Predicting Tennis Matches and Tournaments

Authors: Nathan Dennis, John Breedis, Yiming Chen

Abstract

This paper presents an ELO based approach across tennis surfaces to predict the outcome of tennis matches and tournaments. Our ELO model adjusts the K scaling dynamically for tournament level and year, deriving match probabilities from individual set outcomes in best of three or five sets. A normal distribution is applied to adjust winning probabilities based on players' ages, recognizing that peak performance occurs around 25 years old.[2] This model is applied to predict three Grand Slam tournaments in 2023, comparing our probabilities to betting odds. Results suggest the ELO approach accuracy can vary depending on surface.

1 Introduction

Predicting outcomes in sports like tennis presents an interesting challenge due to the many influencing factors. Elements such as player statistics and rankings to recent performance and surface dominance, every element contributes to the chances of winning a match.[5] Additionally, factors such as health and age play a major influence on who wins a tennis championship each year.

Every year, millions watch as athletes compete against each other to win games and championships. Tennis is no different, with major tournaments year round there is no shortage of excitement around tournament outcomes. The 4 Grand Slams, Roland Garros, Wimbledon, US Open, and Australian Open, are the most prestigious tournaments in tennis, with the Association of Tennis Professionals (ATP) rewarding the most ranking points to players who win these tournaments.[1] However, the question always pops up in people's minds, who is going to win?

In this paper, we hope to not only answer this question using an ELO based approach, but also provide insights as to how different surfaces in tennis may affect match outcomes. We will develop a model to predict who will win tennis tournaments, with a focus on evaluating and validating our results using Grand Slam events and betting odds.

2 Description of Model

To develop a methodology to predict tennis matches and rank players, we decided to use an ELO score metric. The ELO rating system effectively assigns ratings to players and adjusts player ratings over time given in formula 2. We use ELO scores to predict the probability a player will win a given match based on formula 1, the $R_{i/j}$'s denote the ELO ratings for players i and j and scale factor of 800.

$$p_{i,j} = \frac{1}{1 + 10^{(R_i - R_j)/800}} \quad (1) \qquad \begin{aligned} R_i' &= R_i + K(1 - p_{i,j}) \quad \text{i wins} \\ R_i' &= R_i + K(0 - p_{i,j}) \quad \text{i loses} \end{aligned} \quad (2)$$

Utilizing past singles tennis match data between the years 2014 and 2022, we calculate players ELO rating across surfaces. Before we began development of this model, we conducted research into what factors influenced tennis match outcomes and rankings. The first noteworthy discovery we found was how the ATP ranked tennis players, which is widely regarded as the most well respected ranking system in professional tennis. We discovered that in their points breakdown for all tournaments, certain tournaments reward more points for winning than others. The 4 Grand Slams rewarded a total of 2000 points for winning the finals. The ATP Masters 1000 Tournaments rewards 1000 points for the finals winner. Other smaller tournaments reward lesser values of 500 or lower.[1] Based on this analysis, we decided to adjust the "K" factor based on the tournament level when calculating ELO scores. The K factor is a variable in the ELO rating system that controls how quickly and impactful players' ratings change based on a game, a higher K meaning more impactful. We multiply the K factor by 4 for Grand Slams, 2 for masters/ATP tournaments, don't adjust K for all other tournaments and halve the K factor for the Davis Cup as those matches have low stakes. This mimics the respected ATP ranking system, the higher level tournaments contribute to more rapid changes in ratings, similar to our ELO score system.[1]

We further adjusted the K factor based on the given year the match was played in. Matches played further in the future would impact the current day rankings less since many of those players have retired or changed in skill level. Matches closer to the current day, especially within the past year, would have by far the highest influence in present day rankings. We adjusted the K scaling again based on this, with a stepwise function weighing years further than the present day lesser. Additionally, we slightly adjusted the ELO ratings for surfaces not played on for a specific match by a K factor of 0.5, since players skill levels can translate and improve across surfaces allowing for a more holistic approach.

Using these ELO scores derived from the past tennis match data, we then simulated tennis matches. We calculated the expected game score based on the ratings in the ELO system to determine that a given player will beat another based solely on ELO ratings. Since tennis is determined based on sets rather than a single game, we compute the probability of winning based on the probability a player will win a given number of sets in a match, which are usually best out of 3 or 5 sets. We adjusted the players' winning percentage determined by their age and length of the set. Research indicates that tennis players peak age is around 25 years old, with a relatively normal distribution of winning around the peak of 25.[2] We decided to use this information to adjust the probability of a player winning as the number of sets increases, decreasing the probability for players winning as the match goes on to mirror this normal distribution. We applied a normal distribution with standard deviation 25, closely replicating the distribution presented in the research when the number of sets increased to 5. Specifically, we multiplied each player's winning probability for every set by the density of the normal distribution corresponding to their age raised to the power of the set number. The adjusted winning probability in the n-th set for a given player is modeled by:

