the following insights:

1. **From "Form" to "Trajectory" in Scouting:**
   Traditional scouting often relies overly on a player's recent performance in the last 3-5 tournaments, however, this model provides a framework for a "Trajectory Score" that combines a player's Elo rating over the past 24 months to the total volume of matches they have played. Thus, a player with a steadily rising Elo from 1500 to 1700 over 100 matches is a more valuable and reliable asset than a player who jumped from 1500 to 1750 after a single, potentially lucky tournament against lower ranked opponents. This will allow scouts to identify players who are genuinely improving their core skills compared to those experiencing temporary variance.
2. **Quantifying the ROI of Competitive Exposure for Academies:** The importance of total matches played (16.2% for Player 1) allows academies to move beyond gut-feeling and speculation to a data-backed strategy. For example, sending a cohort of 10 promising players to a lower tier Super 100 event can be justified by calculating the aggregate value added to their profiles through "experience". This model argues that the long-term benefit of

accelerating a player's experience curve often outweighs the short-term cost and lack of prestige in these events.

### 4.1.2 Data-Driven Strategy for Coaches and Players

**The "Volatility Index" for Match Preparation:** Coaches can make use of the insights created by the model by calculating a simple Volatility Index for any upcoming match. The index would be high when a match combines a high-stakes context (Super 1000, qualification match) with a player who is susceptible to pressure. For example,

*Volatility Index = (Low Player Experience + High Stakes Context) - (Large Elo Difference)*

A high risk match, flagged by a high Volatility Index, can be prepared for by coaches by a special protocol, for example, extended video analysis focused on the opponent's tactical patterns under pressure, practice sessions dedicated to simulating high-pressure scenarios (e.g., playing points from 16-16 with consequences), and a game plan that will emphasise simple, high-percentage shots to neutralize pre-match nerves and stabilize performance

**A Triage System for Analytical Effort:** Instead of spending equal time on all opponents, the feature importance provides a clear, efficient system for analytical resources:

- **Tier 1 Analysis (The Foundation - 69.3% of signal):** This will be a mandatory part of the evaluation for all opponents, it could include a deep analysis of the opponent’s Elo history, learning capacity, and match volumes to understand the accumulated experience, whether they are veteran players with hundreds of matches or new comers, who could be volatile underdogs.
- **Tier 2 Analysis (The Context - 15.2% of signal):** This could be activated for specific scenarios, including an assessment focussed on how an opponent typically performs in qualification matches or early-round matches against unknown opponents, to better understand psychological factors.

    - **Tier 3 Analysis (The Noise - 15.5% of signal):** A brief review of recent matches and performance, to be aware of changes in play style or new tactics being employed, for example, a new service motion, in place of simply a win/loss result. This would allow coaches to include tactical focus and not solely depend on transient outcomes.

### 4.1.3 For Sports Betting and Predictive Markets: Building an Analytical Edge

In professional betting, rather than predicting every match correctly, it's advantageous when a bettor can identify where the public model is incorrect, and the research done in this paper can provide a framework to do exactly that.

- **A "Model vs. Market" Screener:** A bettor could make use of this model to screen for discrepancies by calculating the model's probability for a player to win and comparing it to the probability offered by a bookmaker and it will create the most significant opportunities when:
    - **The Market Overvalues Recent Form:** The public may tend to give shorter odds to a new favourite because they might have won the previous 3-5 matches, however, the model will recognise this and provide a better advantage.
    - **The Market Undervalues Experience and Context:** For an underdog with 400+ career matches competing in a qualification match, the model might assign 8% "pressure volatility" to boost their chances, ignored by the market, elongating their odds.
- **Parlay and Fading Strategies:** The model's high accuracy (76.5%) makes it reliable for waging confident parlay legs. Additionally, its ability to identify high volatility matches, creates "fade" opportunities, allowing bettors to actively bet against the public favourite.

#### 4.1.4 For Broadcast and Fan Engagement: The New Narrative Toolkit

For media and streaming services, models like the one from this paper, could play a very insightful role to create deeper, more engaging content to cater to a more data-savvy, modern and aware audience.

- **Data-Enriched Storytelling:** Broadcasters can move beyond pure statistics, for example, "He has a 5-2 head-to-head record", instead commentary can include specific details, for instance "While the younger player comes in with a hot streak and good momentum, our analytics give the edge to the veteran here whose 500-match career has built a resilience, crucial for these high-pressure qualification matches." Through this, audiences would be able to move beyond simple numbers and understand more complex nuances of matches. This would also help in deepening their appreciation and understanding and providing them the full story backed with data.

- **Interactive Fan Engagement:** Second-screen applications could be used along with on-screen graphics in real time, with the model. Rather than just showing a score, the broadcaster could display a live "Win Probability" that fluctuates based on the core features of the players involved, not just the scores. A fan could see: "Win chance dropped 10% due to opponent's momentum (Recent Win % factor)", educating views about subtle, non-scoreboard factors that influence the sport.

