Predicting Tennis Matches & Tournaments
Project 2 Expansion

Team 19: Nathan Dennis, John Breedis, Yiming Chen

Problem & Interest

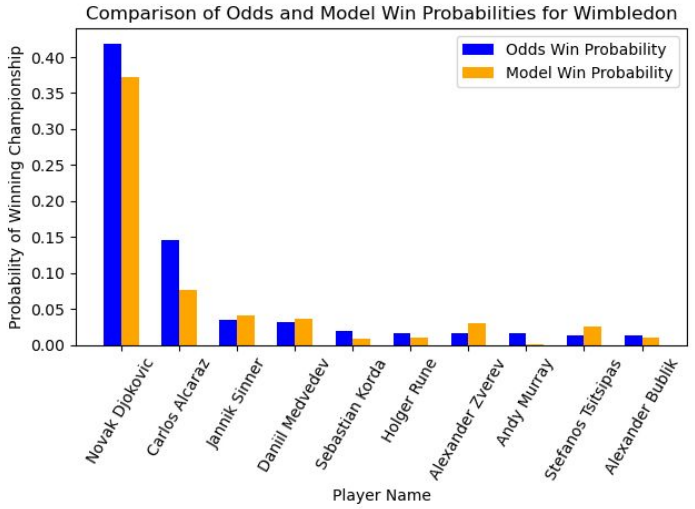
- Continue from project 1
  - Focused on difficulty in predicting tennis matches, many factors that affect match outcomes.
  - Applied different tennis match factors such as age and surfaces differences in project 1.
- More factors
  - Plan to expand and add more features to adjust winning probabilities, such as head-to-head matches.
  - (Potentially) Some playing style metrics, assign players a playing style based on research conducted. Determine which playing styles play well on specific surfaces or against other players.
- Expand user flexibility
  - Allow users to gather data from more tournaments rather than just grand slams (potentially).
  - Users can predict specific one-on-one match outcomes using our data rather than just tournaments.

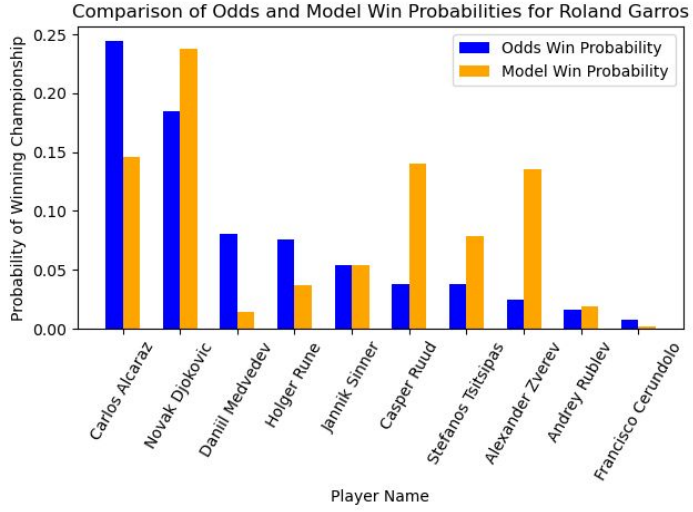
Problem & Methods

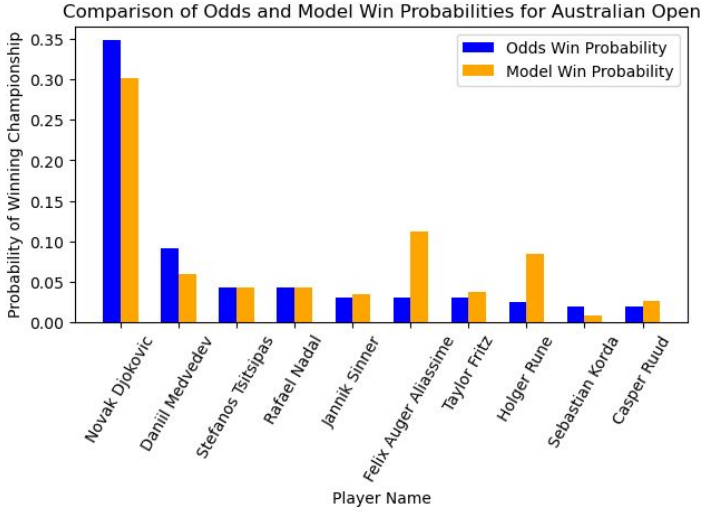
- Improve methods from project 1
  - Utilized different surface ELO calculation on first project, plan to potentially experiment with different scaling factors.
  - Adjusted winning probability based on player ages, refine and improve on this metric where we applied a normal distribution to compute adjusted probabilities.
  - ELO calculation based on year, stepwise.
- Apply new methods
  - Calculate head-to-head winning percentage for players to adjust their winning probabilities against common opponents. Can be very effective for one-sided rivalries.

Project 1 Results

- Our probabilities vs betting odds for Grand Slams.
- Top 10 betting odds probabilities.
- Hoping new metrics can improve our predictions.
- Very inaccurate for Roland Garros.







Error Metrics

- Roland Garros most inaccurate predictions. Clay may be unpredictable in general, we plan to investigate further into why.
- RMSE, max difference between predictions, average absolute difference.

Tournament	RMSE	L	L1
Australian Open	0.00983	0.06162	0.00403
Roland Garros	0.01979	0.10986	0.00609
Wimbledon	0.01153	0.07680	0.00433

Next Step

- Refine age winning probability adjustment factor
  - In the first project, we created a normal distribution with mean 25, which we found to be the prime age of tennis players, standard deviation 25.
  - Idea is that younger and older players winning probabilities decay as the match goes longer due to fatigue and experience level.
  - The density of this distribution is multiplied to a players winning probability for each match in a set. Feedback was to make this more clear and improve this idea.
  - Plan to improve this approach with more background research and experimentation.
- Refine adjustment for K factor in ELO calculation based on year
  - Currently stepwise, will look into exponential function.
  - Weigh recent matches higher, larger K factor. Older matches smaller K factor.

Next Step (continued)

- Calculate head-to-head matchups
  - We plan to calculate winning probabilities players have against each other to determine head-to-head winning percentages.
  - Take into account both winning percentage and quantity of matches played against each other.
  - Planning to apply this within the game prediction section to adjust winning probabilities.
- Investigate playing styles for players

Final Project Plan

- Finalize results for singles on project 2 (this project).
- On final project, implement predictions for doubles matches.
  - Use methods from singles matches to influence doubles match prediction.
- Library where users can predict both singles and doubles matches.
  - Users will be able to extract initial draws and simulate matches/tournaments.
  - Users can simulate specific matches based on statistics, such as ELO scores, age, and head-to-head winning percentage.