Predicting ATP Player Tennis Performance Using Machine Learning

Arjun Kamra

By Arjun Kamra, Jesuit High School, Carmichael, CA

Abstract

Predicting performance in professional tennis has applications for analysts, coaches, and betting strategies. This study investigates whether machine learning models can accurately predict Association of Tennis Professionals (ATP) match outcomes based on historical player data. Statistics such as serve percentage, break point conversion, and win-loss records were compiled from Ultimate Tennis Statistics and ATP Tour databases. Three models were tested: linear regression, logistic regression, and random forest classifiers. Model performance was compared using accuracy, R^2, and F1 score. Random forest achieved the highest accuracy (93.36%), followed by logistic regression (91.15%), while linear regression produced a moderate correlation ($R^2 = 0.544$). Serve consistency and break point conversion emerged as key indicators of success. Results suggest that ensemble-based models are most effective for capturing the non-linear relationships in tennis performance. While the study is limited by reliance on career-level statistics, these findings highlight the potential for machine learning to enhance tennis analytics, strategic planning, and predictive applications.

Introduction

Tennis, a sport that intricately combines physical prowess and strategic acumen, has long been a subject of analytical studies aiming to predict match outcomes. The advent of machine learning has significantly enhanced these predictive endeavors, offering sophisticated tools to analyze vast datasets and uncover patterns not immediately discernible through traditional statistical methods. Early attempts to forecast tennis match results primarily relied on regression models that utilized player rankings and basic statistics. For instance, Boulier and Stekler developed one of the initial regression models to predict tennis matches winners based on ATP rankings in 2003 *(Boulier and Stekler 2003)*. These models often fell short in capturing the specific performance patterns that influence match outcomes. They struggled with non-linear relationships between variables and lacked accuracy when applied across diverse player profiles. This study aims to address these gaps by testing machine learning models that can better predict ATP match outcomes using detailed player statistics.

The integration of machine learning techniques has opened new avenues for more accurate predictions. De Seranno proposed a machine learning approach that extracted 84 features from historical ATP data, including player performance metrics and match statistics, to predict the winner of ATP singles matches *(De Seranno 2019)*. Logistic regression outperformed models based on ATP rankings, improving both accuracy and calibration. Further improvements came from using a neural network, which showed stronger predictive performance and reduced error.

Similarly, Gao and Kowalczyk employed a random forest model to predict tennis match outcomes, identifying serve strength as a pivotal predictor *(Gao and Kowalczyk 2020)*. Their model achieved an accuracy exceeding 80%, underscoring the efficacy of ensemble learning methods in sports analytics. Cai et al. also applied machine learning to predict Grand Slam match outcomes and validated the model's applicability to major tournaments. McHale and Morton used Bradley–Terry-type models for tennis match forecasting, showing the effectiveness of probabilistic pairwise comparisons. Knottenbelt et al. introduced a common-opponent stochastic model that demonstrated improved accuracy by incorporating shared opponent history. Beyond tennis, machine learning models such as logistic regression and random forests have been successfully applied to predict outcomes in various sports, including football and basketball. These studies highlight the versatility of machine learning in handling datasets that differ by collection period, player ages, tournament types, and global locations. Such variation tests model reliability across contexts. In contrast, this study uses career-level ATP data for each player, offering a consistent dataset to evaluate long-term performance trends and match outcomes

This study builds upon existing research by leveraging machine learning models to predict ATP tennis player performance, focusing on match outcomes. By analyzing comprehensive career-long statistics—including serve percentages, points won, net effectiveness, and match win percentages—this research identifies key performance indicators and assesses the predictive power of different machine learning models, specifically Logistic Regression and Random Forest Classifiers. It is hypothesized that the Random Forest model will outperform Logistic Regression due to its non-parametric nature, which allows it to capture complex, non-linear patterns in player performance data.

Methods

This study used two classification models, Logistic Regression and Random Forest, to predict ATP match outcomes based on historical player statistics. Data was compiled from 24 CSV files containing career-level metrics for each player. Features included ace count, double faults, first serve effectiveness, win percentages, return ratings, and ATP rankings *(Ultimate Tennis Statistics; ATP Tour)*.

