Modeling Tennis Player Pathways

Spring Capstone Project

Team Members

Pramit Vyas
Tyler Quinn
Dominick Robinson

Sponsor

United States Tennis Association

Sponsor Representatives

Alex Torres
Lauren Rothstein Wlodawski
David Ramos

Project Advisor / Coach

Will Townes

Carnegie Mellon University

April 2026

1 Abstract

This study identified development pathways for junior tennis players to help the United States Tennis Association (USTA) predict future professional success. By scraping over 3.7 million match results from CoreTennis and integrating historical datasets, we implemented a normalized Elo rating system to quantify player skill and remove era-specific inflation. We developed a Trajectory Similarity Score in order to identify comparable historical players and utilized a Generalized Additive Model (GAM) with linear spline basis expansion to project career results. This approach allowed for flexible, non-linear modeling of player development that remains computationally efficient for tens of thousands of individuals. The project resulted in a functional data pipeline and a Tableau dashboard containing visualizations of player trajectories and uncertainty estimates. Although the model successfully identified a general peak age of approximately 25 and noted that female players reach milestones earlier than males, it still relies greatly on similar players.

2 Introduction

The USTA is constantly interested in identifying junior tennis players who show high potential to perform well in the professional scene. Their larger objective is to grow the game of tennis in the United States, and they believe that developing and marketing more elite professional players will serve this goal. Thus, the purpose of this study is to accurately predict future outcomes for junior players. We achieve this by making models that predict how skilled players will be at future ages, and we visualize the predictions with an interactive Tableau dashboard. Outcomes of junior tennis players are highly variable based on how much they have played, how old they are, and random events such as injuries, so we capture full distributions of projected outcomes for players rather than mere point estimates in these predictions. These projections can help the USTA in identifying players with potential to achieve high ranks (e.g. Association of Tennis Professionals (ATP) top 100, Women's Tennis Association (WTA) top 100) at an early age. USTA can then allocate their Player Development resources accordingly. For instance, they may want to provide specialized coaching to players who demonstrate high potential, and they can use this work to identify these players.

We use a multi-faceted approach in order to generate distributions of future outcomes. First, we create a normalized Elo system in order to estimate player skill. Then for each player, we find their 5 nearest neighbors by calculating the Manhattan Distance from all other players based on their smoothed Elo trajectories. We also fit a GAM with linear spline basis expansion for each player. The predictions for each player are then given by a weighted average of their own GAM coefficients, the GAM coefficients for their 5 nearest neighbors, as well as the coefficients from a global GAM added for regularization.

Finally, we create a dashboard in Tableau to allow the USTA to search for players to query their biographic data, recent and best results, rating history and predictions, as well as comparisons with other (similar) players.

3 Methods

3.1 Data

Our data came from two sources: CoreTennis, and Jeff Sackmann's ATP/WTA Tennis Rankings, Results, and Stats.

3.1.1 CoreTennis

CoreTennis provides match results from singles competitions taking place all over the world for competitive players of all ages, both men and women. The data contained information about each player and tournament, and professional and junior International Tennis Federation (ITF) ranking data, starting from about 2006. However, there was no publicly available API to easily extract these data from CoreTennis, so we scraped the data from the website.

The scraping procedure is outlined in Figure 18. We ran the scraping procedure in mid-March 2026. It took about 10 hours to run fully, including random delays between requests to avoid possible anti-bot detection that CoreTennis may have in place to block web scrapers.

Data cleaning involved the following steps:

1. Tournament Normalization: Extracted unique numeric identifiers for each tournament from source URLs, standardized gender labels (e.g., mapping "Masc." to "M"), and parsed concatenated date strings into distinct start and end date fields.

2. Player Profile Parsing: Sanitized player metadata by extracting unique player IDs and converting physical attributes, such as height and weight, from descriptive strings (e.g., "185 cm (6 ft 1 in)") into standardized integer values.

3. Match Result Structuring: Developed a two-pass processing logic to map opponent names to IDs and transform match records into a "Winner/Loser" ledger. This included a robust score-parsing algorithm to extract individual set scores while handling retirements and walkovers.

4. Entity Reconciliation: Performed a vectorized alignment check between match records and tournament entities. Matches that did not align with existing tournaments were grouped, assigned new unique identifiers, and back-filled into the master tournament list to ensure referential integrity.

5. Specialized Event Flagging: Identified and flagged wheelchair-specific tournaments by searching for "WC" patterns within the tournament tier metadata.

6. Tournament Draw Linking: Implemented a "waterfall" matching logic to link qualifying and consolation draws to their respective main draw tournaments using a tiered hierarchy of attributes (name, year, gender, and location).

7. Redundancy Reduction: Dropped redundant columns, such as player names and tournament details stored within the match table, to normalize the database schema and reduce file size.

8. Match Date Interpolation: Estimated specific match dates by interpolating the tournament duration across the sequence of rounds (e.g., Round of 64 through Finals), allowing for chronological analysis within a single tournament window.

9. Gender and Demographic Refinement: Refined player demographics by inferring gender from tournament category keywords (e.g. "ATP" or "Girls") and cleaned birth year data to ensure consistent numeric formatting for downstream analysis.

After cleaning, we obtained $212,677$ players, $117,347$ tournament editions, and $3,723,578$ match results.

Unfortunately, we found no way to scrape ranking data from CoreTennis, which is why we also used Jeff Sackmann's datasets.

3.1.2 Jeff Sackmann ATP/WTA Tennis Rankings, Results, and Stats

Jeff Sackmann has compiled datasets that include all ATP matches (futures, challengers, and main tour) and WTA matches (futures and main tour) from 1968 through 2024. ATP tournaments include men's tennis results, while WTA includes women's tennis results. Only the following data were relevant to our project:

1. **Player:** Player demographic information, including name, age, country, height, and so on.

2. **Ranking:** Weekly ranking history.

Because CoreTennis and Jeff Sackmann assigned different IDs to each player, we attempted to match players between each dataset with the following method by:

1. Standardized player names by removing accents and special characters (e.g., "Djoković" to "djokovic") and converting all text to lowercase.

2. Reconciled country codes (e.g. "CHL" to "CHI").

3. Extracted birth year from birthdate in Jeff Sackmann data, as CoreTennis data only optionally includes birth year but never birth date.

4. Standardized handedness code (e.g. "Right-handed" to "R").

5. Matched players by name and country code, breaking ties on birth year and then handedness. Instances where all fields matched separate players were dropped.

3.2 Data Analysis Plan

3.2.1 Quantifying Player Success

To quantify player performance, we implemented an Elo rating system based on match results scraped from CoreTennis. Elo is a popular method for estimating player skill in one-versus-one competitions, including chess and tennis. In tennis, Universal Tennis Ratings (UTRs) and the International Tennis Federation's (ITF's) World Tennis Number (WTN) are ELO-based measurements of skill.

Elo calculates an expected win probability E_A for player A against player B using the logistic function:

$$E_A = \left(1 + 10^{(R_B - R_A)/400}\right)^{-1} \tag{1}$$

Following an outcome $S_A \in \{0, 1\}$, the rating R_A is updated via:

$$R'_A = R_A + k(S_A - E_A) \tag{2}$$

The k-factor serves as a sensitivity constant, balancing rating stability against responsiveness to recent form.

To evaluate the predictive power and reliability of our model, we utilized three primary metrics:

- **Accuracy**: The proportion of correct predictions where the winner was the player with the higher win probability ($E_i > 0.5$):

  $$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\text{round}(E_i) = S_i) \tag{3}$$

  where $\mathbb{I}$ is the indicator function.

- **Brier Score**: A proper scoring rule that measures the mean squared error of the probability forecasts (Brier, 1950):

  $$\text{BS} = \frac{1}{N} \sum_{i=1}^{N} (E_i - S_i)^2 \tag{4}$$

  A score of 0 represents perfect prediction, while 0.25 indicates a model no better than random guessing (for a 50/50 event).

- **Expected Calibration Error (ECE)**: Calculated by partitioning predictions into M equally spaced bins (B_m) and taking the weighted average of the gap between the mean predicted probability and the actual win fraction (Guo et al., 2017):

  $$\text{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left|\text{acc}(B_m) - \text{conf}(B_m)\right| \tag{5}$$

  where $\text{acc}(B_m)$ is the true frequency of wins in bin m and $\text{conf}(B_m)$ is the average predicted probability for that bin.

We optimized the Elo model by sweeping the k hyperparameter across 24 values evenly spaced on the interval $[10, 250]$, with the objective of minimizing the Brier score.

To account for longitudinal shifts in the player pool, we normalized raw ratings by calculating a player's Z-score relative to the annual distribution of the active player pool. For each match, a player's rating is benchmarked against the mean and standard deviation of all active players within that calendar year. This Z-score represents skill as a function of standard deviations from the contemporary mean, effectively stripping away era-specific inflation. We then rescaled these scores to a stationary scale ($\mu = 1500, \sigma = 200$) based on each year's internal distribution. This yearly normalization ensures a stable rating scale that allows for fair performance comparisons across different decades by neutralizing the effects of rating drift and pool expansion.

3.2.2 Trajectory Similarity Score

We also created a metric to quantify how "similar" two players are in their development. Using our normalized Elo metric as before, we first created a standardized grid of 0.2-year intervals ranging from ages 12 to 40 in order to account for varying match dates. A player's Elo rating for a certain age point on this grid is given by their calculated Elo rating after their match, if they played one that day. If they did not have a match that day, it is calculated by taking the interpolated Elo between the match immediately before the age of interest and the match immediately after. For example, if a player played games at age 16.1 and 16.3 with none in between, their estimated Elo rating for age 16.2 would be the mean of the two ratings calculated after those matches.

