Data-Driven Prediction of Tennis Ranking Movements with Ensemble Machine Learning Models

Zhiyuan Zhang

Received April 21, 2025
Accepted October 12, 2025
Electronic access November 30, 2025

Tennis rankings are central to evaluating player performance, determining tournament seedings, and influencing sponsorship opportunities. Traditional ranking systems like ATP and WTA often fail to capture players evolving dynamics and subtle performance shifts over time. This study introduces a machine learning-based framework with 9 models in total to predict player ranking movement using time series features derived from match statistics and historical performance data. A comprehensive dataset spanning ATP and WTA matches from 2013 to 2024 was compiled using web scraping methods, incorporating 31 features for the ATP dataset and 49 features for the WTA dataset, such as serve efficiency, return effectiveness, overall match statistics, and tournament-level results. Multiple regression modelsincluding tree-based ensembles (e.g., LightGBM, XGBoost), linear models (e.g., Ridge, ElasticNet), and kernel methods (e.g., Kernel Ridge Regression)were evaluated using MSE, RMSE, and MAE. A stacking ensemble with Kernel Ridge Regression as a meta-learner demonstrated the best predictive performance with the lowest error across all metrics. For the result, we also rank the feature importance for the LightGBM model to identify which match-level statistics contribute most significantly to ranking changes, providing data-informed training focuses, tactical planning, and performance analysis for coaches or players. The findings underscore the potential of ensemble machine learning approaches for accurate and data-driven player ranking forecasts in professional tennis.

Keywords: Tennis ranking, Time series, Machine learning, Data preprocessing, Modelling

Introduction

Tennis is one of the most popular racket sports in the world, with millions of recreational players and a huge population of spectators. In 2024, the combined viewership of the four Grand Slams, the major tennis tournaments, reached almost 2 billion people in more than 200 countries[1]. Due to the competitiveness and global influence of tennis, the ranking of each player plays a central role in different areas. They determine tournament entries, seedings, and sponsorship deals while serving as a fundamental measurement of a players performance and ability. Thus, accurate and timely rankings of players are not only important for the players and coaches, but also for the stakeholders, sponsorships, and event organizers in the tennis community. Traditional ranking systems, such as the ATP and WTA point-based systems, only display raw numbers indicating a players performance in different categories and cannot fully capture the players evolving performance trends, especially when ranking is fluctuating all the time[2].

A possible solution is time series models, a technique that can perceive patterns in datasets to forecast future statistics. In the context of ranking in tennis, the primary features of determining whether a player is good or not are the players statistics and match outcomes in different level matches. By leveraging these key features, authorities can easily predict a players ranking movement, evaluating their potential commercially and tennis-wise. Many ML models have been developed for match outcome prediction because of the huge market for game betting[3]. However, an ML model and data set for tennis ranking prediction are relatively scarce due to the limited availability of comprehensive, high-quality player performance data over time. To address this issue and others, such as the scarcity of machine learning models focused on tennis ranking prediction, a data-driven approach that utilizes yearly player performance data is proposed. Our method begins by identifying the useful features and collecting data based on these features from the official ATP and WTA websites, or tennis datasets, like Ultimate Tennis Statistics, using BeautifulSoup for scraping[4]. Then, we preprocessed the data by normalizing the feature Ranking-Movement due to its large variability. At last, machine learning models that include ensemble models, tree-based models, and linear models are applied to predict the ranking movement of players[5]. By modelling player progression by data, our research aims to provide data-based insights to players, coaches, and tournament organizers.

While prior studies have explored predictive modeling in tennis, they have generally relied on limited datasets. They often restricted to single matches, shorter time periods, or specific tournaments. To our knowledge, no existing work has compiled a comprehensive dataset that spans both ATP and WTA players across more than a decade (20132024). This dataset contribution is therefore a novel aspect of our study. It enables broader and more representative modeling of ranking dynamics in male and female professional tennis.