$$P_{\text{Win Adjusted, n}} = P_{Win} \cdot f(a)^n \tag{3}$$

Where $f(a)$ represents the density of the normal distribution at age a, n the set number, P_{Win} the initial winning probability, $P_{\text{Win Adjusted},n}$ the adjusted winning probability in set n. This approach accounts for the increased fatigue that older players may experience in longer matches, and recognizes that younger players may lack experience in extended play, facing challenges as the match progresses. To compute the probability a player i beats player j in a set, we use the formula given by:

$$\frac{P_{\text{Win Adjusted, n, i}}}{P_{\text{Win Adjusted, n, i}} + P_{\text{Win Adjusted, n, j}}} \tag{4}$$

Where the numerator is the adjusted winning probability on the n-th set for player i, then the denominator adding the adjusted winning probabilities in the n-th set for player i and j. We simulate each set and game using this formula to predict tennis matches, plugging in tournament brackets to predict entire tournaments.

3 Analysis

Using the methods described above, we simulated tournaments using the ELO score calculation and age metrics. We simulated the 2023 Australian Open, Wimbledon, and Roland Garros, 3 of the 4 Grand Slams, all of which were played on different surfaces. We ran this simulation 5000 times to compute probabilities of each player winning the tournament. The Australian Open is the first Grand Slam played in a year, which is on Hard court and played in January. The Roland Garros is typically played between late May and early June, Wimbledon in June to July. We create these plots comparing our models' predicted probabilities against the betting odds, the top 10 players based on the betting odds.

The first tournament we simulated results for was the Australian Open, which is played on the Hard surface. In *Figure 1*, we see a bar plot comparing our predicting probabilities for a given player winning the tournament compared to the probability the odds give the player of winning. As we can see, the results from our predictive model are relatively accurate outside of Rune and Aliassime. Our model gave them much higher probabilities to win the tournament compared to the betting odds. For the rest of the players the model does a fairly good job in predicting whether they will become the champion. The model accurately gives Novak Djokovic by far the highest odds to win.

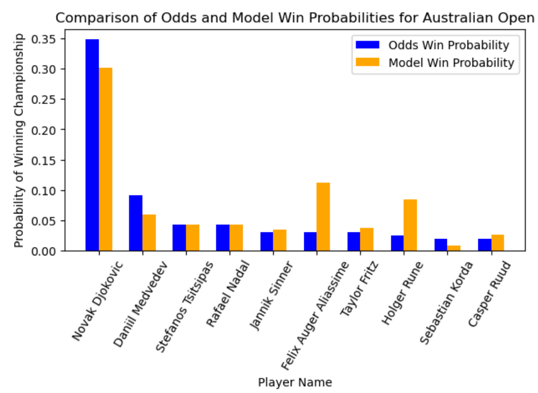

Figure 1: Australian Open

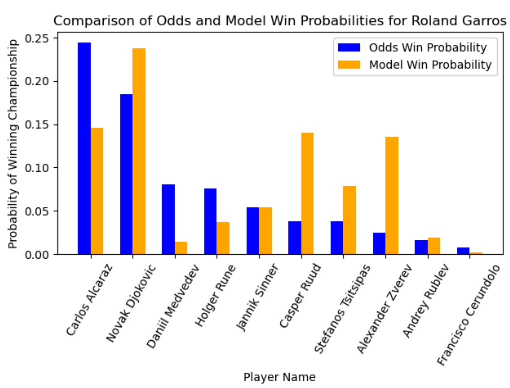

Figure 2: Roland Garros

In *Figure 2*, we now can see a bar plot between our predicted probabilities and the betting odds for the Roland Garros. Our predicted probabilities of winning the tournament are much different from the betting odds, as we can see in the plot many players are being either overestimated or underestimated based on our probabilities. This analysis suggests that the Clay surface may be more difficult to predict. Clay is known as an easier surface to play on due to its rather soft surface and slower style of play, which may contribute to its unpredictability.

The final tournament we predicted was the Wimbledon tournament. This tournament is played on grass and is widely considered to be the most prestigious tournament in tennis.[3] In F*igure 3* our models predicted champions against the bettings odds. As we can see, our model's probabilities are relatively similar to the betting odds, with some minor discrepancies. It seems that both approaches agree Novak Djokovic has significantly better odds of winning than anyone else, with second highest odds Carlos Alcaraz more than 2 times below.

