Finding the Best Tennis Serves with K-Means and GMM Clusters of Ball Tracking data to Interpret Serve Strategies

Kalil Saldanha Kaliffe[1], Reginaldo Santos[2]

[1] Universidade Federal do Pará, Brazil
kalilsaldanha@gmail.com
[2] Universidade Federal do Pará, Brazil
regicsf2010@gmail.com

Abstract. The serve is a crucial shot in tennis, that dictates a player's advantage. However, there has been a noticeable gap in recent data analysis focused on player behavior during serves, when compared to data analysis adoption in other sports. With high speeds, precision, and small margins, ball-tracking systems like Hawkeye are essential for capturing serve steps with fidelity. This data is crucial for decision-making improvements, performance enhancement, and knowledge discovery. However, the Full Hawkeye data is not publicly available. In this manner, this article uses scraping techniques to harness Hawk-Eye serve tracking data from the Australian Open (2020-2024) and Roland Garros (2019-2024), consisting of 152.761 serves from 951 matches. K-Means and Gaussian Mixture Model (GMM) clustering models were employed to discover clusters that summarize thousands of servers into interpretable serve strategies. The best serve strategies optimize success percentages, risk of missing the serve (fault), and may vary from first to second serves, or be affected by pressure in breakpoints, thus the best serve is a serve that best fits a situation and matches a desired outcome. The relation between the serve success and best players was checked, by correlating the server ranking with cluster success using serves from these clusters in different context scenarios. We discovered that the success rate in the clusters increases with player ranking points in high-pressure situations, such as breakpoints and tiebreaks, also that, the hard courts at the Australian Open have greater success rates, while the slower clay courts at Roland Garros have lower first and second serve success rates, despite using similar serve strategies, and that rankings had little bearing on serve performance on these slower courts, indicating that in this surface, other factors may matter more for player advantage in the end than just winning points with the serve right away.

CCS Concepts: • **Clustering** → **Data Mining**.

Keywords: Ball-Tracking, Clustering, Data Mining, GMM, K-Means, Serve Strategies, Successful Serves, Tennis Serves

1. INTRODUCTION

Data analysis has revolutionized sports by improving decision-making, enhancing performance, and discovering knowledge. This is especially true in sports like tennis, where players' actions generate extensive data that reflect their choice patterns, which results in performance metrics such as success winning points and rankings.

The serve, is a crucial shot in tennis, it not only initiates points but also directly influences a player's advantage [O'Donoghue and Brown 2008], the serve is unique since it is the only time a player has complete control over the ball's toss and strike, this highlights the value of the player's individual choices when adapting against different opponents and game situations.

Tennis Grand Slam tournaments are tracked by the Hawkeye System [Hawk-Eye Innovations 2024], which is a multi-camera optical tracking system used to improve fairness and viewer experience. It provides precise ball-tracking to determine if shots are in or out, handling the high speeds and small margins of professional serves, thus capturing every step of a serve due to its accuracy and reliability [Innovations 2015]. Therefore, this precise ball tracking is fundamental for Tennis Serves data analysis. However, despite its potential for player preparation, performance analysis and strategic planning, the Full Hawkeye data is not publicly available. Thus, tennis remains significantly behind in adopting comprehensive data analytics strategies. Different from other sports like baseball, where data is widely available and adopted [Society for American Baseball Research 2024], [Fangraphs 2024] and has real life impact through data analysis as depicted in the Moneyball story [Lewis 2003].

To bridge this gap, this article employed scraping techniques, and harnessed Hawk-Eye serve tracking data from the Australian Open (2020-2024) and Roland Garros (2019-2024), consisting of 152.761 serves from 951 matches. One of the contributions of this study is the collection of this dataset[1], now easily available for future research. However, this represents only a subset of Hawk-Eye's data, considering tennis was an early adopter of Hawk-Eye technology since 2006 [BBC Sport 2006].

