Tennis Server Quality Metric

Aiwen Li, Amrita Balajee, Audrey Kuan

Contents

Motivation

- Existing research focuses extensively on match-level prediction (mostly ELO models)
- Basic serving statistics exist, but they tell an incomplete story of server dominance
- Room in the literature for an evaluation of servers in men’s and women’s tennis

Research Objectives

Group players by serve style: using serve speed, ace%, location entropy, and modal serve location

Model serve outcomes: Fit a model using server statistics to predict point outcomes

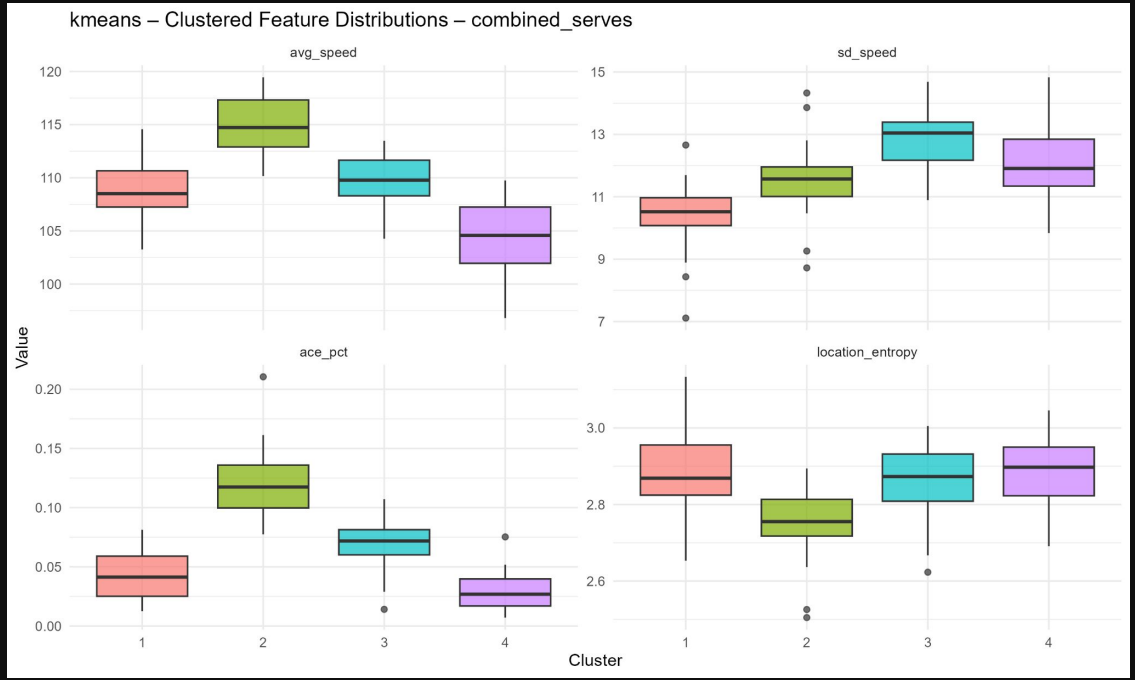
Create a server quality metric: Define a single, interpretable score to evaluate serve performance.

Raw Data

Clustering

Modeling

K-Means Clustering: Speed, Aces, Entropy



	cluster	avg_speed	sd_speed
1	1	108.960916920795	10.4145155732768
2	2	114.815266722397	11.5776122609671
3	3	109.746322772352	12.9167511071142
4	4	104.32923706341	12.0743616614806

	cluster	ace_pct	location_entropy
1	1	0.0433578436481444	2.872178922517
2	2	0.119837475337068	2.75073665561808
3	3	0.0706959470167434	2.86197448159623
4	4	0.0295168777819073	2.8866707644689

Server Quality Metrics

Weighted Average of Server Statistics

- Formula:
  *Σ [(Importance Score) * (player statistic)]*
- Importance scores from random forest model
- Player statistics: speed (average and sd), entropy, modal serve location (one-hot encoded)

Residuals from Prediction Models

- General form of models:
  (*Win_pct* or *serve_efficiency*) ~ (*avg_speed* + *sd_speed* + *location_entropy* + *as.factor(modal_location)*
- 3 metrics from residuals using linear regression, random forest, and XGBoost

Separate metrics using different data (Wimbledon/U.S. Open, males/females)
All metrics scaled to be between [-1, 1]

Outcomes for Prediction Models

Serve Efficiency

- Formula: *(# points won as server where rally count <= 3) / (total # points played as server)*
- How often a serve leads **rapidly** to a point win
- Might undervalue players with good serves who didn't win points quickly

(Point) Win Percentage

- Formula: *(# points won as server) / (total # points played as server)*
- How often a serve leads to a point win, regardless of rally count
- More confounded by other factors (e.g., opponent skill; other hits during rally)