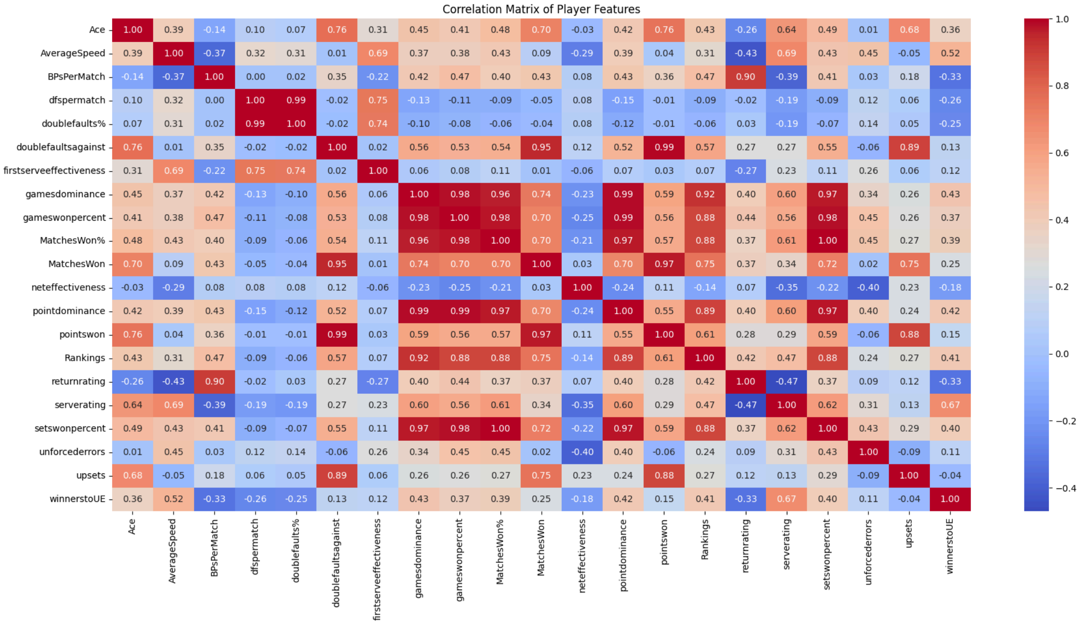
After merging and cleaning the data, players with missing values were removed. All possible player matchups were generated, and each matchup was labeled using ATP rankings. A matchup was labeled 1 if player one had a higher ranking than player two, and 0 otherwise.

For each matchup, the features of both players were combined into a single input vector. The dataset was split into 80% training and 20% testing. Features were scaled using StandardScaler for logistic regression. The Random Forest model used 500 trees, a maximum depth of 6, and considered 50% of features at each split.

Model performance was evaluated using accuracy, F1 score, and precision-recall curves. Additional analysis included a correlation matrix of features and F1 score plotted against decision thresholds.

Results

A correlation matrix revealed that features such as points won, match win percentage, games won percent, and serve rating were most strongly correlated with player rankings.