The Trajectory Similarity Score for two players is then calculated by taking the Manhattan Distance between their Elo ratings on this grid. If we let t denote a 0.2-year interval where both players have an interpolated Elo rating, and we let n denote the number of these intervals for which the players' careers overlap, the formula for the Trajectory Similarity Score between players i and j is as follows:

$$TPS_{i,j} = \frac{\sum_{t=1}^{n} |Elo_{i,t} - Elo_{j,t}|}{n}$$

We used these Trajectory Similarity Score results to find five comparables for each player in the dataset, to be used in our dashboard deliverable. There were added restrictions for these, as limiting possible comparables to those who have observed trajectories past the player of interest is desirable for projecting junior players' careers. For example, comparing a young player to a similarly aged young player is not useful because neither of their outcomes are known. The restrictions are as follows:

- Overlap period must span at least 2 years
- Comparable player's ending age must be greater than or equal to 3 years past the player of interest's ending age

3.2.3 Modeling – Mixed Effects

We wanted to use our measure of player skill to then build a model using it as an input as well as other features (age, handedness, height, region, and possibly other factors depending on what was available) to predict Elo at every future age year by year, as well as uncertainty estimates for those Elo ratings. Initially when predicting Elo by itself, we did not see smooth curves for the next year's prediction due to the input we take from other players. Thus, we decided to predict ΔElo (defined as the change in Elo from one game to the next for mixed effects modeling) as a way to measure only the changes in skill. We wanted to predict the trajectory of a player rather than predicting how good they will be, and then follow their trajectory up until a given age to predict with their skill at that age.

We considered several approaches for this. The first was a Mixed Effects Model with player effects pooled to the mean for players with few observations. We started by including age and a quadratic term for age as fixed effects in order to find a general quadratic curve for a player's trajectory. This model did a good job of modeling player trajectories, but it was extremely computationally expensive in predicting the random slopes for over 86,000 players giving us 6.4 million match rows. With such a large number of players, the model has to simultaneously estimate a massive covariance matrix just to track the individual variances and the correlations between the random intercepts and slopes. This was very computationally expensive. Therefore, we simplified the model such that the intercept term was the only random effect. Since the response was ΔElo, this model would still change slopes for predicted Elo. This model had a key weakness though, in that it could not handle non-linear curves for players. Non-linear patterns were often present in observed trajectories (e.g. a player who peaked and then slowly declined in ability throughout their 30's), so we moved to another approach.

3.2.4 Modeling – GAM

After running into the computational bottlenecks of the Mixed Effects model, we transitioned to a Generalized Additive Model approach, which allowed for more flexible, non-linear trajectories while still being computationally efficient enough to handle tens of thousands of players.

We applied a linear spline basis expansion rather than estimating a global parametric curve, fixing six knots at ages 18, 22, 26, 30, 34, and 38 across the entire population. Because of the fixed basis, every player's career trajectory was reduced to a vector of eight coefficients (an intercept, a linear age term, and six truncated power terms), making trajectories directly comparable as low-dimensional feature fingerprints and solving each player's curve as a simple ordinary least squares problem. We first fitted a global average trajectory over pooled yearly-aggregated delta-Elo values to serve as the population prior. For each player, we aggregated their match-level ratings to integer age-years, computed year-over-year differences (ΔElo), and fitted an individual spline via Empirical Bayes shrinkage toward the global mean, specifically:

$$\beta_{\text{blended}} = (1 - \lambda)\beta_{\text{own}} + \lambda\beta_{\text{global}},$$

where the shrinkage weight decays exponentially as a function of the number of observed age-years rather than raw match volume:

$$\lambda = \exp(-n_{\text{years}}/\tau).$$

This ensured that players with sparse careers were pulled toward the population curve to prevent overfitting, while those with long histories more heavily retained their own estimated curve.

Rather than constructing a KD-Tree over coefficient fingerprints at runtime, we leveraged the similarity tables from the player similarity score calculations at $k = 5, 10,$ and 30 nearest neighbors, each encoding the top matches and their Trajectory Similarity Scores for every player. We found the most accurate results using only the 5 nearest neighbors. For a target player, we retrieved their list of neighbors and computed a weighted average of neighbor coefficient vectors, yielding an ensemble coefficient that represents the career trajectory of their closest historical analogs. This ensemble was further regularized through a time-aware blending function: in the near horizon (within roughly two years of the player's last observed age), the player's own fitted spline was given additional weight, decaying linearly to zero as the forecast extended further into the future. We can reference figure 16 to examine how the global spline compares to player specific splines to get an idea of what the differences will look like when weighting. At longer time horizons, where neither the player nor their nearest neighbors had observed data, the ensemble blended smoothly toward the global prior curve, with the global weight ramping up to a ceiling of 35 percent over six years. Neighbors with insufficient historical coverage at the target age are themselves partially regularized toward the global prior before entering the ensemble, preventing the propagation of unreliable extrapolations.

Projection proceeded by evaluating the blended ΔElo curve on a dense age grid from the player's last observed age to the target age, then by accumulating these predicted annual changes via a cumulative sum anchored at their last observed Elo rating. To construct confidence intervals, we evaluated each retrieved neighbor's coefficient vector independently on the same age grid, computed the standard deviation of these neighbor-level predictions at each step (floored at five Elo points per year to prevent degenerate intervals when neighbor predictions converge), and propagated uncertainty forward as:

$$\sigma^2_{\text{total},T} = \sum_{j=\text{start}}^{T} (\sigma_{\text{age},j} \cdot \Delta t)^2.$$

The 90 percent confidence interval is then:

$$\text{Elo}_T \pm 1.645 \cdot \sqrt{\sigma^2_{\text{total},T}}.$$

The same confidence interval construction was applied to rolling-window backtests, where for each historical year of data Y, the player's spline is re-estimated on data up to and including Y only, ensuring strict causal constraints, and the one-step prediction uncertainty is derived from the same neighbor spread at that anchor age.

Then we derived an error metric to be used for tuning our model. This error metric needed to measure how well the model performed on unseen data without any possibility of feature leakage. Thus we fit the model at each age for all players, used the most similar players based on their trajectories up to the age for which we projected, and then fit the model to their future ages. Again we predicted ΔElo and then the calculated cumulative sum of their ΔElo ratings to project their future ratings. Then we calculated the error based on the differences between the projected Elo ratings and the observed Elo ratings for years after the data used to make the projections.

3.3 Tableau Dashboard

We created a dashboard in Tableau to allow users to search for players to query their biographic data, recent and best results, rating history and predictions, as well as comparisons with other (similar) players. We decided to use Tableau over alternative tools because USTA primarily uses Tableau for internal projects.

4 Results

4.1 Elo System

We optimized the Elo k factor using a three-stage iterative zooming sweep that narrowed the search window around the best-performing values. By minimizing the Brier score through successive refinements, we pinpointed an optimal $k \approx 62$ that maximizes probabilistic accuracy. This configuration achieved a Brier score of 0.1884, an accuracy of 71.03%, and an ECE of 0.0003, all of which were near or at the optimum across all Elo variations in the search. A more comprehensive view of the optimization results is illustrated in Figure 11. Furthermore, Figure 12 provides the calibration plot for the selected model, demonstrating high alignment between predicted win probabilities and observed frequencies.

We then normalized the Elo ratings over time to allow for rating comparisons across time, as outlined in the Methods section. Following normalization, the ratings become much more constant for each ranking over time, allowing us to accurately compare the ratings over time, as seen in Figure 1.

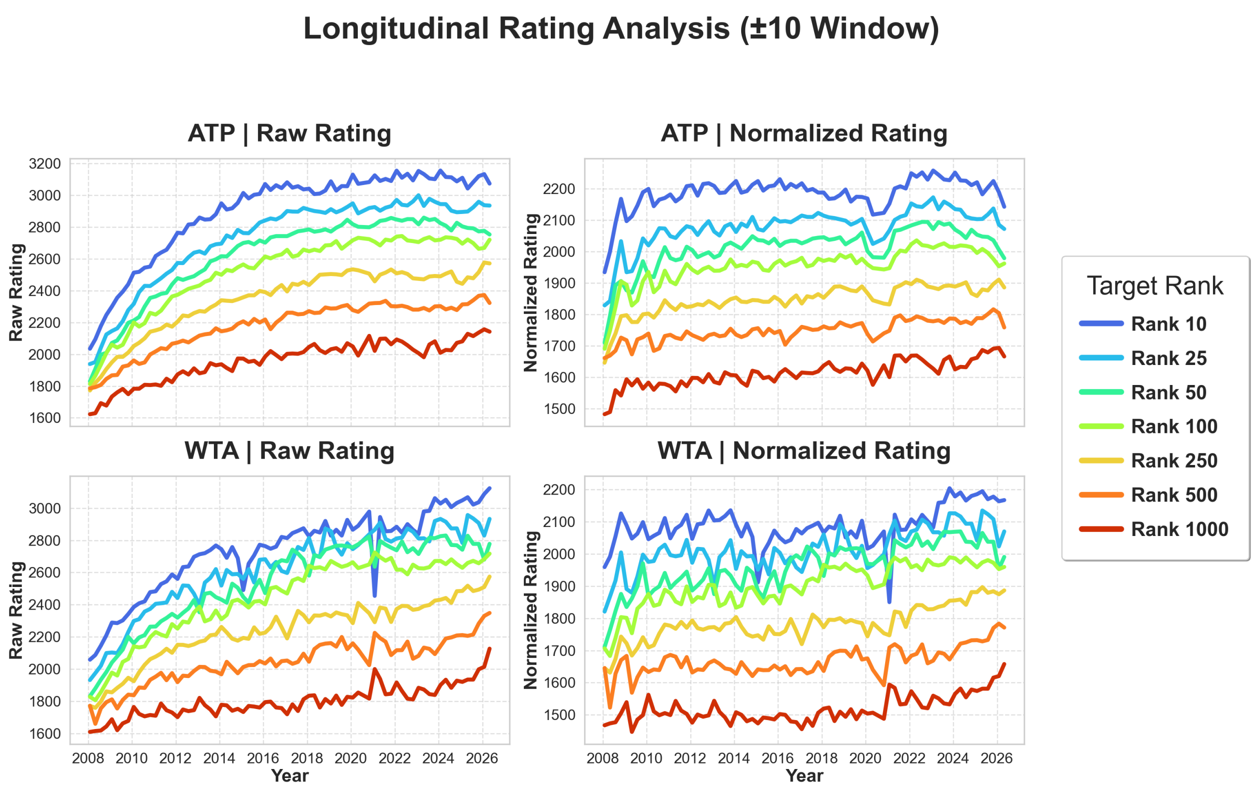

Figure 1: Rating by ranking before and after normalization for ATP and WTA tours. Our normalization process successfully normalizes ratings over time, allowing for better comparisons of ratings across time.