Out-of-sample Predictive Performance

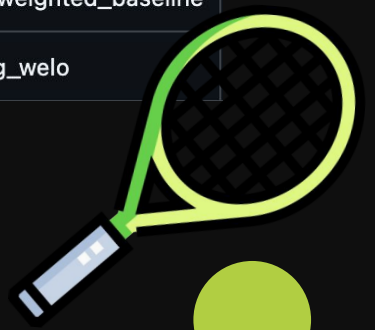
Serve Efficiency

1	cor	p_value	rmse	avg_metric	n	Model
2	0.0658026039090986	0.629920945584615	1.35463301283568	1.33008373660275e-17	56	overperf_lm
3	0.133539290993594	0.326503426907044	1.30459817288357	-1.21430643318376e-17	56	overperf_rf
4	0.172178479388628	0.204474519792002	1.27517762955409	5.4752209710518e-18	56	overperf_xgb
5	0.534139987204592	2.23691723505814e-05	0.956599272418174	-1.6146868068802e-18	56	rf_weighted_baseline
6	0.0214043186327044	0.875575490603796	0.990804165032057	-1.85565751597868e-15	56	avg_welo

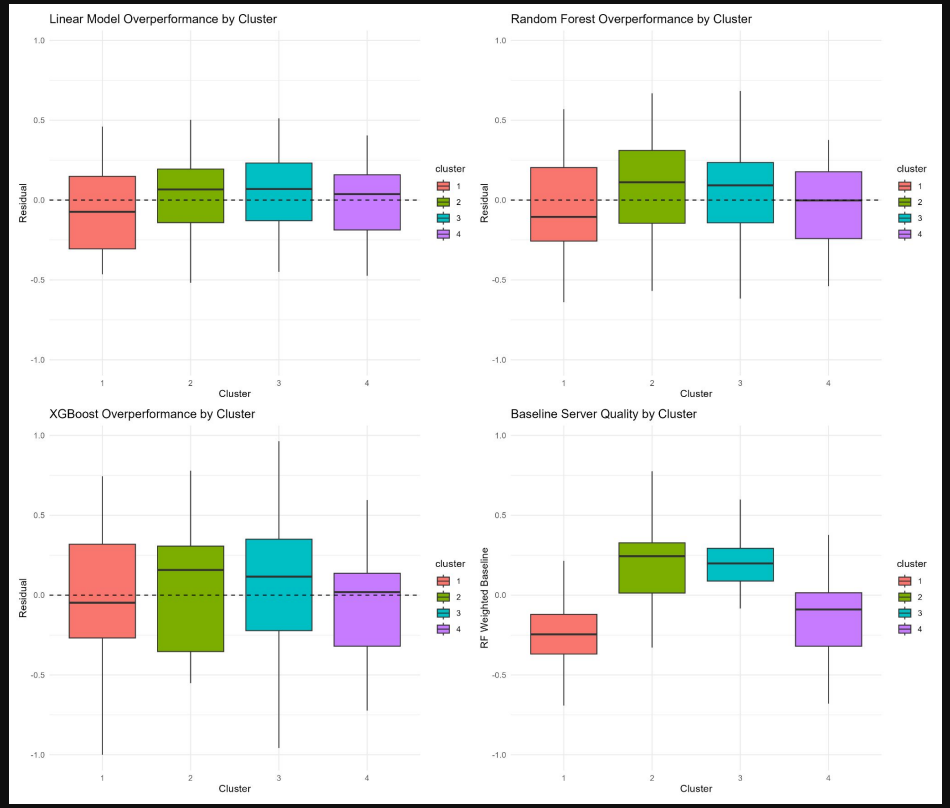
Out-of-sample Predictive Performance

Win Percentage

1	cor	p_value	rmse	avg_metric	n	Model
2	0.232420756872523	0.0847544652815722	1.22790265976483	1.36880524276295e-17	56	overperf_lm
3	0.279088666180536	0.0372521451697804	1.18998984629628	-2.45513709808726e-17	56	overperf_rf
4	0.258426014659132	0.0544710213163756	1.20692302384659	2.94360889829814e-17	56	overperf_xgb
5	0.159330091827412	0.240821390165845	1.28503536568193	1.24625167576593e-17	56	rf_weighted_baseline
6	0.284334647347935	0.0336846431777086	0.950126490239438	-1.85565751597868e-15	56	avg_welo



Distn of Server Quality Per Cluster



Server metrics based on win_pct prediction models

Alexander Zverev

Serve Details

Average Serve Speed
119.448 mph

SD Serve Speed
11.783 mph

Ace Percentage
0.116

Location Entropy
2.833

Modal Serve Location
WC - DNCTL

Cluster
2

Metric Scoring

RF Residuals (better for win percentage)	0.669
Weighted Stats (better for serve efficiency)	0.792

- Placed in “Powerhouse” server cluster
- Known for high speed, powerful serve

Future Work

- Adding serve spin and placement data to the metrics
- Ace percentage as an outcome variable?
- Generate actionable insights for players including personalized training plan for certain serve styles and opponents
- For Betting: server quality metric that updates live throughout a match