### 4.2 Theoretical Contributions: Challenging Conventional Wisdom in Sports Analytics

Beyond its practical utility, this study makes two significant theoretical contributions that challenge established narratives in sports performance analysis.

### 4.2.1 Quantifying the "Hot Hand" Fallacy in Racket Sports

The relatively low importance of the recent win percentage (15.5%) cumulatively for both players, compared to long term skill and experience provides strong evidence for a “hot hand” fallacy. “Hot hand” fallacy, although it is debated heavily in sports like basketball, the results from this study show that even in the context of an individual, “form”/ “momentum” is often overvalued and wrongly attributed as a result of statistical noise or weaker opponents.

This doesn’t disregard the psychological impact that momentum and  “hot streaks” may have, but it shows that it is not as important as a predictive feature, and is statistically shadowed by a player’s raw skill and experience. This finding urges a re-evaluation of how “form” is weighed both analytically and intuitively for match predictions.

### 4.2.2 Experience as a Quantifiable Intangible Skill

The strong performance of total career matches as a predictive feature with 28.4% importance, makes “experience” not just a vague cliché, but a measurable indicator for performance. This metric is also a proxy for intangible skills which are difficult to capture solely off traditional statistics, but are critical in evaluating match predictions:

- **Pressure Management:** The effect of navigating hundreds of high stakes points and matches, which creates a distinguishing resilience from raw skills and ability.
- **Strategic Adaptation:** Additionally, players with a long term experience have played a wider variety of play styles as well as game situations, which allows them to make faster and more effective in-match tactical changes.
- **Tournament Recovery:** The physical and mental endurance required to compete repeatedly across a long career is itself a skill that contributes to consistent performance.

This shows how in elite level badminton, experience isn’t just a subtle detail, but is a distinct and measurable dimension of athletic ability and capacity that is built from a volume of matches and competitions and contributes to providing a significant and measurable advantage.

## 4.3 Limitations and Avenues for Future Research

The research conducted for this paper, while it does provide a powerful macroscopic view, there are still certain limitations that could be overcome upon further research. The analysis is based on match outcomes and does not currently incorporate in-play details like shot type, player placement, movement. Additionally, the Elo system implements standard parameters; however, future work could develop a badminton-specific rating system incorporating margin of victory or surface type (indoor vs. outdoor).

- **Integration of Tracking Data:** Future work could integrate player and shot tracking data from official providers to unlock features related to technical prowess (e.g., smash speed, shot placement).
- **Optimized Rating Systems:** Developing an Elo system specific to badminton related features and analytics, potentially with a dynamic K factor, or a margin of victory component, could further help to improve the skill assessment for players.
- **Expansion to Other Disciplines:** Applying this framework to women's singles and doubles matches, would help to further test the applicability of the model and assess the universality of performance drivers across the sport

# 5. Conclusion

This study successfully demonstrates that machine learning models, particularly XGBoost, can effectively predict elite badminton match outcomes with 76.49% accuracy, outperforming traditional methods like Elo ratings and logistic regression. More importantly, analysing feature importance has helped to understand the hierarchy of various performance drivers and how long-term skill and career experience are paramount, while short-term form and match context are secondary factors.

The findings challenge conventional wisdom by providing evidence against the predictive value of the "hot hand" narrative, instead revealing that accumulated experience and consistent skill development are more reliable indicators. By moving beyond predictions, to uncover the fundamental determinants of victory, this research offers a practical framework for coaches, scouts, analysts, and broadcasters to make more informed decisions.

The methodology and insights not only increase the understanding of badminton performance but also contribute to the broader sports analytics literature by validating established theories in newer contexts. This work establishes a foundation for more sophisticated analytics and highlights the value of machine learning for generating strategic insights beyond simple prediction for underrepresented sports like badminton.

# 6. Acknowledgements

I would like to express my gratitude to the Badminton World Federation (BWF) for making such comprehensive match data publicly available, making this research possible. I also extend my appreciation to the developers of the open-source Python libraries,including scikit-learn, XGBoost, and scikit-optimize, all of which were instrumental in implementing the analytical models.

I am thankful to my computer science teachers at Strawberry Fields High School for fostering my interest in data analytics and machine learning. Finally, I would like to acknowledge the Wharton Sports Analytics Journal for providing a platform that encourages student research in sports analytics.

# 7. References

Elo, A. E. (1978). The rating of chessplayers, past and present. Arco Publishing.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

Klaassen, F. J. G. M., & Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454), 500–509. https://doi.org/10.1198/016214501753168217

Lasek, J., Szlávik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1(1), 27–46. https://doi.org/10.1504/IJAPR.2013.052339

Lewis, M. (2003). Moneyball: The art of winning an unfair game. W. W. Norton & Company.

Silver, N. (2012). The signal and the noise: Why so many predictions fail—but some don't. Penguin Press