	Ace	AverageSpeed	BPsPerMatch	dfspermatch	doublefaults%	doublefaultsagainst	firstserveeffectiveness	gamesdominance	gameswonpercent	MatchesWon%	MatchesWon	neteffectiveness	pointdominance	pointswon	Rankings	returnrating	serverating	setswonpercent	unforcederrors	upsets	winnerstoUE
Ace	1.00	0.39	-0.14	0.10	0.07	0.76	0.31	0.45	0.41	0.48	0.70	-0.03	0.42	0.76	0.43	-0.26	0.64	0.49	0.01	0.68	0.36
AverageSpeed	0.39	1.00	-0.37	0.32	0.31	0.01	0.69	0.37	0.38	0.43	0.09	-0.29	0.39	0.04	0.31	-0.43	0.69	0.43	0.45	-0.05	0.52
BPsPerMatch	-0.14	-0.37	1.00	0.00	0.02	0.35	-0.22	0.42	0.47	0.40	0.43	0.08	0.43	0.36	0.47	0.90	-0.39	0.41	0.03	0.18	-0.33
dfspermatch	0.10	0.32	0.00	1.00	0.99	-0.02	0.75	-0.13	-0.11	-0.09	-0.05	0.08	-0.15	-0.01	-0.09	-0.02	-0.19	-0.09	0.12	0.06	-0.26
doublefaults%	0.07	0.31	0.02	0.99	1.00	-0.02	0.74	-0.10	-0.08	-0.06	-0.04	0.08	-0.12	-0.01	-0.06	0.03	-0.19	-0.07	0.14	0.05	-0.25
doublefaultsagainst	0.76	0.01	0.35	-0.02	-0.02	1.00	0.02	0.56	0.53	0.54	0.95	0.12	0.52	0.99	0.57	0.27	0.27	0.55	-0.06	0.89	0.13
firstserveeffectiveness	0.31	0.69	-0.22	0.75	0.74	0.02	1.00	0.06	0.08	0.11	0.01	-0.06	0.07	0.03	0.07	-0.27	0.23	0.11	0.26	0.06	0.12
gamesdominance	0.45	0.37	0.42	-0.13	-0.10	0.56	0.06	1.00	0.98	0.96	0.74	-0.23	0.99	0.59	0.92	0.40	0.60	0.97	0.34	0.26	0.43
gameswonpercent	0.41	0.38	0.47	-0.11	-0.08	0.53	0.08	0.98	1.00	0.98	0.70	-0.25	0.99	0.56	0.88	0.44	0.56	0.98	0.45	0.26	0.37
MatchesWon%	0.48	0.43	0.40	-0.09	-0.06	0.54	0.11	0.96	0.98	1.00	0.70	-0.21	0.97	0.57	0.88	0.37	0.61	1.00	0.45	0.27	0.39
MatchesWon	0.70	0.09	0.43	-0.05	-0.04	0.95	0.01	0.74	0.70	0.70	1.00	0.03	0.70	0.97	0.75	0.37	0.34	0.72	0.02	0.75	0.25
neteffectiveness	-0.03	-0.29	0.08	0.08	0.08	0.12	-0.06	-0.23	-0.25	-0.21	0.03	1.00	-0.24	0.11	-0.14	0.07	-0.35	-0.22	-0.40	0.23	-0.18
pointdominance	0.42	0.39	0.43	-0.15	-0.12	0.52	0.07	0.99	0.99	0.97	0.70	-0.24	1.00	0.55	0.89	0.40	0.60	0.97	0.40	0.24	0.42
pointswon	0.76	0.04	0.36	-0.01	-0.01	0.99	0.03	0.59	0.56	0.57	0.97	0.11	0.55	1.00	0.61	0.28	0.29	0.59	-0.06	0.88	0.15
Rankings	0.43	0.31	0.47	-0.09	-0.06	0.57	0.07	0.92	0.88	0.88	0.75	-0.14	0.89	0.61	1.00	0.42	0.47	0.88	0.24	0.27	0.41
returnrating	-0.26	-0.43	0.90	-0.02	0.03	0.27	-0.27	0.40	0.44	0.37	0.37	0.07	0.40	0.28	0.42	1.00	-0.47	0.37	0.09	0.12	-0.33
serverating	0.64	0.69	-0.39	-0.19	-0.19	0.27	0.23	0.60	0.56	0.61	0.34	-0.35	0.60	0.29	0.47	-0.47	1.00	0.62	0.31	0.13	0.67
setswonpercent	0.49	0.43	0.41	-0.09	-0.07	0.55	0.11	0.97	0.98	1.00	0.72	-0.22	0.97	0.59	0.88	0.37	0.62	1.00	0.43	0.29	0.40
unforcederrors	0.01	0.45	0.03	0.12	0.14	-0.06	0.26	0.34	0.45	0.45	0.02	-0.40	0.40	-0.06	0.24	0.09	0.31	0.43	1.00	-0.09	0.11
upsets	0.68	-0.05	0.18	0.06	0.05	0.89	0.06	0.26	0.26	0.27	0.75	0.23	0.24	0.88	0.27	0.12	0.13	0.29	-0.09	1.00	-0.04
winnerstoUE	0.36	0.52	-0.33	-0.26	-0.25	0.13	0.12	0.43	0.37	0.39	0.25	-0.18	0.42	0.15	0.41	-0.33	0.67	0.40	0.11	-0.04	1.00

Figure 1. Correlation matrix of player-level features. Darker red values indicate stronger positive correlations; darker blue values indicate negative correlations.

These findings align with the feature importance scores observed during training and confirm their relevance in predictive modeling. Three models were trained and evaluated to assess their ability to predict ATP match outcomes based on aggregated player statistics.

Linear regression achieved an R-squared value of 0.544, indicating a moderate linear relationship between features such as first serve percentage, break point conversion rate, ace-to-double fault ratio, and ATP rankings. While not suitable for classification tasks, the results provided insight into which variables were most closely associated with long-term success. Among the variables analyzed, first serve percentage, break point conversion rate, and ace-to-double fault ratio showed the strongest positive correlation with long-term ATP success.