4.2 Trajectory Similarity Score

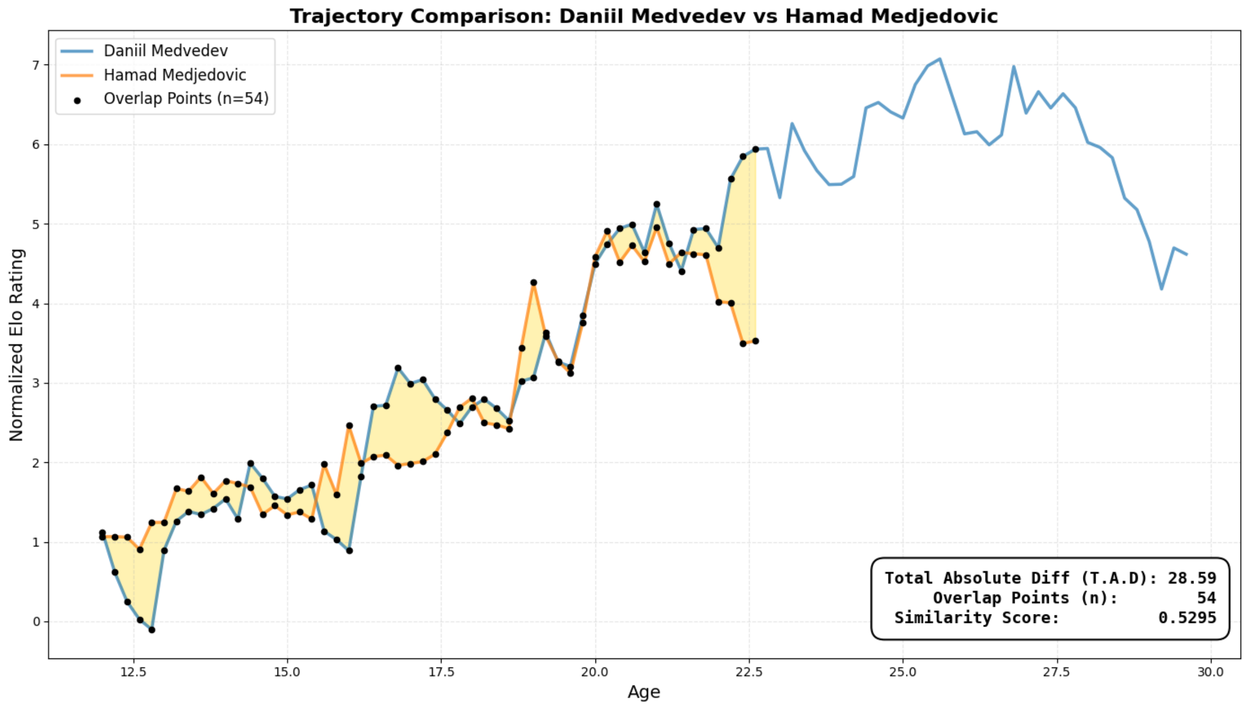

Figure 2: Comparison of Daniil Medvedev's trajectory and that of his closest comparable, Hamad Medjedovic. We see that this formula and criteria gives a directionally accurate way to measure how similar two trajectories are.

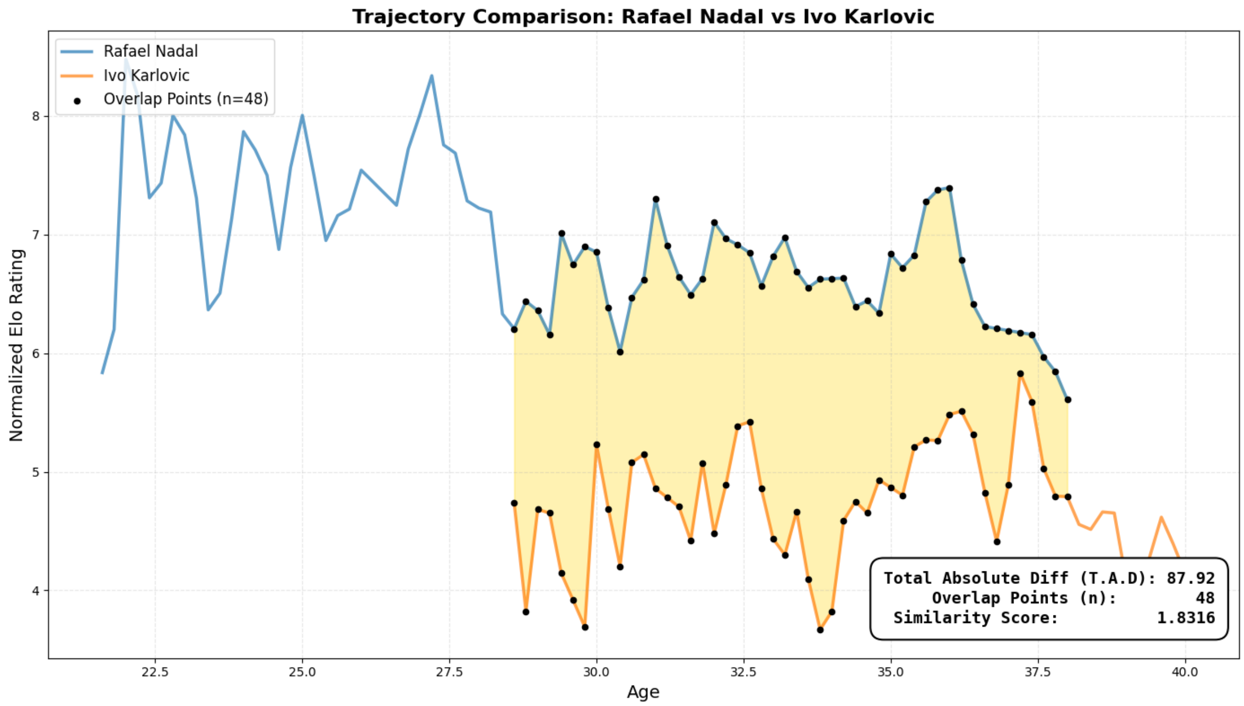

Figure 3: Comparison of Rafael Nadal's trajectory with another longtime player, Ivo Karlovic. We see that their ratings are farther apart than the previous 2 players, and thus their Similarity Score is much higher (i.e., the distance between their curves is higher, so they are less similar).

4.3 Modeling

With our fitted GAM model, we have identified a global age curve for tennis performance.

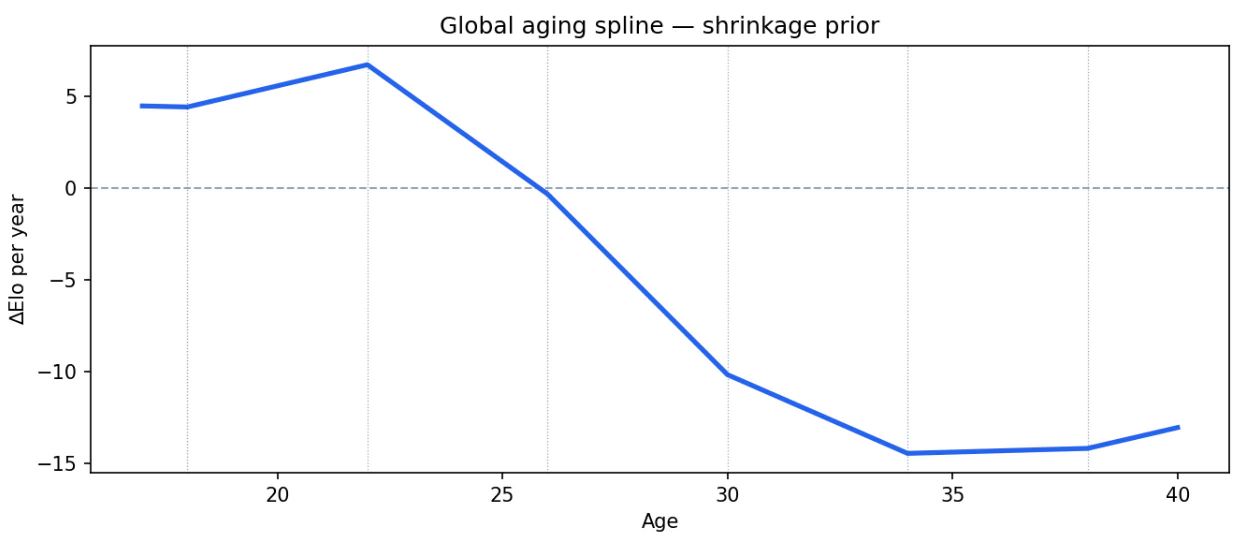

Figure 4: This shows the predicted change in Elo per year from ages 17 through 40

We can see in Figure 4 that players are expected to gain Elo on average until approximately age 25, implying that the typical peak performance for players in the data is around 25 years old. This aligns with tennis intuition confirming that this modeling technique performs as expected.

We also wanted to investigate our GAM coefficients to see that they properly identify clusters and outliers of players, so we created and viewed Figure 15.

From this Figure we can see that the majority of the coefficients for the trajectories of the players cluster around the center. This aligns with previous knowledge about how most players should have similar spline curves as most players are mediocre juniors who do not make it far. We can also see that some of the players who do make it further in the rankings or who have very different trajectories than the others tend to have more games played. This shows that these "exceptional players" tend to also be longer lasting players. This shows that the GAM coefficients do help in identifying certain trajectories.

We also applied the model findings to Teodor Davidov, a junior player who is known to be a very strong player for his age, to see what our model's projections for him looked like.