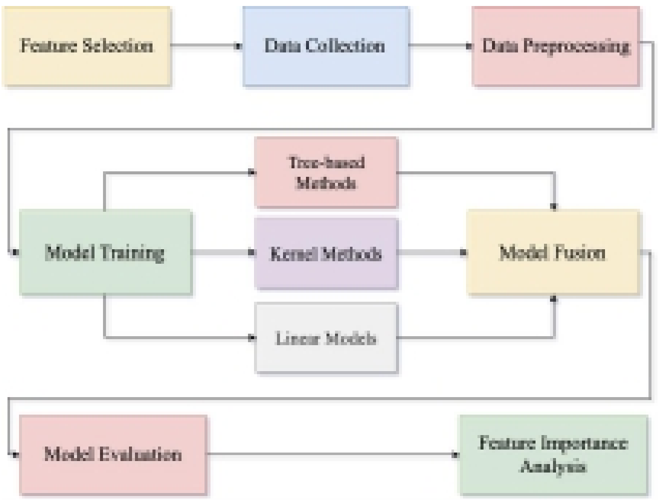

Fig. 1 A Simplified General Process of the Tennis Ranking Prediction Task

Our main contributions are as follows:

1. We compile, to our knowledge, the first comprehensive dataset spanning both ATP and WTA players from 2013 to 2024, incorporating yearly player performance and match information from official and public sources.
2. We design a machine learning pipeline with 9 models that leverages key player statistics to predict ranking movement, incorporating ensemble learning methods and model fusion for improved accuracy.
3. We provide empirical results and feature importance insights that can support players, coaches, and game organizers in understanding ranking dynamics and performance trends.

Literature Review

Player performance forecasting increasingly relies on machine learning modeling, integrating key factors, such as match statistics and historical rankings. Bozdch et al. introduced multinomial logistic regression and neural networks to identify key factors of ATP rankings, such as age, height, and service-related metrics5.Their study analyzed data from 1,990 tennis players during the 2022 season, with a total of 20,040 data points, highlighting the potential of neural networks in predicting ATP rankings. Buhamra et al. uses various regression-based models, including random forests, spline-based approaches, and logistic regression to model and predict outcomes of mens Grand Slam matches[6]. With a dataset of 5013 matches, the study considers diverse features, such as, ATP ranking and points, Elo rating, and betting odds. One notable thing this research introduced is the creation of two new variables, Age.30 and Age.int, which measure the deviation of age from 30 and an optimal interval. These features slightly improved the overall performance of the model. Bunker et al. compare the traditional Elo and WElo rating methods with machine learning models, including ANN, SVM, and Random Forest models, for predicting tennis match outcomes[7]. The study concludes that certain machine learning models, such as the Logistic Regression and ADTree models, outperformed the traditional rating method, matching the accuracy of betting odds. This underscores the potential of machine learning models in the field of tennis match outcome prediction, which is closely related to the prediction of player rankings.

Li introduced a logistic regression model grounded in Bayesian probability to predict point-level outcomes in a tennis match[8]. To consider the psychological factors in a match, the study employs the Analytic Hierarchy Process (AHP), with features, including rally length and first serve success rates, to quantify the momentum of a player. Pham et al. focus on the use of classification models, including XGboost, logistic regression, and random forest models, to forecast outcomes of ATP singles match[9]. The study utilizes a decade-long dataset ranging from 2012 to 2022, with up to 49 explanatory variables constructed of player statistics and match features. The random forest model in this study successfully predicted the winner, Novak Djokovic, of the 2023 Australian Open men's singles tournament, demonstrating its accuracy. Lastly, Rui et al. introduced a supervised machine learning approach, using models such as, K-Nearest Neighbors, XGBoost, and logistic regression models, with XGBoost to predict the flow of points in a tennis match[10].The study utilizes Grey Relational Analysis and fuzzy set theory to rank several important features, including serve status, games won in the current set, ranking difference, distance covered, serve speed, previous victory status, and unforced errors. The study exhibited the potential of machine learning models when it comes to tennis-related forecasting. Despite these advances, existing studies focus more on match outcome prediction rather than ranking dynamics, and often rely on datasets with limited time spans, or narrow feature sets. Few works incorporate both ATP and WTA players in a unified framework, while exploring the ranking dynamic of professional tennis in a comprehensive way.