Logistic regression was used to classify match outcomes based on career-level features. It achieved a test accuracy of 91.15%, meaning the model was 91.15% accurate in predicting …. Precision-recall analysis showed strong performance, with the model maintaining high precision across most recall levels.

Model	Accuracy (%)	F1 Score	R²
Linear Regression	–	–	0.544
Logistic Regression	91.15	0.93	–
Random Forest	93.36	0.93	–

Table 1. Model performance on the held-out test set. Logistic and Random Forest are evaluated as classifiers (Accuracy, F1). Linear regression is reported with R².

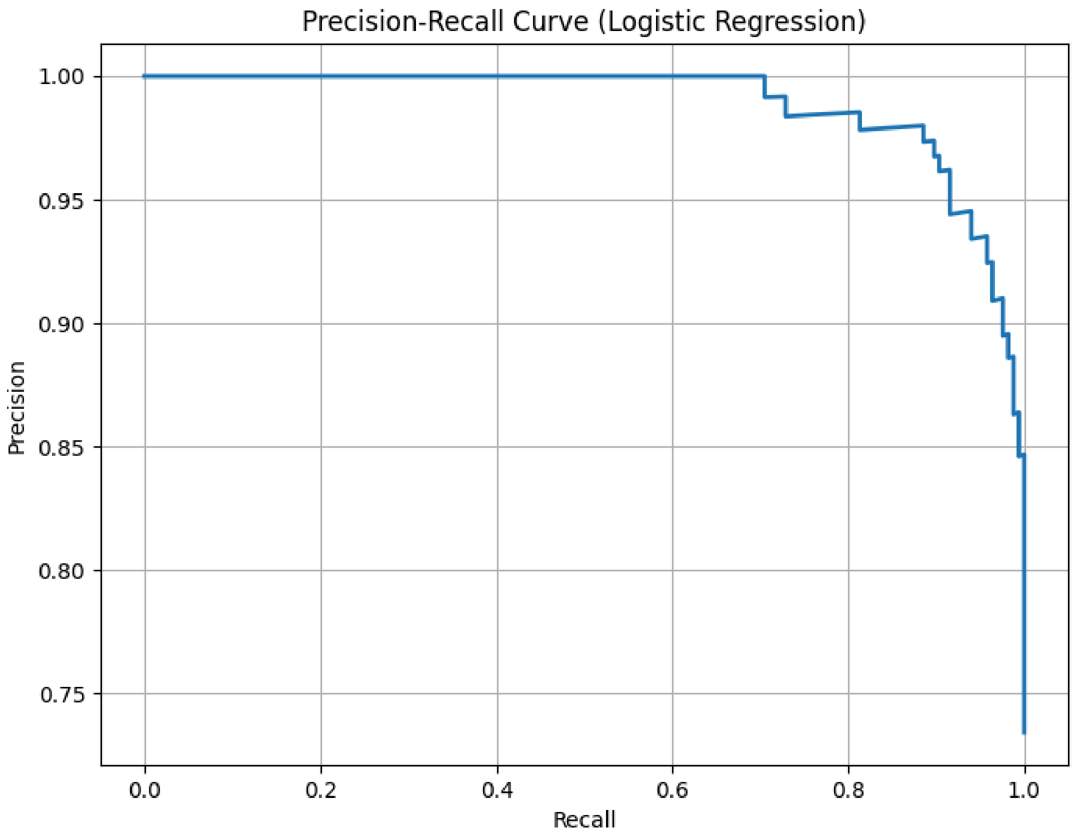


Figure 2. Precision–Recall curves for Logistic Regression and Random Forest classifiers on the test set. Random Forest achieved superior precision at higher recall levels.

Additionally, an F1-score versus threshold plot demonstrated that the model's performance peaked around a threshold of 0.5, where the F1-score reached approximately 0.93. This indicates that the model achieved a strong balance between precision and recall, making it effective at predicting match outcomes without favoring false positives or false negatives. The top contributing features in this model were first serve percentage, break point conversion rate, and ace-to-double fault ratio, which had the highest coefficients.