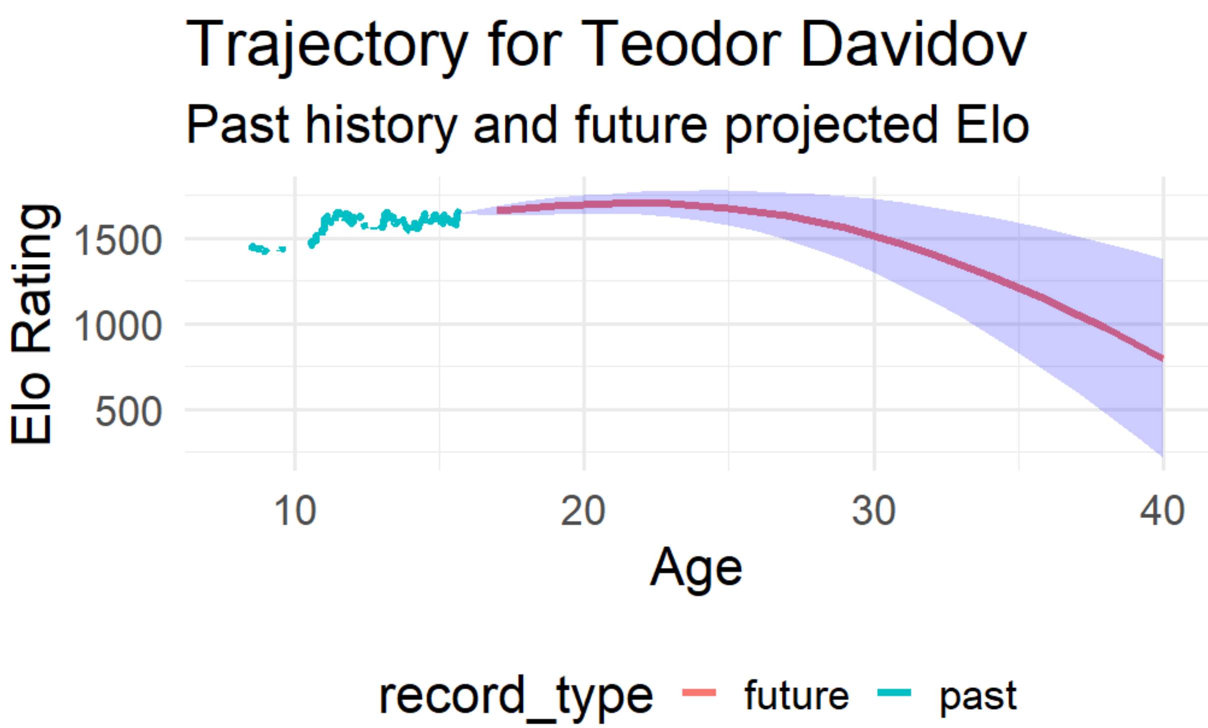

Figure 5: Teodor Davidov's elo forecast uncertainty increases by age

We can see in Figure 5 that the model projects the trajectory for Teodor Davidov to get more volatile as he gets older and it still projects him to be around 1600 Elo for his career. This aligns with expectations of more uncertainty at an older age as well as less room for error at a younger age. If needed, one can also see the top neighbors that were used to fit Davidov's spline in figures 13 14. We can see that although it may seem surprising that Davidov's trajectory isn't projected to increase as much, many of his neighbors had similar trajectories in fitting the model. The phenomenon of increasing variance farther out and seeing separate elo prediction curves based on player past trajectories, neighbors and global spline isn't just seen on certain players, we can also view randomly selected player forecasts in Figure 17.

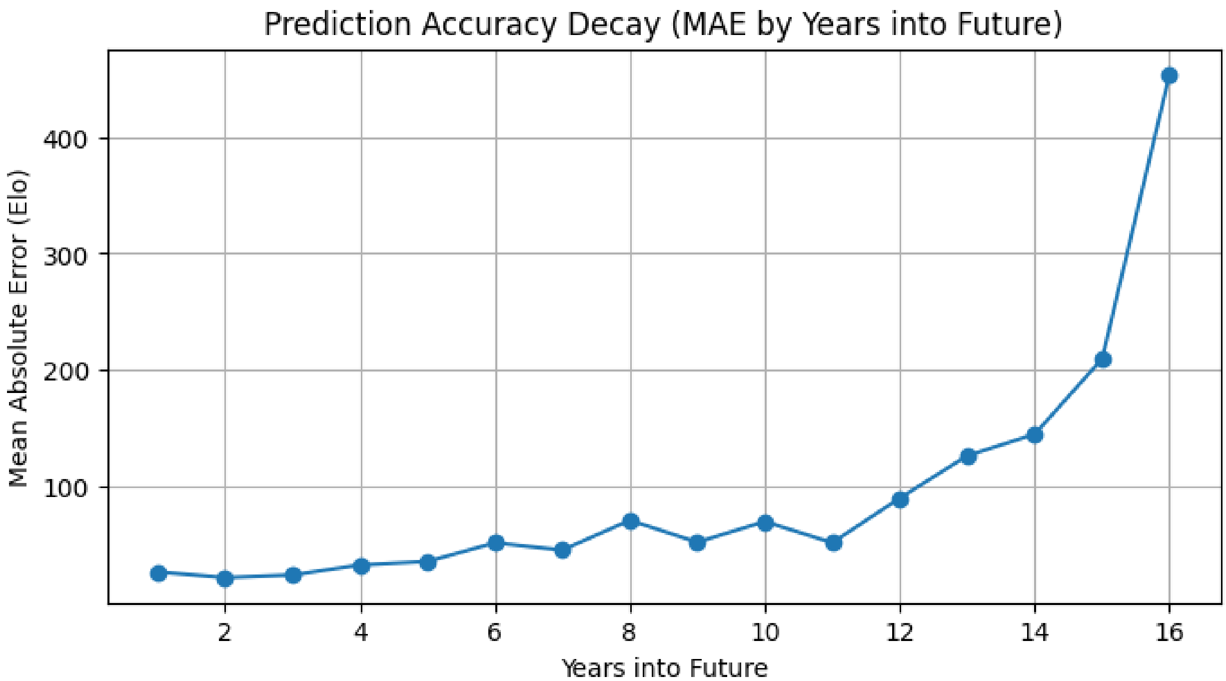

Figure 6: Model back-tested errors by age

In Figure 6, we can see that as the predictions are closer to the future, we become more uncertain about our estimates. Our Mean Absolute Error (MAE) increases at an exponential rate, starting around 11 years after the last observed data point. This shows that our model performs the best until approximately 10 years into the future. Overall, we observe a testing MAE rate of 27.2 Elo points and an RMSE of 36.25 Elo points. This is a significant improvement over other modeling frameworks that we employed. In other words, the test error was around 1.8 percent of the mean Elo, and well within a standard deviation of what it should be.

Now, we display some diagnostics to show that our model is fit properly.

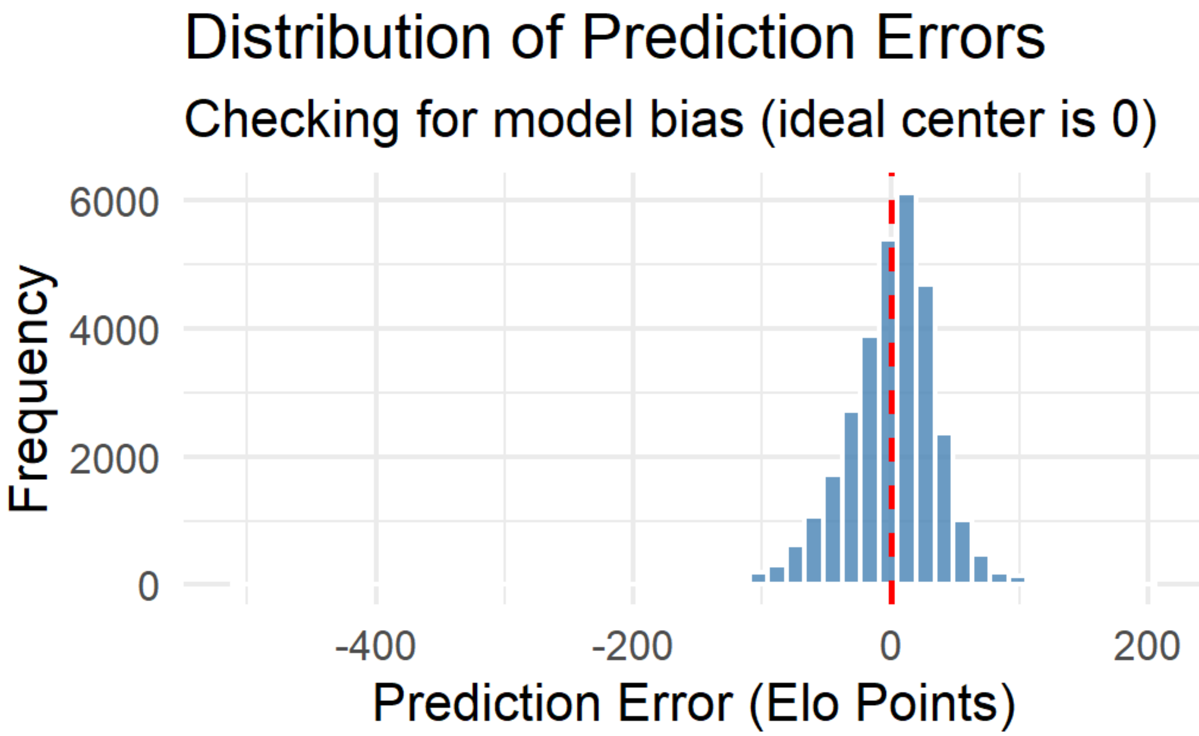

Figure 7: Here, we see that the model is not biased in any direction. We can see that the test errors for the model is normally distributed. This means that our GAM model met the necessary normality assumption.

Next, we examine the residuals. Note that as our predictions get further into the future, we tend to underestimate Elo more often. However, we contend that this is not too problematic, as the model will be used primarily for predictions within 10 years.