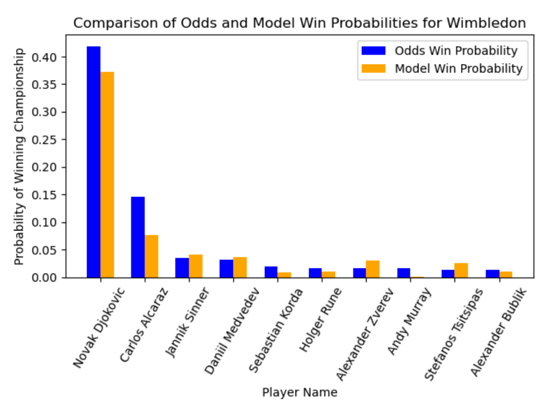

Figure 3: Wimbledon

3.1 Overall Tournament Prediction Analysis

Using the results from the first 3 figures, we can compare how accurate our predicted winning probabilities are compared to the betting odds. For every tournament our model gave Novak Djokovic the highest probability to win, but the odds makers disagreed for the Roland Garros giving Carlos Alcaraz a higher chance. This could be because our model doesn't account for the matches happening throughout the year, as Alcaraz was playing exceptionally well on Clay by the time the Roland Garros was played in May, which our model does not account for. It seems like our model is able to account for some variation in the different surfaces. Across our model probabilities and the betting odds, nearly every player has varying winning probabilities based on the different surfaces. The question regarding if surfaces make a difference in predicting tournament champions is most likely true based on this analysis, as different players are given higher odds to win on some surfaces than others for both the betting odds and our predicted probabilities.

3.2 Parameter Analysis

We used several different parameter combinations for our analysis. We tested using different K scaling factors to compute ELO scores, in particular the scaling factor for the tournament level and year. Scaling the tournament levels higher than we intended ended up harming analysis, as it gave those who have performed well in past high level tournaments significantly higher probabilities of winning again, which is not always the case in tennis. In *Figure 2*, the predictions for the Roland Garros were the most random due to the fact that in the 2022 tournament, Casper Ruud made the finals and Alexander Zverev the semi-finals. This makes our ELO calculation heavily increase their clay surface ELO, hence predicting they will be successful again in 2023. Testing different parameter combinations, such as weighting the past 3 years equally rather than only the past year, heavily overestimated players such as Novak Djokvic, who have had success in many tournaments recently. This made this analysis difficult, as the decision to choose the K factor to scale the ELO calculation based on year occasionally made the model biased to players who performed well further in the past.

3.3 Betting Odds Difference

We calculated several statistics, comparing our results to the betting odds. We used 3 error metrics, including root mean squared error (RMSE), maximal element difference to calculate the max difference between predictions (L_∞), and the average of absolute differences (L_1). We treat the actual value as the betting odds probabilities and the predicted value as our models predictions. We observe a table of the calculated statistics for each of the 3 tournaments we made predictions on:

Tournament	RMSE	L_∞	L_1
Australian Open	0.00983	0.06162	0.00403
Roland Garros	0.01979	0.10986	0.00609
Wimbledon	0.01153	0.07680	0.00433

Table 1: Error Metrics for all 3 Tournaments

In *Table 1*, the 3 error metrics used are calculated for the models predictions compared to the betting odds across all three tournaments. One notable observation is that Roland Garros has the highest error values across all metrics, further showing how inconsistent predictions on the Clay surface are. Our model demonstrated the best performance at the Australian Open across all metrics, indicating our predictions were much closer to the betting odds. In particular the average of absolute difference was 0.00403, which suggests our probabilities deviated from the true probabilities by about 0.403%. This may suggest that our model provides a reliable assessment of player performance on hard courts. It is noteworthy that the Australian Open is the first tournament played out of these 3 in January, since our model incorporated data until December 2022 it would be most accurate predicting tournaments in the beginning of 2023. However, the model's performance on the Wimbledon tournament was relatively intermediate, a tournament played in the summer on grass, not as accurate as the Australian Open but far more accurate than the Roland Garros.

The wide variation of difference in error metrics highlights the sensitivity of our model to surface types and suggests the model may need improvements to account for specific conditions. We tested different parameter combinations to reduce errors across all tournaments and surfaces, including changing the age decay factor since clay is seen as easier on players, but none seemed to fix the poor predictions on clay. Overall, there is a clear effect of surfaces on match outcomes as indicated by the alignment of our model with betting odds on the hard and grass surfaces, but not clay.

4 Discussion

Through our analysis of developing an ELO-based scoring system separated by surfaces to predict tennis matches and tournaments, we discovered several key insights. One significant finding with regards to the ELO calculation itself was how effective adjusting the K factor was to formulate our predictions. We effectively modeled how higher stakes tournaments such as the Grand Slams in tennis have a greater impact on player ratings, where those who are successful in those tournaments are rated more highly. This makes sense as it clones the idea used in the ATP rankings as well, where certain higher level tournaments are weighted more significantly in their rankings. We learned that the K scaling adjustment for tournament level heavily influenced and improved our models predictions, revealing this scaling factor plays a major role in tennis tournament prediction, which was not clear to us at first.