Methods

Model Selection

This research evaluates the performance of several machine learning models, including LightGBM (LGB), XGBoost (XGB), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Extra Trees Regressor (ETR), Kernel Ridge Regression (KR), Ridge Regression (Ridge), ElasticNet (EN), and Bayesian Ridge Regression (BR). These models can be classified into five groups based on their underlying methodologies: Tree-Based Ensemble Methods (Random Forest Regressor, Extra Trees Regressor), Boosting Methods (LightGBM, XGBoost, Gradient Boosting Regressor), Bagging Methods (Random Forest Regressor, Extra Trees Regressor), Linear Models (Ridge Regression, ElasticNet, Bayesian Ridge Regression), and Kernel Methods (Kernel Ridge Regression). Tree-Based Ensemble Methods are machine learning models that aggregate the predictions of multiple decision trees to enhance performance and reduce overfitting. This research uses these diverse methods in a stacking framework, applying a 5-fold cross-validation repeated twice for better evaluation.

1. **Tree-based ensemble methods** are machine learning models that aggregate the predictions of multiple decision trees to enhance performance and reduce overfitting. The Random Forest Regressor fits a large number of decision trees on various subsets of data and averages their output to improve accuracy and control overfitting [11]. In comparison, Extra Trees Regressor introduces more randomness by selecting splits randomly during tree construction, which may be better at avoiding [12]. These models capture non-linear relationships between player statistics, such as the interaction of first serve percentage and return game won percentage, which strongly influence ranking changes. Their averaging mechanisms reduce variance and make them robust to noise in yearly performance data.

2. **Boosting methods** sequentially refine predictions by correcting prior errors, which is particularly valuable for ranking prediction where small errors can accumulate into significant ranking shifts. Gradient Boosting Regressor builds a predictive model in the form of an ensemble of weak predictive models and optimizes a differentiable loss function over function space [13]. XGBoost improves gradient boosting with second-order derivatives (Hessian) and parallel computations for faster and more accurate optimizations [14]. Lastly, LightGBM uses a leaf-wise growth strategy rather than the classic level-wise methods and histogram-based techniques for a lower memory usage, offering greater efficiency for large datasets [15].

3. **Bagging methods** focus on training multiple base models independently on different random subsets of the original dataset. Random Forest and Extra Trees are notable examples that apply bootstrapping (sampling with replacement) to create various models whose outputs are averaged to improve generalization and reduce overfitting [16].

4. **Linear models** offer interpretability and serve as strong baselines for understanding how individual features, such as aces or break points converted, contribute to ranking changes. Ridge regression is a linear regression model that uses L2 regularization to shrink coefficients and reduce overfitting [17].ElasticNet is a linear model that incorporates both L1 (lasso) and L2 (ridge) regularization, which offers both coefficient shrinkage and feature selection Lastly, Bayesian Ridge Regression is a type of conditional modeling in which coefficients are treated as random variables, with the goal of estimating their distributions to capture uncertainty in predictions [17].

5. **Kernel methods** are a class of algorithms of pattern analysis, which utilizes kernal function to map input data into higher-dimensional spaces, allowing the linear models to capture non-linear relationships. Kernel Ridge Regression combines Ridge regression (linear model) with kernel methods, which gives the model more flexibility when fitting and less overfitting with the help of the L2 penalty [17].

Dataset and Splits

The dataset includes ATP and WTA player yearly performance data from 2013 to 2024, sourced from the official ATP and WTA Tour website and Ultimate Tennis Statistics, which is an online ATP database made by Cekovic [18].The ATP dataset includes up to 31 explanatory variables, ranging from match statistics such as first serve percentage and aces to historical features like match records and player rankings.

Serve performance, return performance, overall match performance, historical match records, and ranking information are the most substantial features in determining a players overall performance. Thus, it is important to capitalize on these features when training an ML model to predict the ranking of players.

Precise ranking predictions deeply rely on the consideration of the quality and effectiveness of players serve. Key variables include Aces, DoubleFaults, 1stServe percentage, and points won on both first and second serves. Additional features such as BreakPointsFaced, BreakPointsSaved, ServiceGamesPlayed, ServiceGamesWon, and TotalServicePointsWon provide an overall view of a players serve consistency and ability to maintain service under pressure. All of the above features either reflect the direct effectiveness of serve, such as the number of aces, which is a successful serve that the opposing player does not touch, or indirectly reflect the effectiveness of a serve, for example, service games won. Similar to the serve, the return also plays a crucial role in a tennis match. It determines how effectively a player can handle pressure from the opponents serve and service games, which indirectly indicates the opportunity a player has to break the opponents serve. Variables such as 1stServeReturnPointsWon, 2ndServeReturnPointsWon, BreakPointsOpportunities, and BreakPointsConverted measure return efficiency and conversion rates. More directly, ReturnGamesPlayed, ReturnGamesWon, and ReturnPointsWon give insight into return dominance and consistency.