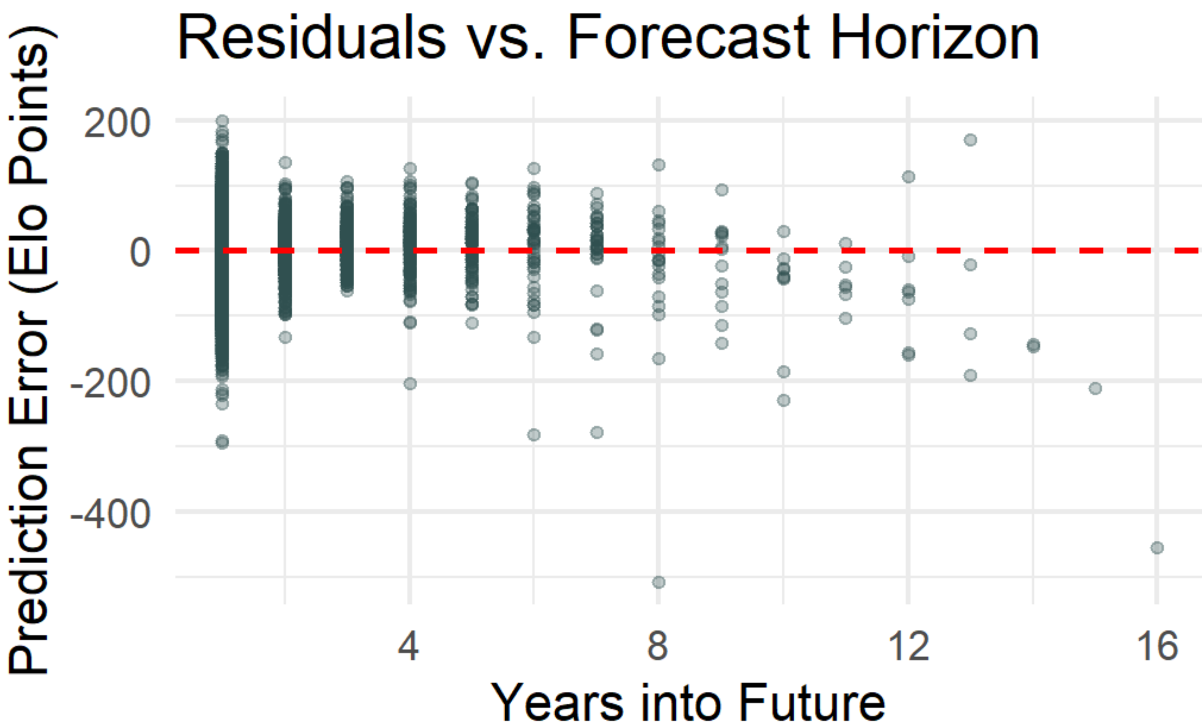

Figure 8: Here we can see that the residuals were mostly randomly distributed by age, which means that the model is not biased and fits the assumptions of random errors

We also visualize this for a single player that has many games played so far and is viewed as one of the top player (Jannik Sinner).

As we see in Figure 9, our model attempted to predict Jannik Sinner's trajectory by just looking at the errors from predicting his trajectory at every age using his closest neighbors at each age. We can see here that the projected trajectory follows Sinner's real trajectory quite close, as expected.

4.4 Dashboard

Finally, after combining our various data sources, calculating ratings, and generating predictions, we developed a dashboard, shown in Figure 10. The dashboard allows the user to search for any player in the data to obtain demographic information, recent and significant results, ranking history, and rating history and projections. It also allows users to compare the trajectories of multiple players at once as well as display the trajectories of players with the most similar trajectories to a player the user searches for.

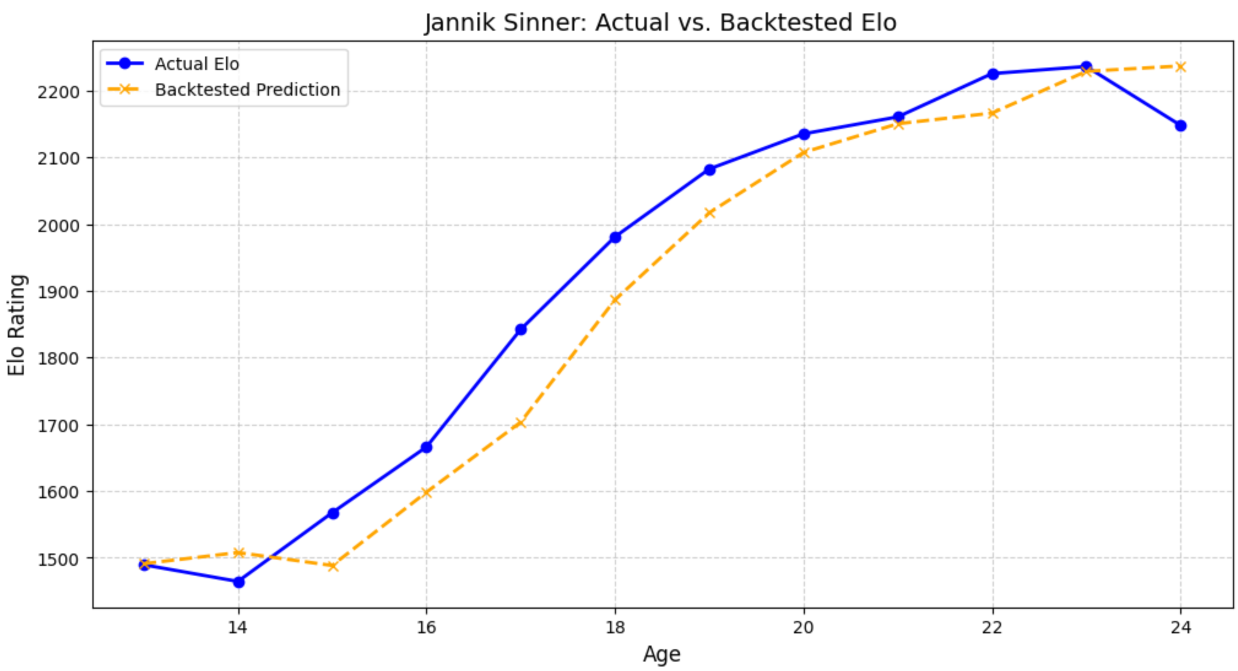

Figure 9: How the model does at predicting Jannik Sinner's Trajectory

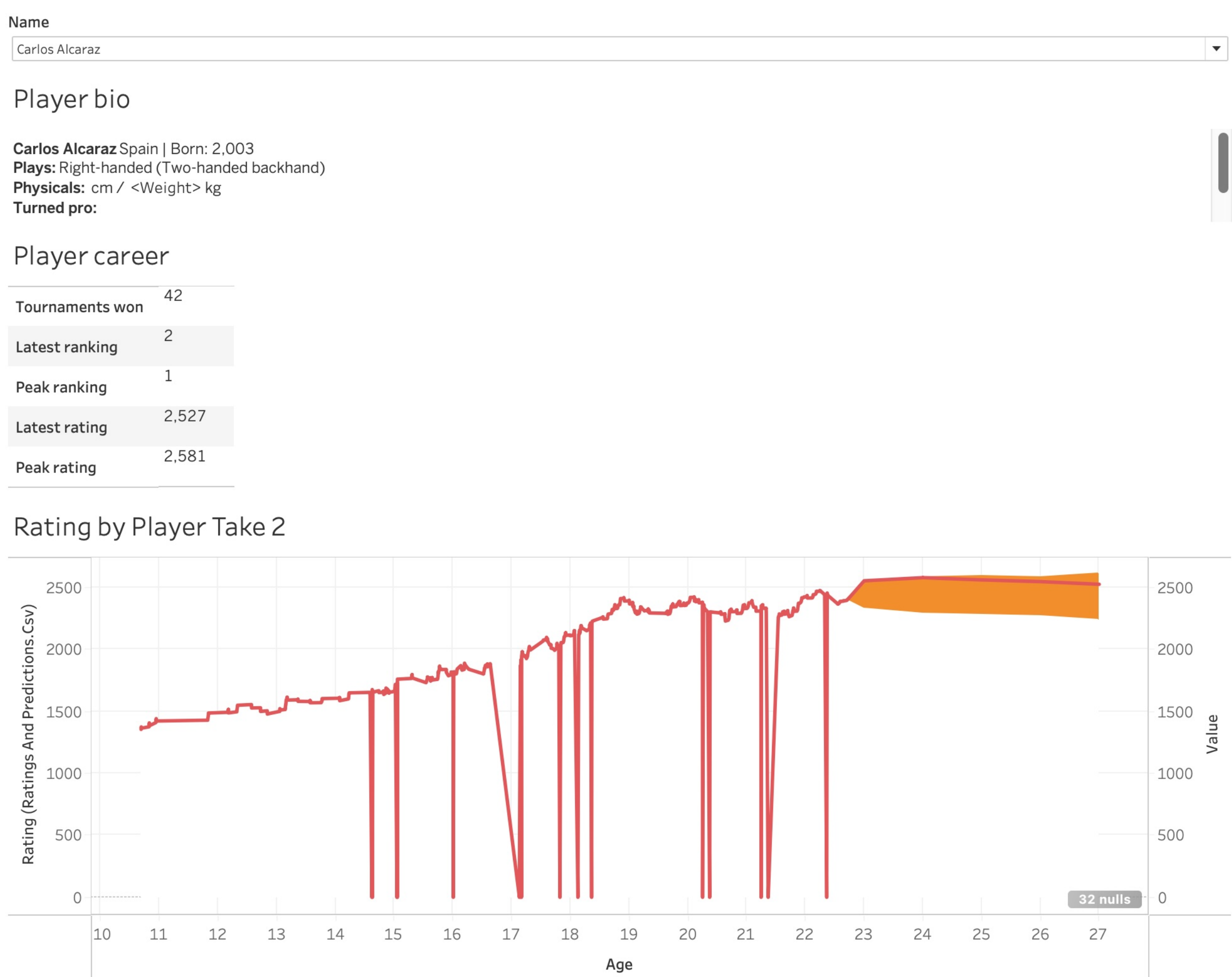

Figure 10: Prototype of dashboard. The dashboard will allow users to search for players and display their biographic and performance history, rating projections, and comparisons with other players.

4.5 Deliverable

We have created a deliverable pipeline that:

1. Scrapes data from Core Tennis website (can be replaced by API as USTA stakeholders see fit)
2. Cleans that data into a usable format
3. Creates a similarity score based on the trajectory curve and the player Elo similarities through time
4. Creates a Generalized Additive Model (GAM) that uses age at 6 knots throughout a player's trajectory and fits a regression between each. Uses the 30 most similar players' spline coefficients as well as a global shrinkage if needed (low sample size) to fit a prediction of ΔElo for a player.
5. Uses those predictions for ΔElo and then does a cumulative sum across the years they have until the age of 40 to find out their predicted Elo by age.
6. Uses the model output, the similarity output and the initial EDA we had done to create a Tableau Dashboard that shows the users what a trajectory may look like for a specific player.

