Predicting the Ace: Machine Learning for Professional Tennis Serve Direction

Rino Cattabiani

University of Connecticut

December 3, 2025

Research Question

- Can machine learning models accurately predict the direction of a professional tennis player's first serves? That is, can the models do better than guessing (random chance)?

Research Question Importance

- Players win about 69%–75% of points behind a strong first serve, compared with only 55%–57% on second serves across the ATP (Prieto-Lage et al.).
- Helps coaches and players evaluate the predictability of their serves (leading to new patterns of serving to stay unpredictable).
- Knowing where the server is going to serve 'most of the time' can greatly help the returner in anticipation to hit a great return.

Tennis/Tennis Serve Preliminaries

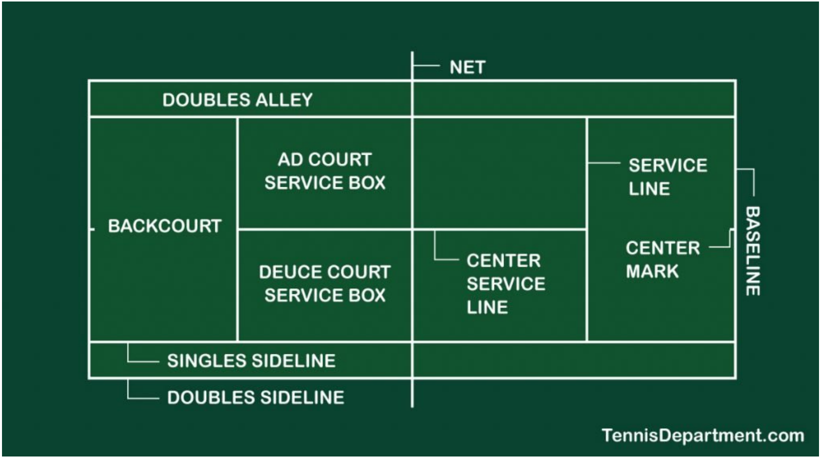
Tennis

- A tennis match is divided into sets, games, and points.
- A tennis court is divided into two sides: the deuce side and advantage (ad) side.
- Tennis matches are played on three kinds of surface: Hard, Clay, and Grass.

Tennis Serves

- Each player gets two serves to start every point (if the first serve is a fault, the player uses their second serve).
- There are three 'main' areas to serve the ball: **T** (down the center service line), **Body** (to the returner), and **Wide**.
- First serves are faster and more risky, but rewarding; Second serves are slower and safer, but less rewarding.