With regards to the K factor adjustment based on year, we were pleased with the approach we took in making a stepwise adjustment where further years are weighted less. The decision to scale this parameter and weigh older years less makes sense, as tournaments played further in the past have a lesser effect on the results of tournaments today. Factors such as age, style of play, and new opponents contribute significantly to a player's current capability. Every year players go through changes and our model is able to capture some of that change by weighting the older matches less.

The adjustment of winning probability based on age also made sense, but could also have been done differently. Our thought process was that newer and younger players have less experience than middle aged and older tennis players, and older players may get more fatigued as matches go on for longer. However, there are older players who have had significant success even when they have reached old ages. Our decision to use a normal distribution to model this adjustment was based on research we had conducted, but different methods such as an exponential function could be used instead. We learned that this age factor drastically changed model prediction for different parameters and distributions, which we found surprising. Without this age factor, we found that older players exceeded their predicted performance in comparison to the betting odds.

One key shortcoming of this analysis was the betting odds used. Since the analysis relied on betting odds for the 2023 tournaments, we had trouble finding correct betting odds as many websites we looked at did not show previous 2023 betting odds. We were able to find betting odds on 2 websites, CBS sports and Sportsbettingooddshistory[4] which has betting odds for past tennis matches. The CBS odds were not always complete, so we decided against using their odds. One drawback we had was we didn't have the ability to compare multiple booking companies' odds, such as FanDuel or DraftKings, and take the average based on their odds to compute players' winning probabilities. Using the betting odds on this site alone may not give us an accurate analysis of the betting odds given to these players and we are relying on the validity of their betting odds.

Another shortcoming was the amount of variables included in our analysis. In reality there are many factors that affect tennis matches, such as the weather, player health, or coaching changes. While we thought our methods seemed sensible, there are many other parameters we would need to include in order to make accurate predictions. The sports betting odds take all of these factors into account, which we were unable to. As noted in the analysis, when predicting tournaments such as Roland Garros which is played in May - June, our model would be less accurate. This is because there are hundreds of matches played from the end of 2022 until the Roland Garros, the betting companies are able to take these factors into account when forming their models, which our model did not account for.

5 Summary

Predicting tennis matches and tournaments is of interest to fans across the world. This project scraped the surface of one method to predict tennis matches, there are many other factors that must be taken into account when predicting these matches. It seems that the ELO rating system was a solid measure of player performance, but more work is needed to develop an accurate model. Our model captures differences that surfaces can make in predicting Grand Slams, as some players are expected to perform better on specific surfaces than others. Future analysis may focus on incorporating more factors into these models predictions to better match sports betting odds. Research into factors that heavily influence tennis matches as well as sensitivity analysis for their parameters are possible directions for extension of this model. Including different surface specific value parameters may enhance prediction, as surfaces themselves have many different traits that influence how players perform.

6 Attribution of Effort

We all contribute together on this project, working together in person at every step. We all began by figuring out what dataset to use and what questions we wanted to answer. Once deciding on a dataset, we began implementing the ELO calculation metrics and discussing parameters to adjust in these calculations. Nathan worked specifically more on the ELO calculation side and tournament simulation. John worked more on the error metrics and plotting. Yiming focused on figuring out strategies to simulate games based on different factors in our dataset, where we all decided to include only age. We worked together on each of these steps as well so that everyone understood what was going on each step of the way. We met weekly over the duration of the project to update each other and work on the project.

7 References

[1] ATP Tour. “FAQ about PIF ATP Rankings: ATP Tour: Tennis.” *ATP Tour*, www.atptour.com/en/rankings/rankings-faq. Accessed 11 Oct. 2024.

[2] Ferrer, David. “Home.” *Data Action Lab*, 5 Feb. 2022, www.data-action-lab.com/2021/11/29/analysis-of-age-in-tennis-data-2/.

[3] Admin. “Why Wimbledon Is the Greatest Tennis Tournament in the World.” *World Tennis Magazine*, 2 July 2023, www.worldtennismagazine.com/archives/22172.

[4] “Archived Tennis Futures Odds.” *SportsOddsHistory.Com | Archived Futures Lines of the Super Bowl, World Series & More*, 8 Apr. 2020, www.sportsoddshistory.com/tennis-odds/.

[5] Prieto-Lage, Iván, et al. “Match Analysis and Probability of Winning a Point in Elite Men’s Singles Tennis.” *PloS One*, U.S. National Library of Medicine, 28 Sept. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10538650/.