5 Discussion

Through our modeling and player propensity work we observe that tennis players typically peak around age 25, and then their year to year change in Elo starts to decline. This matches subject matter knowledge with tennis professionals, as well as previously conducted research by the U.C Berkeley Sports Analytics Center (Brewer, 2023). We also observed via backtesting that our model predicts players' Elo ratings for the next 10 years within 100 on average, which is well within the overall standard deviation for our calculation of Elo of 200.

5.1 Limitations and Future Work

The current player-then-tournament scraping logic requires a full pipeline re-run to capture updates, as we can't identify new results without visiting every player profile. Switching to a tournament-then-player approach would allow us to skip completed tournaments and only ingest new matches and players. This pivot would be essential for enabling efficient, incremental updates in a production environment.

We currently use a single Elo system for all matches, but male and female players often follow different development trajectories– females, for instance, tend to peak earlier. Future iterations should explore optimizing separate ELO systems for each sex to better reflect these distinct properties among the sexes.

Additionally, we did not have access to player birth dates in the CoreTennis data, but it did contain birth year. Thus, our "age" field is not totally accurate. We estimated these ages as if every player was born on July 1st of their listed birth year in order to minimize the bias, but the player ages that we used can be off by as much as 6 months for players born near the beginning or end of their birth year.

We also have limitations that can be addressed through future modeling work. First, the GAM is very conservative due to the shrinkage parameters that attempt to alleviate some issues of variance with the model. Although this leads to more accurate overall errors, we still have to account for the fact that this may lead to inaccurate estimates on players who are truly exceptional outlier talents. This is a problem, since the main use case of the project is to project the trajectories of the top players. This leads to future work where there will be experimentation on reducing the shrinkage and further tuning the number of nearest neighbors to acquire a more variable projection that could project greater trajectories.

We also have a limitation of survivorship bias in the data that can be addressed in future work. Many players dropping out due to poor performance leads to later years for peak ages. This occurs due to tennis players who may not perform as well stopping their play and then us not being able to see their data past a certain age. These missing data are almost always biased, as dropping out strongly correlates to subpar performance in matches. This impacts when we use certain players who may have dropped out as nearest neighbors in our GAM model, as the model then goes towards global spline for those years, even though the dropping out of a nearest neighbor gives information. We believe this could be future work, working to impute dropped out Elo values, but we also believe that this will not impact our main use case. Our main use case is to predict top players, and most top player neighbors are not players who have dropped out. More data will also help with this– 18 years of data gives us many players and games, but when taking a longer term look at tennis careers like this, it only gives us one "generation" of data. Several more years of data will give more observed trajectories that can potentially be used as nearest neighbors, which may make the algorithm more accurate.

Future work might also look at other factors that can influence success in the top levels of tennis (e.g., adult height, handedness, serve speed). We strongly suspect that these are valuable in modeling future outcomes for junior players, but limited data accessibility and time made it impossible for this study.

6 References

- Brewer, Natalie (2023, November 12). *Beating the Clock: Age vs Performance in Professional Tennis.* Berkeley Sports Analytics. `sportsanalytics.studentorg.berkeley.edu/articles/beating-the-clock.html`

- Brier, G. W. (1950). *Verification of forecasts expressed in terms of probability.* Monthly Weather Review. `https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml`

- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017, August 3). *On Calibration of Modern Neural Networks.* `https://arxiv.org/pdf/1706.04599`

- Jeff Sackmann. *Tennis ATP Match Results Dataset.* Available at: `https://github.com/JeffSackmann/tennis_atp`

- Jeff Sackmann. *Tennis WTA Match Results Dataset.* Available at: `https://github.com/JeffSackmann/tennis_wta`

- Jeff Sackmann. *Tennis Match Charting Project*. Available at: https://github.com/JeffSackmann/tennis_MatchChartingProject
- Association of Tennis Professionals (ATP). *ATP Rankings*. Available at: https://www.atptour.com/en/rankings/singles
- Ultimate Tennis Statistics. *Elo Ratings*. Available at: https://www.ultimatetennisstatistics.com/eloRatings.

7 Technical Appendix

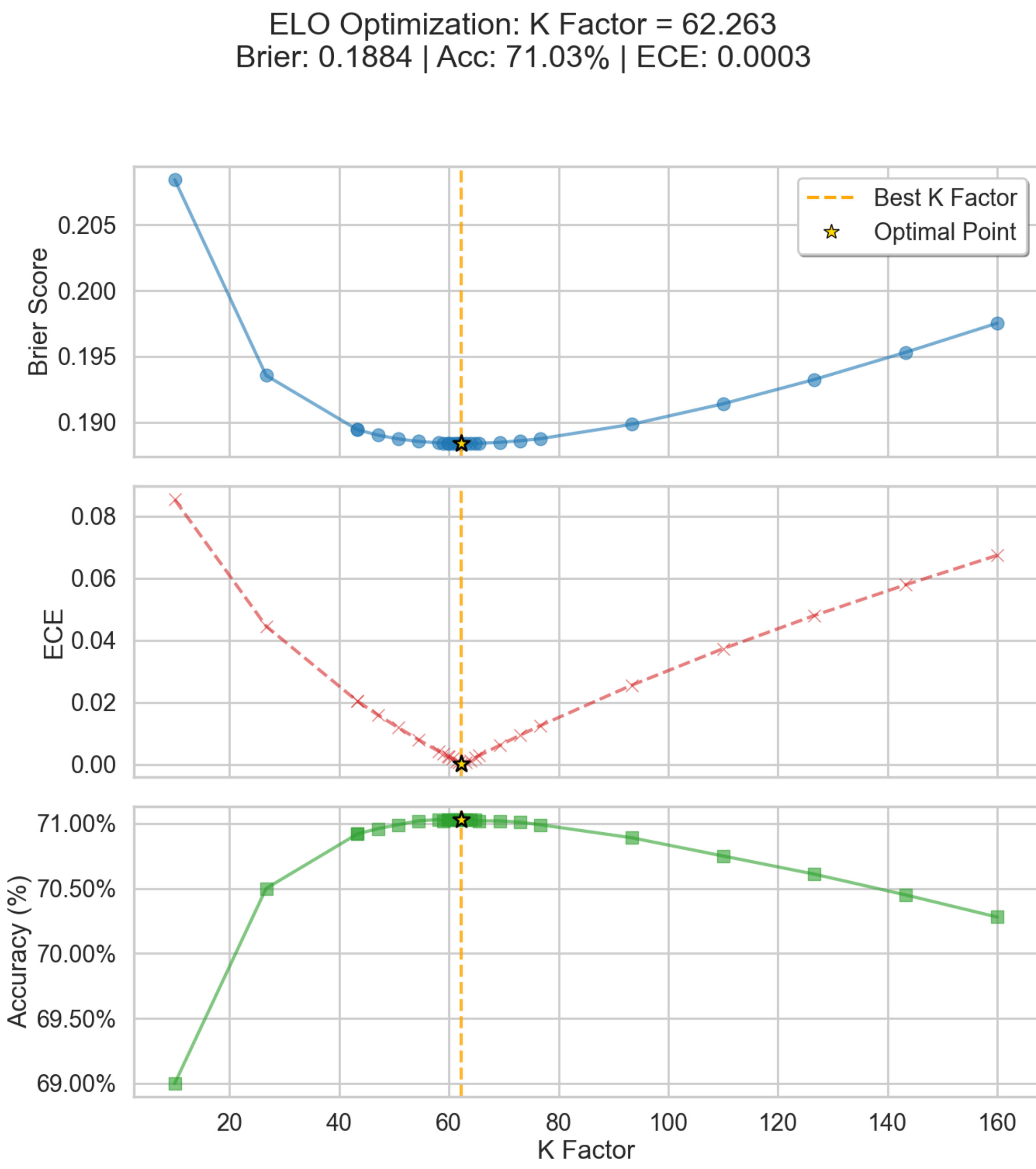

Figure 11: Elo optimization sweep over the k-factor parameter. The optimal point (marked with a star) represents the minimum Brier score.

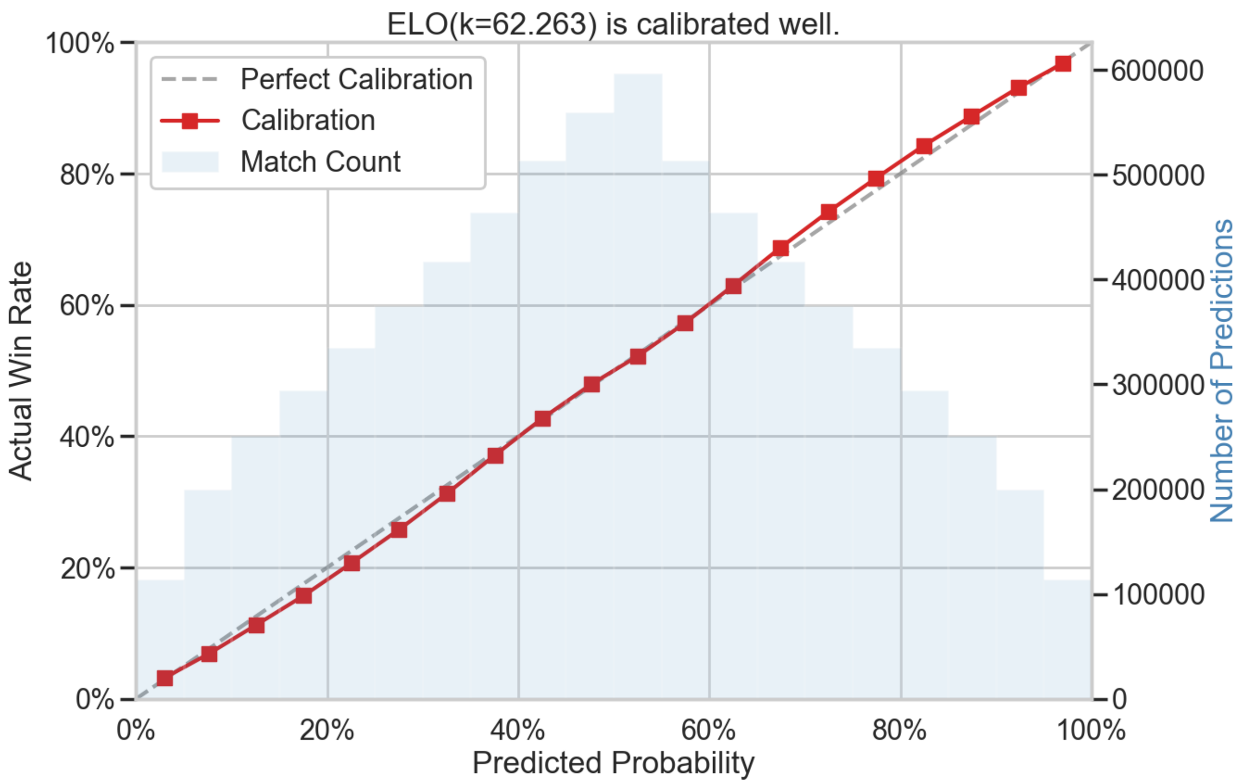

Figure 12: Calibration plot for the Elo model with $k = 62.263$. The red line closely follows the ideal diagonal, indicating a well-calibrated model.