Certain players may have a high percentage in the above features due to the lower-level tournament they have participated in. Thus, it is important to consider the level of matches they are playing in and how they are performing. These include counts of matches won and lost in Grand Slams, Masters 1000, ATP 500, and ATP 250 events. This also indirectly reflects how a player handles pressure and different levels of opponents. Lastly, since we are predicting the ranking of players, it is crucial to include their ranking in the previous year and the ranking movement. This is the most direct reflection of how the player has performed in the entire year. The dataset is then separated into a training dataset and a test dataset with a ratio of 4:1.

Data Preprocessing

Due to the larger variation of the feature Rankingmovement in the WTA dataset, every value in the feature Rankingmovement is scaled in the range of [-80,80], meaning that any value that is greater than 80 will be scaled down to 80, and vice versa. The data points of the feature ranking movement is standardized according to the formula:

$$z = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the original data point, x_{min} is the minimum value in the dataset, and x_{max} is the maximum value in the dataset. This standardization scales all the data in [0,1] and ensures that the feature contributes appropriately to model training without being impacted by the scale of other variables, especially when the ranking movement has a wide range of values, with a maximum absolute value of up to 100[19].

Hyperparameter Tuning

To optimize model performance, we applied Exhaustive Grid Search for each model. This method systematically explores all possible combinations within the generated grid to identify the optimal configuration that yields the best cross-validation performance. Each combination is evaluated using 5 fold-cross validation, and the optimal set of parameters was selected based on the lowest validation MSE. This exhaustive grid search allowed the model to perform at its best.

Fusion Model

In this pipeline, each of the nine base models (tree-based ensembles, boosting methods, linear regressors, and kernel methods) was first trained on the training dataset using 5-fold cross-validation repeated twice. For each fold, the base models generated out-of-fold predictions that were not used during their own training. These predictions were then fused to form a collective output of the base models. The choice of KRR as the meta-learner is motivated by its ability to capture nonlinear interactions among the base model outputs. KRR also applies L2 regularization to prevent overfitting. Unlike other meta-learners, KRR emphasizes stronger base learners in the pipeline while diminishing the influence of weaker ones. A fusion model efficiently combines the advantages of different models, producing a superior result.

Results

We take Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) into account when analyzing the performance of the different models.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y}_i|$$

MSE is a cost function that measures the average of the squares of the errors. The function emphasizes larger errors, which may be helpful for certain data, like financial forecasting, where large errors are more detrimental. However, it may perform worse with the occurrence of outliers, so it could be better to consider the MAE. MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. Similar to MSE, RMSE is a cost function that can be represented as the square root of the mean square error. Looking at the square root of MSE can be advantageous because of the similar scaling of RMSE and the data, and the lesser sensitivity towards outliers. The models trained on the ATP dataset performed significantly better compared to the models trained on the WTA dataset across all metrics (MSE, RMSE, and MAE). The largest MAE increase between WTA and ATP models scales up to +0.1205 and +0.1204 in LightGBM and Random Forest Regressor. This indicates that the WTA dataset is much noisier and more heterogeneous compared to the ATP dataset, meaning it has a less predictive structure. In reality, in the feature RankingMovement, there is a much wider range, with an absolute value of up to 400, in the WTA dataset as in other features compared to the ATP dataset. The large difference in errors may suggest that male players have a more consistent performance and fewer performance swings when compared to female players, as well as less variability in overall outcomes. Due to the diversity of models used in the process, accompanied by the differences, a fusion model, which combines the