Image of Tennis Court

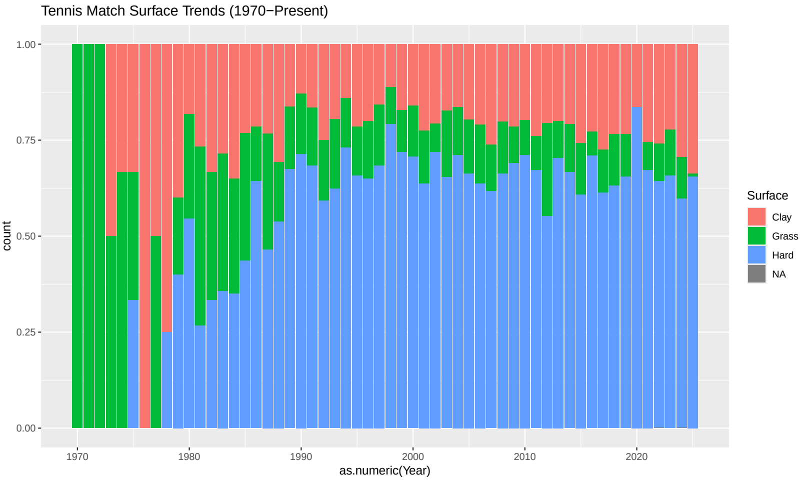


Data

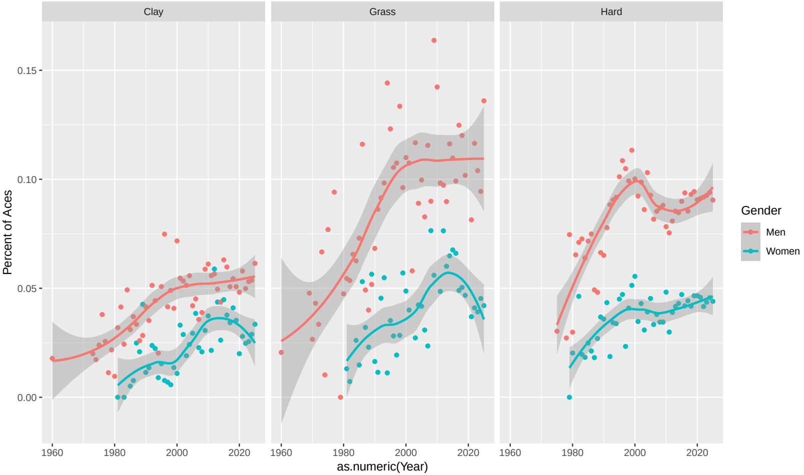
Data Set

- Data is taken from the Match Charting Project (Open Source, 1960-Present).
- Here are some of the files I used:
  - charting-m/w-points-2020s
  - charting-m/w-stats-ServeBasics
  - charting-m/w-stats-ServeDirection
  - charting-m/w-stats-ServeInfluence
- For the ML modeling, I chose to work with data from 2020-Present so as to keep up with trends.
- Worked with men and women who played 60+ matches.
- Total serves per player: Men $\sim$ 10,000, Women: $\sim$5,500

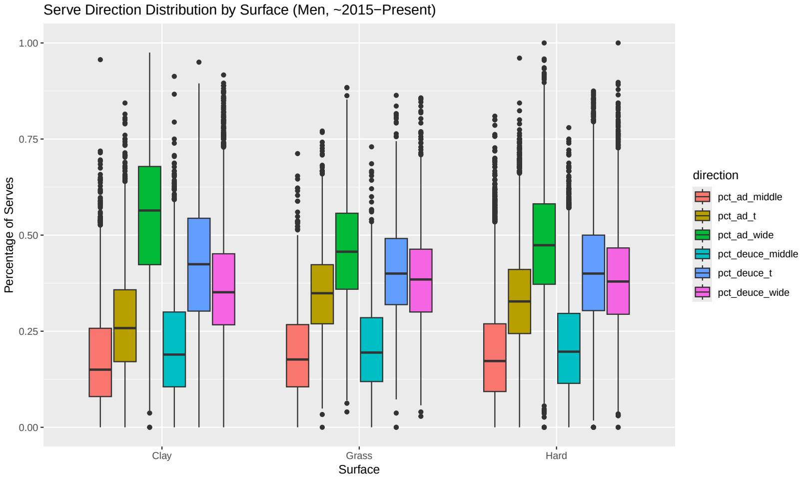
Tennis Match Surface Trends (1970-Present)



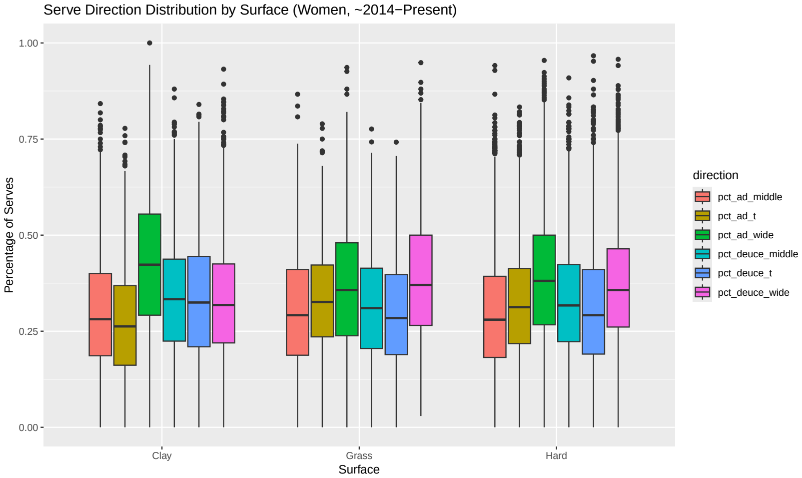
Ace Percentages Over Time Across Every Surface (1960-Present)



Serve Direction Distribution by Surface (Men, 2015-Present)



Serve Direction Distribution by Surface (Women, 2014-Present)



Example of Features the Models Used to Learn:

Serve History Features:

- t_count, body_count, wide_count
- t_won, body_won, wide_won
- consecutive_same_direction

Score and Pressure Features:

- is_break_point, is_game_point
- is_deuce, is_tiebreak

Match Progression Features:

- serve_number_in_match
- match_stage

Models

- Multinomial Logistic Regression (LR)
- Random Forest (RF)
- Decision Tree (DT)
- Neural Network (NN)
- Used each model to predict serve direction (T, Body, Wide) on each player (10 men, 7 women) on each service side (Deuce, Ad).