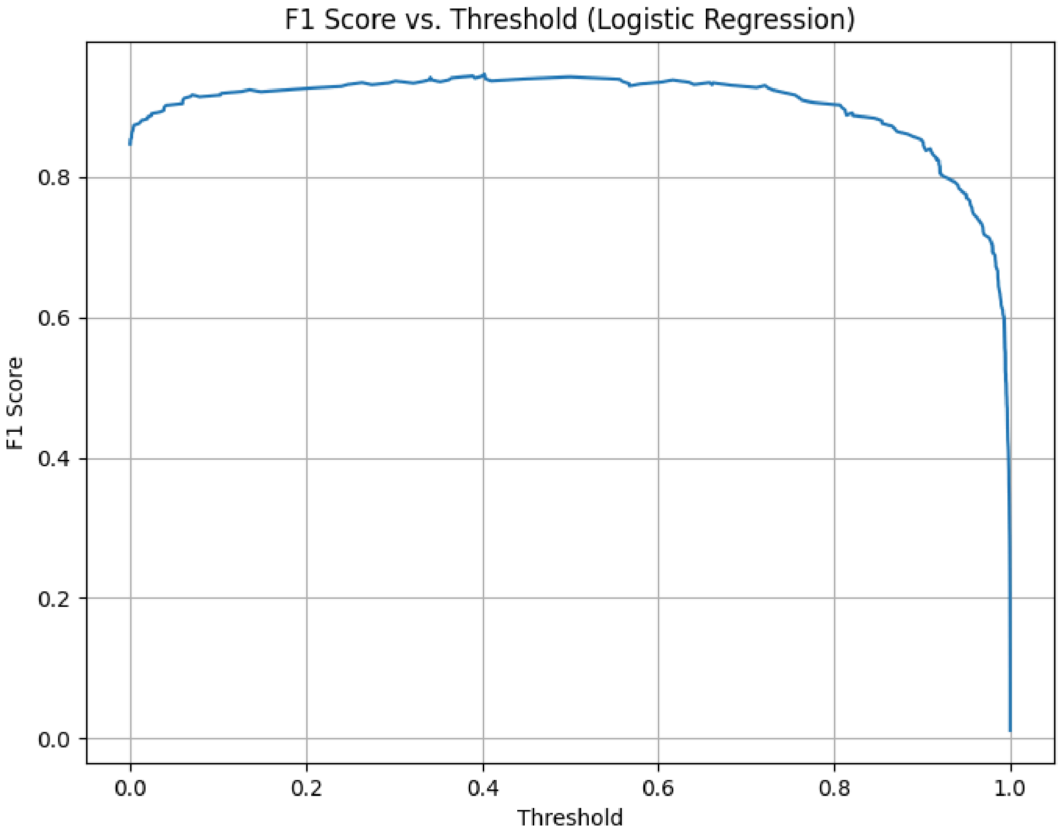


Figure 3. F1 score versus decision threshold for Logistic Regression. Performance peaked near 0.5, balancing precision and recall.

The Random Forest Classifier outperformed both previous models, reaching a test accuracy of 93.36%. Its ability to capture non-linear relationships between features made it effective for this dataset, which included varied performance metrics with complex interactions. For example, serve percentage and break point conversion may impact outcomes differently depending on a player's overall style or consistency. The Random Forest model handled these interactions without assuming a fixed form, improving accuracy across diverse player profiles. The model showed robustness across various thresholds and proved to be the most reliable approach for predicting outcomes using long-term performance data, as seen by….

To improve model performance and reduce overfitting. Overfitting would have caused the model to perform well on training data but fail to predict new matchups accurately, reducing its value for real-world forecasting. Several hyperparameters were tuned using manual experimentation and grid-style variation. For the Random Forest Classifier, the number of trees was a key parameter. Increasing the number of estimators generally improved performance, with diminishing returns after 500 trees. A value of 500 trees was selected as the optimal point balancing predictive accuracy and computational efficiency. The maximum tree depth was also tested across a range of values. A depth of six provided the best results, capturing enough complexity without overfitting to the training data. Finally, the max_features parameter, which controls the number of features considered at each split, was set to 0.5. This value ensured that each tree had access to a randomized but diverse subset of features, which helped improve generalization.

Threshold tuning was also applied to the Logistic Regression model. Adjusting the classification threshold allowed for trade-offs between precision and recall. Performance peaked near a threshold of 0.5, where the model maintained a strong balance between false positives and false negatives. Lowering the threshold increased recall at the cost of precision, while higher thresholds resulted in more conservative predictions with improved precision but reduced recall.

The Random Forest Classifier achieved the highest test accuracy at 93.36%, outperforming both logistic and linear regression. Its performance was consistent across matchups, correctly predicting outcomes for players with different playing styles and ranking levels. Key features driving accurate predictions included first serve percentage, break point conversion, and ace-to-double fault ratio. These variables captured player efficiency and consistency, which Random Forest used effectively due to its non-parametric structure. The model handled complex interactions between features and maintained strong generalization by using 500 trees, a depth of 6, and a 0.5 max feature split. These settings balanced predictive accuracy with control over overfitting, allowing the model to deliver robust and reliable predictions.