Table 1 A List of All the Features in ATP Dataset with Explanations

Feature Name	Description
PlayerID	Unique identifier for each player.
Aces	Number of service aces made by the player.
DoubleFaults	Number of double faults committed by the player.
1stServe	Percentage of first serves successfully made.
1stServePointsWon	Percentage of points won on first serve.
2ndServePointsWon	Percentage of points won on second serve.
BreakPointsFaced	Total number of break points faced on serve.
BreakPointsSaved	Number of break points saved on serve.
ServiceGamesPlayed	Total number of service games played.
ServiceGamesWon	Number of service games won.
TotalServicePointsWon	Total percentage of service points won.
1stServeReturnPointsWon	Percentage of opponent's first serve points won.
2ndServeReturnPointsWon	Percentage of opponent's second serve points won.
BreakPointsOpportunities	Total break point opportunities against opponents.
BreakPointsConverted	Number of break points successfully converted.
ReturnGamesPlayed	Number of return games played.
ReturnGamesWon	Number of return games won.
ReturnPointsWon	Percentage of points won while returning serve.
TotalPointsWon	Percentage of total points won (serve + return).
GrandSlamWon	Number of Grand Slam matches won.
GrandSlamLost	Number of Grand Slam matches lost.
MastersWon	Number of ATP Masters 1000 matches won.
MastersLost	Number of ATP Masters 1000 matches lost.
ATP500Won	Number of ATP 500 matches won.
ATP500Lost	Number of ATP 500 matches lost.
ATP250Won	Number of ATP 250 matches won.
ATP250Lost	Number of ATP 250 matches lost.
CurRanking	Current ATP ranking.
PrevRanking	Previous ATP ranking.
RankingMovement	CurRanking - PrevRanking.

predictions of individual base models using Kernel Ridge Regression (KRR), is added. In both the ATP and WTA based models, the fusion model is seen to have the best performance over all metrics. This is likely because KRR captures nonlinear relationships between the base model outputs and the true target. This allows it to learn how to optimally weight and adjust their contributions, leading to better generalization on both ATP and WTA datasets. Moreover, the gap of +0.0708 between ATP and WTA models further suggests that the ATP dataset is more predictive. Among the models tested from both the ATP and WTA datasets, tree-based models and ensemble methods such as LightGBM, Random Forest, and Gradient Boosting are observed to have a better overall performance compared to other linear models. Across the linear models, Elastic Net performs the worst, which may be due to the lesser flexibility of the model. This suggests that the relationship between ranking outcomes and the features used to measure players performance have a nonlinear relationship and involves complex interactions, which is better captured by tree-based models and ensemble methods. In comparison, linear models are more limited in modelling such patterns despite being more interpretable.

We take Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) into account when analyzing the performance of the different models.

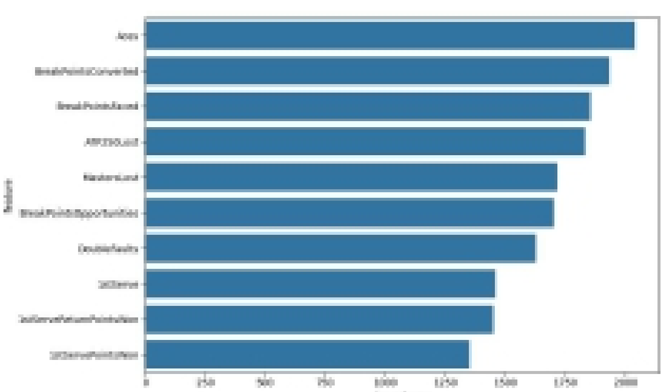

Fig. 2 A Chart Showing the Top 10 Most Important Features for ATP Ranking Prediction Based on LightGBMs Feature Importance Scores.