This study aims to find and analyze tennis serve strategies, by clustering serve features such as coordinates, direction, speed, serve types and correlate them with success statistics for each cluster in order to find the best serves. Clustering is fundamental to group similar serves with interpretability, thus finding serve strategies, which we propose to be approximately described by grouping similar combinations of serves, and the degree of this approximation is the model fit. Then, after discovering knowledge from clusters, from the statistical analysis of these clusters we identified which results does these serves strategies lead to and under which context lead to the best serves. To this extent, we also compared how well K-Means and Gaussian Mixture Model (GMM) clustering models represent these serve strategies.

2. RELATED WORK

[Mecheri et al. 2016] validated using Ball tracking data from elite tournaments to examine the impact of serve characteristics on winning probabilities, finding that higher serve speeds and optimized strategies improved winning-point rates, with notable gender differences. [Wei et al. 2016] predicted serve trajectories using Hawk-Eye data, highlighting the usefulness of clustering algorithms for opponent preparation. [Tea and Swartz 2023] investigated serve tactics using Bayesian models, identifying serve pattern differences between genders. [Whiteside and Reid 2016] focused on spatial characteristics of serves contributing to aces but did not consider faults or second serves. To the best of our knowledge, our work is the first to evaluate success with rankings of serve strategies found via clustering, specially using GMM and our scraped broad ball tracking Hawkeye dataset from Australian Open (2020-2024) and Roland Garros (2019-2024).

3. METHODOLOGY

3.1 Data Description

Table I, shows a statistical summary of the preprocessed data used in this study. Top and Bottom serves are from the players sorted by ranking points. Breaks Won are the break points defended by the server over the number of total number of breaks, T, Wide and Body Aces are over the total number of aces, Successful is the number of aces and forced opponent errors when serving over the total number of serves, double faults are over the number of total second serves. Also, out of the 257 ATP men's singles players, 145 played in both tournaments at some point.

Table I. Men's Singles Serve statistics for the preprocessed data. Serves from Australian Open (2020-2024) - 630 matches and Roland Garros (2019-2024) - 321 matches. Total is the sum of either AO and RG Serves, or First and Second Serves.

	Serves	Successful	Aces	Faults	Double Faults	Breaks Won	T	Wide	Body	T Ace	Wide Ace	Body Ace	Avg Spd	Max Spd	Min Spd
Total	152761	19.89%	6.27%	15.29%	4.26%	55.73%	41.83%	42.26%	15.91%	53.27%	46.59%	0.15%	176.22 KPH	251 KPH	97 KPH
Roland Garros	40318	15.63%	3.59%	17.51%	3.95%	52.21%	42.89%	43.04%	14.07%	56.04%	43.75%	0.21%	174.79 KPH	251 KPH	97 KPH
Australian Open	112443	21.41%	7.22%	14.50%	4.36%	57.21%	41.45%	41.98%	16.56%	52.77%	47.09%	0.14%	176.74 KPH	243 KPH	97 KPH
First Serves	106468	25.33%	8.72%	20.09%	0%	57.53%	44.50%	47.48%	8.02%	53.24%	46.64%	0.12%	186.75 KPH	243 KPH	97 KPH
Second Serves	46293	7.37%	0.61%	4.26%	4.26%	51.67%	35.71%	30.25%	34.04%	54.06%	44.88%	1.06%	152.01 KPH	251 KPH	98 KPH
Top 10 serves	28135	20.57%	6.27%	15.22%	4.33%	58.59%	43.96%	43.05%	12.99%	55.19%	44.81%	0%	179.90 KPH	231.18 KPH	98 KPH
Top 30 serves	57710	20.96%	6.67%	15.42%	4.13%	58.76%	42.55%	43.10%	14.35%	51.98%	47.95%	0.08%	179.15 KPH	239.29 KPH	98 KPH
Bottom 30 Serves	4264	18.01%	5.25%	16.79%	4.86%	50.84%	41.96%	40.95%	17.10%	50.89%	48.66%	0.45%	172.67 KPH	220.84 KPH	97 KPH

[1]Scraped dataset available at: https://github.com/hawkilol/tennisSlamBallTracking

3.2 Tools

Clustering analysis was conducted using Python's scikit-learn library, with data manipulation handled by Pandas, visualization was performed using Matplotlib and Plotly, while numerical computations were executed with NumPy, and scraping with Requests and Selenium Webdriver.