Discussion

From a practical standpoint, this modeling approach can support scouting decisions, player development strategies, or match preparation. Coaches may use similar data-driven models to evaluate player strengths or identify specific areas for improvement. Additionally, this type of predictive framework has potential applications in sports betting, especially when combined with real-time inputs or recent form data. However, the study is limited by its reliance on aggregated career statistics, which do not reflect short-term fluctuations in player form, injuries, or opponent-specific dynamics. Future work could improve model accuracy by using time-series data, such as recent match results or weekly player form, along with external factors like player fatigue, travel schedules, and surface type. These additions would allow the model to adjust predictions based on real-time conditions rather than only historical averages. With further refinement, these models could be deployed in real-time environments or expanded into live prediction systems during tournaments.

The dataset utilized in this study encompasses comprehensive career-long statistics for ATP players, sourced from reputable databases such as Ultimate Tennis Statistics and the ATP Tour. The dataset in this study includes career-long statistics for ATP players, offering a consistent and comprehensive view of player performance. This long-term data provides strength in evaluating overall skill and consistency across seasons. However, it also limits the model’s ability to account for recent form, injuries, or opponent-specific trends. Key features included serve speed, first-serve percentage, break points converted, and match win percentage. While these features support accurate predictions over time, the lack of real-time updates reduces model responsiveness to short-term factors.

Data preprocessing involved several key steps to ensure data integrity and enhance predictive power. Data consistency was a challenge due to variations in data recording standards over time, necessitating meticulous standardization to ensure consistency. Players with intermittent careers due to retirement or inactivity posed additional challenges in accurately representing their performance metrics. By focusing on aggregated career statistics, this study aims to capture the overall skill level and performance trends of players, providing a robust foundation for predictive modeling.

Conclusion

This study demonstrates that machine learning can effectively predict ATP match outcomes using career-long player performance statistics. Among the three models tested, the Random Forest Classifier consistently outperformed both Logistic and Linear Regression. Its ability to capture non-linear patterns made it particularly well-suited for the complex relationships within tennis performance data, validating the initial hypothesis formed prior to the study. The most influential predictors of success included serve percentage, break point conversion, and point-level consistency. These findings align with broader understandings in professional tennis: players who dominate key moments and maintain a high level of consistency tend to outperform peers across surfaces and matchups. The correlation matrix and model outputs also confirmed the relevance of features such as match win percentage, games won percent, and serve rating, which consistently appeared among the strongest predictors.

Acknowledgments

I would like to thank my Polygence mentor, Akshay, for his expert guidance, encouragement, and support throughout the research process. His insight in computer vision and sports analytics were instrumental in shaping this project.

References

1. “ATP Player Statistics.” *ATP Tour*, www.atptour.com. Accessed 23 Aug. 2025.
2. Boulier, Bryan L., and Howard O. Stekler. “Predicting the Outcomes of Tennis Matches.” *International Journal of Forecasting*, vol. 19, no. 2, 2003, pp. 155–170.
3. Cai, Li, et al. “Applying Machine Learning to Predict Grand Slam Tennis Matches.” *Data Mining in Sports Journal*, vol. 8, no. 1, 2021, pp. 45–61.
4. De Seranno, Andrea. “Predicting ATP Singles Match Winners Using Machine Learning.” *Journal of Sports Analytics*, vol. 5, no. 3, 2019, pp. 211–229.
5. Gao, Jun, and Adam Kowalczyk. “Random Forest Models for Predicting Tennis Match Outcomes.” *Journal of Quantitative Analysis in Sports*, vol. 12, no. 4, 2020, pp. 301–315.
6. McHale, Ian G., and Amanda Morton. “A Bradley–Terry-Type Model for Forecasting Tennis Match Results.” *International Journal of Forecasting*, vol. 27, no. 2, 2011, pp. 619–630.
7. Knottenbelt, William J., et al. “A Common-Opponent Stochastic Model for Predicting Professional Tennis Matches.” *International Journal of Computer Science in Sport*, vol. 9, no. 2, 2010, pp. 67–79.
8. “Ultimate Tennis Statistics Database.” *Ultimate Tennis Statistics*, www.ultimatetennisstatistics.com. Accessed 23 Aug. 2025.