Table 2 A List of All the Features in WTA Dataset with Explanations

Feature Name	Description
PlayerNbr	Unique identifier for each player.
Aces	Number of service aces made by the player.
DoubleFaults	Number of double faults committed by the player.
FirstServesWon	Number of points won on first serve.
FirstServesPlayed	Total number of first serves attempted.
SecondServesWon	Number of points won on second serve.
SecondServesPlayed	Total number of second serves attempted.
BreakPointsFaced	Total number of break points faced on serve.
BreakPointsLost	Number of break points lost on serve.
ServiceGamesPlayed	Total number of service games played.
ReturnGamesPlayed	Total number of return games played.
BreakPointChances	Total number of break point opportunities against opponents.
BreakPointsConverted	Number of break points successfully converted.
FirstServeReturnChances	Number of opponent's first serves returned.
FirstReturnWon	Number of points won on opponent's first serve.
SecondReturnChances	Number of opponent's second serves returned.
SecondReturnWon	Number of points won on opponent's second serve.
FirstServeWonPercent	Percentage of points won on first serve.
SecondServeWonPercent	Percentage of points won on second serve.
FirstReturnPercent	Percentage of points won on opponent's first serve.
SecondReturnPercent	Percentage of points won on opponent's second serve.
BreakpointConvertedPercent	Percentage of break points converted.
FirstServePercent	Percentage of first serves successfully made.
ReturnGamesWonPercent	Percentage of return games won.
BreakpointSavedPercent	Percentage of break points saved.
ServiceGamesWonPercent	Percentage of service games won.
ServicePointsWonPercent	Percentage of total service points won.
ReturnPointsWonPercent	Percentage of total return points won.
TotalPointsWonPercent	Percentage of all points won (serve + return).
MatchCount	Number of matches played.
AcesPerMatch	Average number of aces per match.
DoubleFaultsPerMatch	Average number of double faults per match.
FirstServesWonPerMatch	Average number of first serve points won per match.
FirstServesPlayedPerMatch	Average number of first serves attempted per match.
SecondServesWonPerMatch	Average number of second serve points won per match.
SecondServesPlayedPerMatch	Average number of second serves attempted per match.
BreakPointsFacedPerMatch	Average number of break points faced per match.
BreakPointsLostPerMatch	Average number of break points lost per match.
ServiceGamesPlayedPerMatch	Average number of service games played per match.
ReturnGamesPlayedPerMatch	Average number of return games played per match.
BreakPointChancesPerMatch	Average number of break point opportunities per match.
BreakPointsConvertedPerMatch	Average number of break points converted per match.
FirstServeReturnChancesPerMatch	Average number of opponent's first serves returned per match.
FirstReturnWonPerMatch	Average number of first serve return points won per match.
SecondReturnChancesPerMatch	Average number of opponent's second serves returned per match.
SecondReturnWonPerMatch	Average number of second serve return points won per match.
CurSinglesRanking	Current WTA singles ranking.
PrevSinglesRanking	Previous WTA singles ranking.
RankingMovement	CurSinglesRanking - PrevSinglesRanking.

Table 3 An Overview of All the Hyperparameter Values in the Model Pipeline for Both ATP and WTA Dataset.

Dataset Model	ATP	WTA
LightGBM	num_leaves(4) min_data_in_leaf(20) objective(regression) max_depth(-1) learning_rate(0.003) boosting(gbdt) feature_fraction(0.4) bagging_freq(1) bagging_fraction(0.99) bagging_seed(14) metric(mse) lambda_l1(1.5) lambda_l2(1.5) verbosity(1)	num_leaves(4) min_data_in_leaf(20) objective(regression) max_depth(-1) learning_rate(0.003) boosting(gbdt) feature_fraction(0.6) bagging_freq(1) bagging_fraction(0.8) bagging_seed(14) metric(mse) lambda_l1(1.5) lambda_l2(0) verbosity(1)
XGBoost	eta(0.02) max_depth(6) min_child_weight(3) gamma(0) subsample(0.7) colsample_bytree(0.3) lambda(2) objective(reg:linear) eval_metric(mse) silent(True) nthread(-1)	eta(0.01) max_depth(8) min_child_weight(5) gamma(0) subsample(0.8) colsample_bytree(0.6) lambda(2) objective(reg:linear) eval_metric(mse) silent(True) nthread(-1)
RF	n_estimators(1600) max_depth(9) min_samples_leaf(9) min_weight_fraction_leaf(0.0) max_features(0.25) verbose(1) n_jobs(-1)	n_estimators(1600) max_depth(9) min_samples_leaf(5) min_weight_fraction_leaf(0.0) max_features(0.25) verbose(1) n_jobs(-1)
GBoost	n_estimators(400) min_learning_rate(0.01) subsample(0.65) max_depth(7) min_samples_leaf(20) max_features(0.22) verbose(1)	n_estimators(200) min_learning_rate(0.01) subsample(0.8) max_depth(5) min_samples_leaf(20) max_features(0.3) verbose(1)
ExtraTrees	n_estimators(1000) max_depth(9) min_samples_leaf(12) min_weight_fraction_leaf(0.0) max_features(0.4) verbose(1) n_jobs(-1)	n_estimators(500) max_depth(11) min_samples_leaf(12) min_weight_fraction_leaf(0.0) max_features(0.5) verbose(1) n_jobs(-1)
KRR	alpha(1.0) kernel(linear) gamma(None) degree(3) coef0(1) kernel_params(None)	alpha(1.0) kernel(linear) gamma(None) degree(3) coef0(1) kernel_params(None)
Ridge	alpha(1200)	alpha(1200)
ElasticNet	alpha(1.0) default l1_ratio(0.06)	alpha(1.0) default l1_ratio(0.06)
BayesRidge	default	default