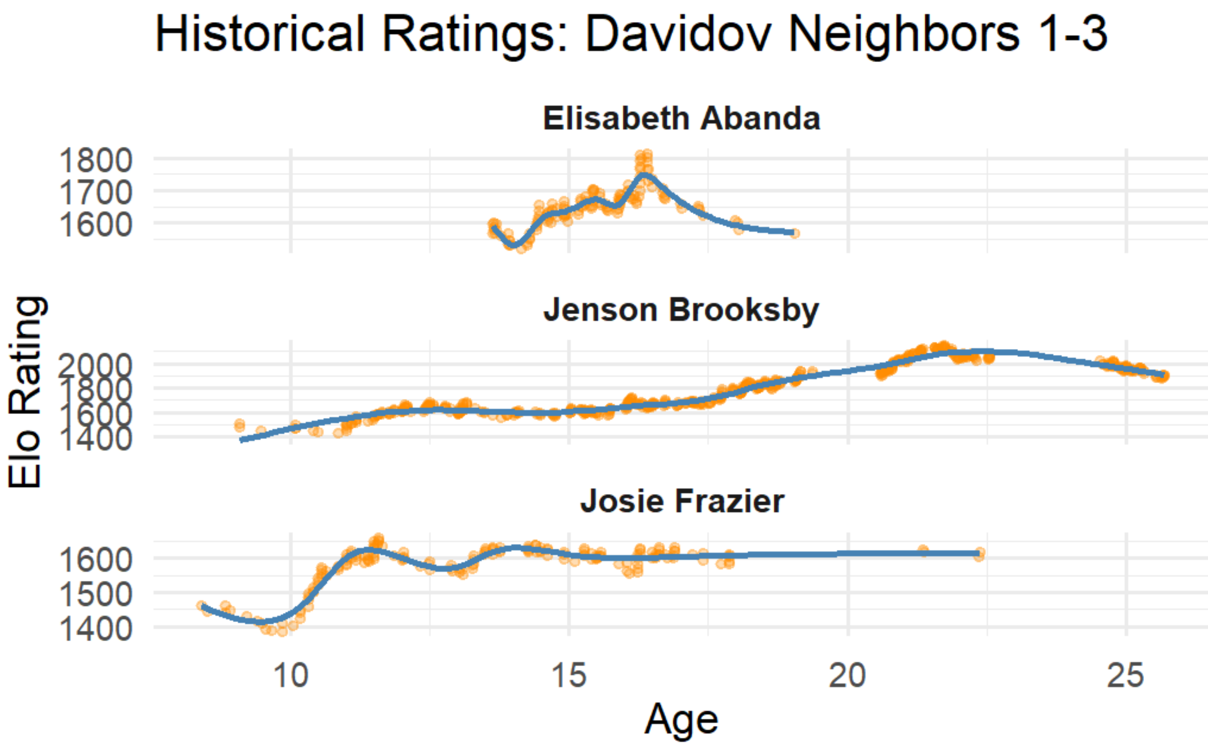

Figure 13: Teodor Davidov's 1st through 3rd nearest neighbors' Elo trajectories with fitted Spline

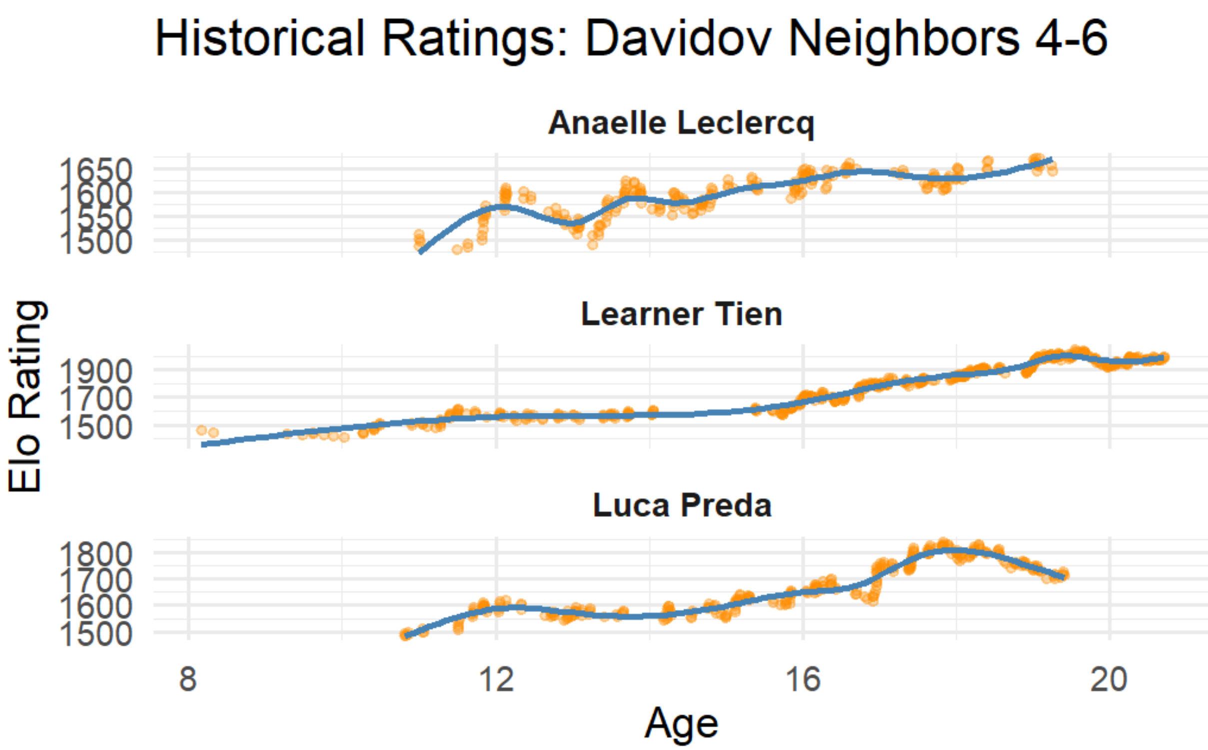

Figure 14: Teodor Davidov's 4th through 6th nearest neighbors' Elo trajectories with fitted Spline