Men's Deuce Side Prediction Accuracy

Player	**LR**	**RF**	**DT**	**NN**	**MEAN**
Hubert Hurkacz	0.51	0.51	0.51	0.51	0.51
Daniil Medvedev	0.51	0.49	0.50	0.48	0.50
Jannik Sinner	0.46	0.46	0.45	0.46	0.46
Carlos Alcaraz	0.46	0.42	0.45	0.44	0.44
Novak Djokovic	0.48	0.51	0.47	0.48	0.49
Andrey Rublev	0.55	0.55	0.54	0.52	0.54
Casper Ruud	0.49	0.45	0.48	0.49	0.48
Stefanos Tsitsipas	0.58	0.55	0.58	0.58	**0.57**
Alexander Zverev	0.50	0.51	0.52	0.52	0.48
Botic van de Zandschulp	0.49	0.47	0.47	0.48	0.48
MEAN	**0.50**	0.49	**0.50**	**0.50**	0.50

Women's Deuce Side Prediction Accuracy

Player	**LR**	**RF**	**DT**	**NN**	**MEAN**
Iga Swiatek	0.43	0.42	0.41	0.41	0.42
Bianca Andreescu	0.38	0.33	0.37	0.34	0.36
Elena Rybakina	0.52	0.46	0.50	0.48	**0.49**
Coco Gauff	0.48	0.44	0.50	0.48	0.48
Elina Svitolina	0.43	0.39	0.44	0.41	0.42
Jessica Pegula	0.42	0.38	0.42	0.38	0.40
Mirra Andreeva	0.40	0.39	0.42	0.43	0.42
MEAN	**0.44**	0.40	**0.44**	0.42	0.43

Men's Advantage (Ad) Side Prediction Accuracy

Player	**LR**	**RF**	**DT**	**NN**	**MEAN**
Hubert Hurkacz	0.51	0.50	0.52	0.51	0.50
Daniil Medvedev	0.52	0.51	0.49	0.48	0.50
Jannik Sinner	0.58	0.53	0.59	0.57	**0.57**
Carlos Alcaraz	0.49	0.45	0.49	0.50	0.48
Novak Djokovic	0.52	0.53	0.52	0.51	0.53
Andrey Rublev	0.49	0.50	0.51	0.48	0.50
Casper Ruud	0.51	0.45	0.52	0.51	0.49
Stefanos Tsitsipas	0.50	0.48	0.50	0.51	0.49
Alexander Zverev	0.50	0.55	0.48	0.52	0.52
Botic van de Zandschulp	0.47	0.42	0.47	0.44	0.44
MEAN	**0.51**	0.49	**0.51**	0.50	0.51

Women's Advantage (Ad) Side Prediction Accuracy

Player	**LR**	**RF**	**DT**	**NN**	**MEAN**
Iga Swiatek	0.37	0.36	0.37	0.38	0.37
Bianca Andreescu	0.37	0.33	0.39	0.36	0.36
Elena Rybakina	0.53	0.47	0.53	0.52	**0.50**
Coco Gauff	0.45	0.46	0.46	0.43	0.44
Elina Svitolina	0.46	0.45	0.49	0.46	0.46
Jessica Pegula	0.49	0.43	0.50	0.46	0.47
Mirra Andreeva	0.39	0.36	0.41	0.39	0.38
MEAN	0.44	0.41	**0.45**	0.43	0.43

Concluding Remarks

- All models outperform random chance (33%).
- Men's models: $\sim$ 50% accuracy (best 57%).
- Women's models: $\sim$ 43% accuracy (best 50%).
- No single model dominates:
  - unexplained variation $\rightarrow$ serve kinematics (Jacquier-Bret and George, 2024).

Future Research

- Which contextual factors lead to accurate serve prediction?
- Are taller players' serve more predictable than shorter players?
- New data $\rightarrow$ new findings/confirm old findings?

Sources

- Jeff Sackmann, Tennis Match Charting Project, GitHub.
- Preito-Lage et al. (2023), Match analysis and probability of winning a point in elite men's singles tennis.
- Jacquier-Bret Gorce (2024), Tennis Serve Kinematics Review.