Table 4 Model Performance on Men's Tennis (ATP) Dataset

Model	MSE	RMSE	MAE
LightGBM	0.04438	0.21067	0.16519
XGboost	0.04612	0.21476	0.16785
RandomForest	0.04370	0.20905	0.16392
GradientBoosting	0.04370	0.20904	0.16421
ExtraTrees	0.04400	0.20975	0.16459
Kernel Ridge	0.04410	0.20999	0.16623
Ridge	0.04389	0.20949	0.16591
ElasticNet	0.04688	0.21652	0.17132
BayesianRidge	0.04394	0.20962	0.16599
Fusion	0.04326	0.20799	0.16318

Table 5 Model Performance on Women's Tennis (WTA) Dataset

Model	MSE	RMSE	MAE
LightGBM	0.11385	0.33742	0.28568
XGboost	0.11825	0.34388	0.28580
RandomForest	0.11437	0.33818	0.28430
GradientBoosting	0.11406	0.33772	0.28461
ExtraTrees	0.11377	0.33729	0.28456
Kernel Ridge	0.11550	0.33986	0.28444
Ridge	0.11527	0.33952	0.28531
ElasticNet	0.11472	0.33871	0.28603
BayesianRidge	0.11457	0.33848	0.28565
Fusion	0.11406	0.33773	0.28453

For the explanation about the models, we use LightGBM for feature important ranking, as shown in Fig. 3. For the ATP player, serves plays a crucial role in a tennis match, and a player with a successful serve can win large percentages of their service games. One great example is John Isner, who has a career-highest ranking at number 8 and had 1260 aces in 2015[20]. Aces are a great indicator of players serving performance, and a great serve reduces pressure while saving more games. It is evident that most of the top players have a strong and accurate serve. This suggests that coaches should prioritize training sessions focused on serve consistency and accuracy under pressure, as improvements here are most likely to translate into ranking gains. The second important feature, BreakPointsConverted, measure how effectively a player deals with the service games of their opponent, their ability to handle critical points, and most of all, how effectively a player turns break point opportunities into actual breaks which, in many cases, a single break wins the player an entire round. On the opposite, the third important feature, BreakPointssaved, indicates dominance on serve, which is a common trait for high-ranking players. The next two traits are direct measurements of a players performance in large tournamentsATP 250 and Masters (ATP 1000). Frequent early losses in lower-tier (ATP 250) or high-tier (Masters) events can prevent ranking progression. It is reasonable for the model to rank ATP250 before the Masters tournament despite that the Masters being more prosperous and competitive than the ATP250 tournament, because only the top 50 players are the major participants in Masters events, while we took the top 300 players for both the ATP and WTA datasets.

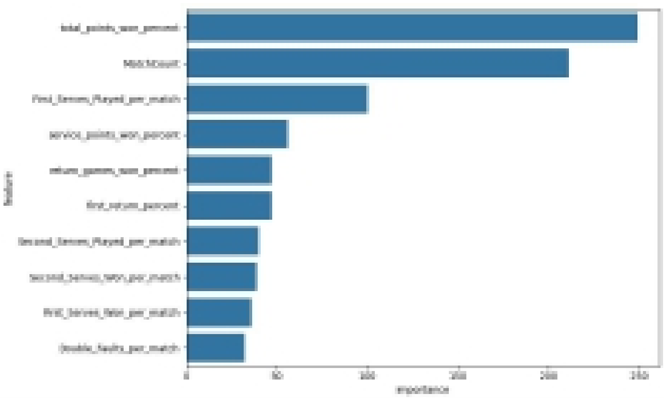

Fig. 3 A Chart Showing the Top 10 Most Important Features for WTA Ranking Prediction Based on LightGBMs Feature Importance Scores.

For the WTA players, Total_Points_Won_Percent plays the most important role in determining ranking movement. This is reasonable because the higher the percentage a player wins in total, the more dominance a player is showing in a match. The second most important feature, MatchCount, measures the number of matches a player has participated in in a year. A greater value for MatchCount indicates better performance stability and a higher ranking in general for those who progress more across all tournaments. This may also be caused by the nature of LightGBM, which often favors numerical figures with a broader range. However, based on the models ranking, players should prioritize playing more matches in different level to gain more experience.