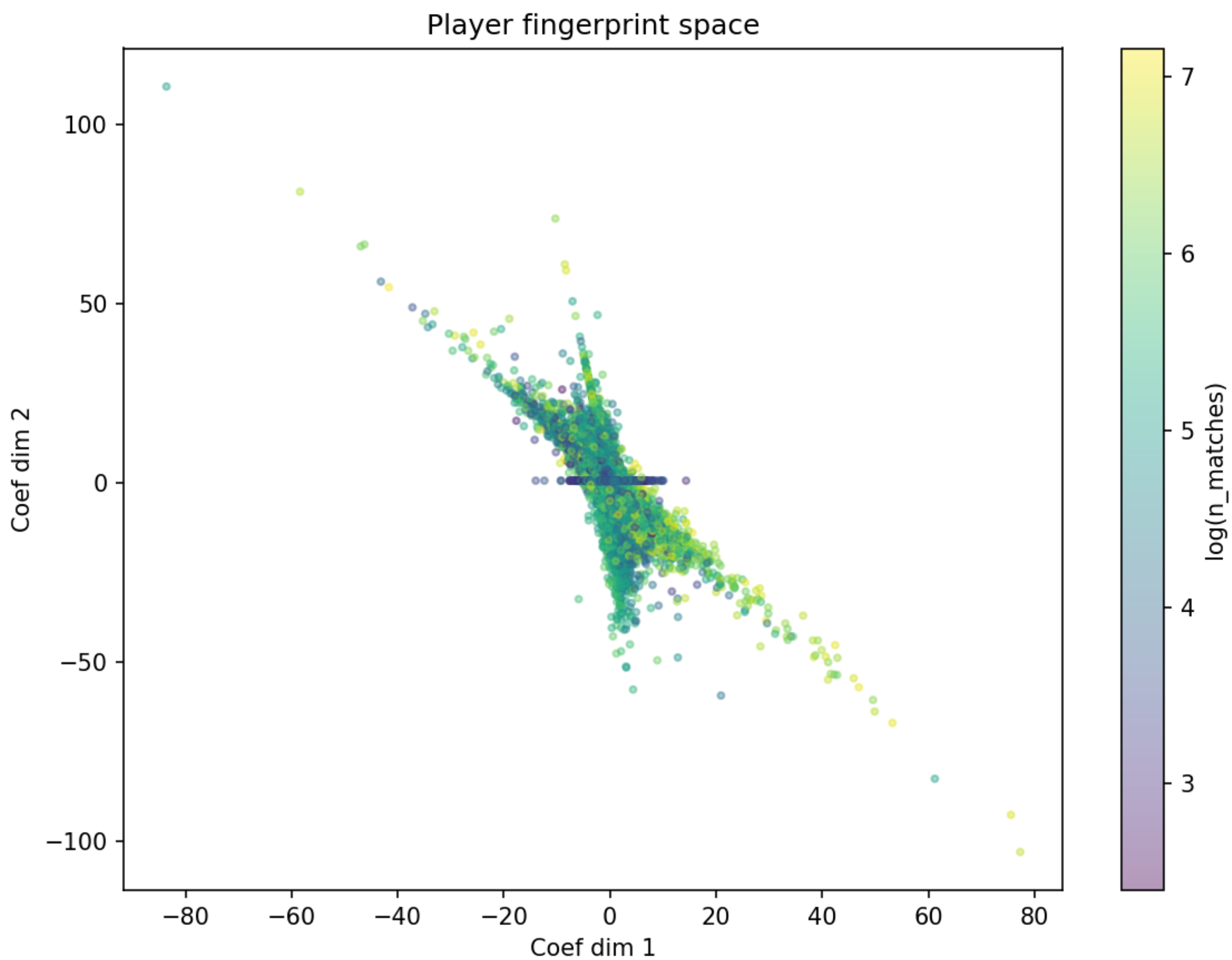

Figure 15: Finger Print Space of the GAM coefficients on a 2 Dimensional Scale

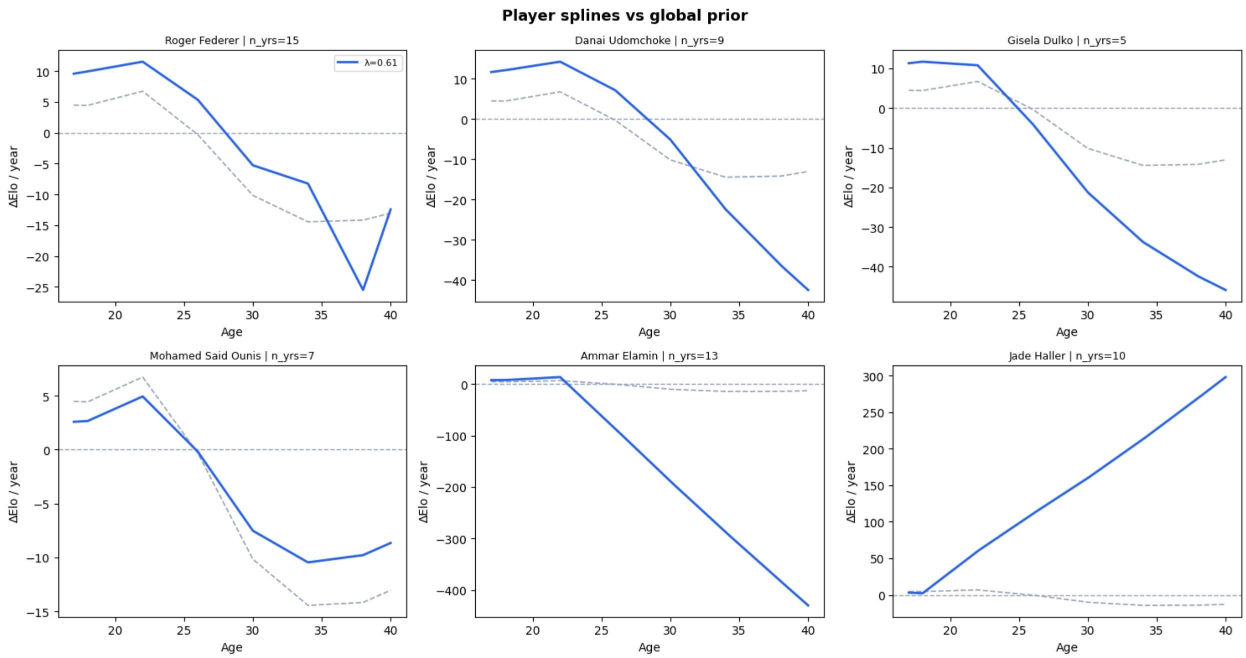

Figure 16: Individual player splines and how they compare to the overall spline (model took a weighted average of both)

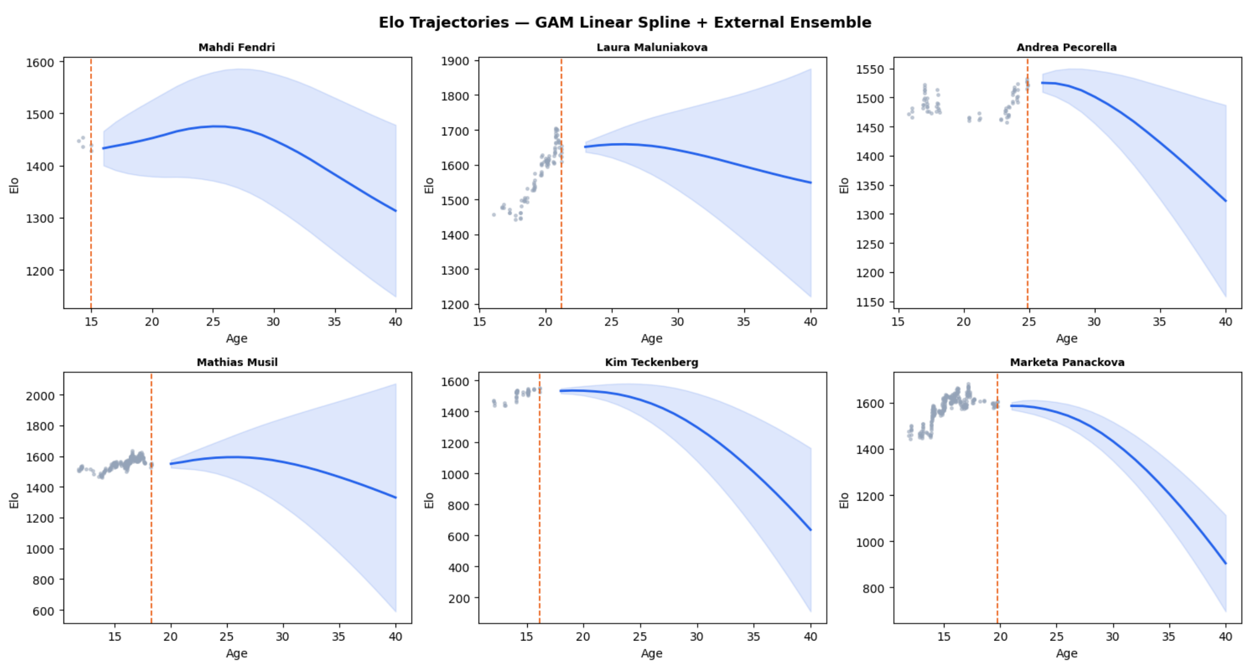

Figure 17: Splines and Forecasts for randomly selected players

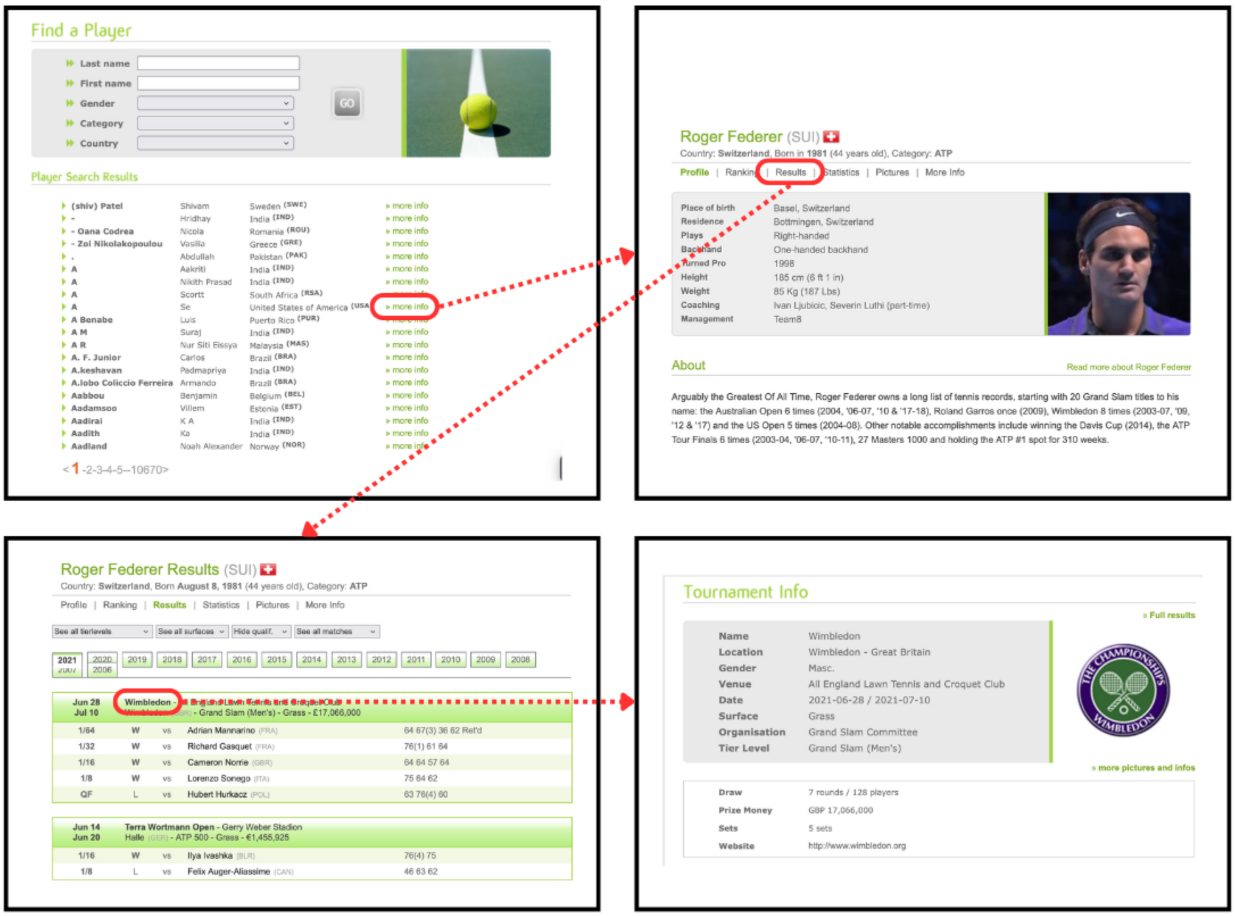

Figure 18: Overview of the multi-stage CoreTennis scraping pipeline. Player profile URLs are first harvested from the search index to access individual profile pages. From these, we extract links to specific player results, which in turn provide the URLs for individual tournament pages. The process concludes with the extraction of detailed data from each tournament profile.