After the two more dominant features in the WTA dataset, the third-ranked figure is rather similar to the ATP feature ranking, showing a feature related to serving. The number of first serves played per match directly indicates the percentage a player has to hold a serve or serve an ace. It can be a deciding factor in winning games. The fourth-ranked feature, Service Points Won Percent, quantifies a player's overall effectiveness when serving by calculating the percentage of total service points won. Unlike features like aces or first serves won, this feature captures the complete view across all situations. A high value in this feature typically reflects strong serving consistency and the ability to hold serves. Both the third and fourth features indicate that serves play a crucial role in WTA tennis. The fifth feature, Return_Games_Won_percent, measures the ability of a player to break serves of the opponent, and in many cases, the player with a single break in a set takes the set. Overall, most of the features rankings are very reasonable and explainable based on tennis knowledge, which indicates that the model is making decisions based on meaningful and interpretable signals, not noise.

Conclusion

This study demonstrates the feasibility and effectiveness of using machine learning models to predict tennis player ranking movements based on match statistics and historical performance data. Unlike the ATP/WTA point systems, which only reflect past results and provide static standings, our framework offers forward-looking insights into ranking movements. By applying and comparing a diverse set of modelsranging from boosting techniques to linear regressionswe found that tree-based ensemble methods consistently outperform others, reflecting the nonlinear and complex nature of performance factors in tennis. Furthermore, the fusion model using kernel ridge regression yielded the most accurate predictions by capturing nonlinear relationships across model outputs. The performance gap between ATP and WTA datasets reveals a more stable and predictable structure in mens tennis rankings, highlighting differences in performance variance and data quality. These insights have practical implications for stakeholders across the tennis ecosystem, from analysts and coaches to media and sponsors. Future work could expand to include real-time data, psychological metrics, and player health indicators to further enhance model robustness and predictive accuracy. These enhancements would allow the model to generate more precise, context-aware ranking forecasts, while including certain aspects that is difficult to incorporate for our current model. Beyond methodological gains, such predictions can more comprehensively support players and coaches in training focus, tournament scheduling, and workload management.

References

1 Roland-Garros, *Worldwide audiences for Grand Slam tennis reach new records*, https://www.rolandgarros.com/en-us/article/worldwide-audiences-grand-slam-tennis-record-onsite-tv-social-media.

2 R. Mayew, *Which global tennis rating better measures player skill? Evidence from the 2022 USTA Junior National Championships*, https://thesportjournal.org/article/which-global-tennis-rating-better-measures-player-skill-evidence-from-the-2022-usta-junior-national-championships.

3 I. McHale and A. Morton, *Journal of Mathematical Psychology*.

4 A. Kamra.

5 M. Bozdch, D. Puda and P. Grasgruber, *PLoS ONE*, **19**, e0310246.

6 N. Buhamra, A. Groll and S. Brunner, *J Sports Analytics*, **10**, 1733.

7 R. Bunker, C. Yeung, T. Susnjak, C. Espie and K. Fujii, *Journal of Sports Analytics*.

8 C. Li, *Research on dynamic analysis and prediction model of tennis match based on Bayesian probability and analytic hierarchy process*, arXiv preprint arXiv:2407.07116.

9 C. Pham and K. Bufi, *Master's Essay*, Department of Economics, Lund University, Lund, Sweden, 2023.

10 B. Rui, C. Hing, L. Haoyuan and T. Yew, Proceedings of 2024 International Conference on Machine Learning and Intelligent Computing, PMLR, p. 393403.

11 J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Burlington, MA, 3rd edn, 2011.

12 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, 2nd edn, 2009.

13 C. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.

14 T. Chen and C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 785794.

15 G. Ke, *Advances in Neural Information Processing Systems*, **30**, 31463154.

16 L. Breiman, *Machine Learning*, **24**, 123140.

17 F. Pedregosa, *Journal of Machine Learning Research*, **12**, 28252830.

18 *Ultimate Tennis Statistics. Ultimate men's tennis statistics destination for die-hard fans*, `https://www.ultimatetennisstatistics.com`.

19 L. B. Amorim, G. D. Cavalcanti and R. M. Cruz, *Patterns*, **3**, 100401.

20 *John Isner Player stats*, `https://www.atptour.com/en/players/john-isner/i186/player-stats?year=2015&surface=all`.