3.3 Data Collection

The Hawkeye data is not publicly distributed via standardized API, to our knowledge, the only available data is from a web interface that samples from the Hawkeye system, called CourtVision, available for some tournaments managed by InfoSys [Infosys 2024] such as Roland Garros and Australian Open.

Ranking history of the players and Ids were scraped from the official ATP (Association of Tennis Professionals) website [Association of Tennis Professionals 2024]. ATP rankings are determined by a points system based on a player's performance in tournaments over the previous 52 weeks. Players earn points by advancing through rounds in tournaments, with the most points awarded in Grand Slams such as Australian Open and Roland Garros. Points drop off after 52 weeks, meaning players must consistently perform well to maintain or improve their ranking [ATP Tour].

3.3.1 *CourtVision Scraping.* Although, this data can be visualized in the CourtVision interface, the API response JSON is obscured into an illegible AES cipher code [Dworkin et al. 2001]. However, since this response is processed on the client-side, we investigated the client JavaScript revealing how it decrypts API responses: by extracting a decryption key from predictable simple transformations made to the response timestamp. This process allows access to the original JSON data.

3.4 Data Processing and Cleaning

When player points and ranking changes, the ranking history gets updated along with a new date entry. To measure a player's strength per serves in a match, each match was assigned the ranking and points for both players based on ranking at the date of the match.

We filtered out, serves without available ball tracking, serves with missing any features from Section 3.5.1, as well as serves with ambiguous information, such as those marked both as a fault and an ace, or with implausible coordinates. This study initially collected 311.865 serves, after cleaning we left with 152.761 as seen in Table I, it goes to show how not having the original data affects the data quality. Roland Garros has fewer serves and matches despite the extra year, since only matches and serves with tracking available data were selected, and the full adoption of Hawkeye may not be essential for fair play on the clay courts of Roland Garros, which naturally leave clear ball marks, when the hard court at the Australian Open do not, with functionality in mind it would be more of a luxury for the main courts. The preprocessed and cleaned data is shown in Table I.

To ensure proper clustering, the coordinates were mirrored, into a one AdCourt and DeuceCourt, we can visualize this new court in Figures 3 and 4. Also, the categorical features, court side, serve direction and serve type, were encoded into numbers. Then after fitting the model, to improve cluster interpretability, we rounded these back to integers and inverse transformed them back to the original string categorical resulting in the legends of cluster centers in Figures 3 and 4.

3.5 Clustering

3.5.1 *Cluster features.* We choose columns that reflect player input, plus the court side which is different from other contextual features, because when it changes, other serve features change, since depending on the side, the players are required to land the serves in spatially different sides. Unlike the clustering done by [Wei et al. 2016] which employed two different models for each side, we choose to let the model separate the clusters based on court side feature, to evaluate how K-Means and GMM manage the separation, it also allows comparison between serves from different court sides.

Serve Speed: Serve Speed in Kilometers per Hour.
Serve Direction: T (inner edge, between the two serve boxes), Body (middle of the serve box), Wide (outer edges).
Serve Type: 'Flat', 'Pronated', 'Slice', 'Kick', changes based on the spin applied to the ball.
Ball Serve Impact (x, y, z): Location coordinates where the serve ball was hit by the server.
Net Serve (x, y, z): Coordinates of serve ball as it reached the net.
Serve Bounce (x, y, z): Coordinates of serve bounce on the ground.
Court Side: The two serve boxes, to serve at Deuce Court on even points or Ad Court on odd points.

3.5.2 *Outcome and Context.* These columns are used to provide meaning to the clusters, and are not directly controlled by the players, only correlated.

Serve Number: 1 (First Serve) or 2 (Second Serve).
Break Point: If the server loses the point they lose the game, if they win they continue serving.
Ace: If the serve won the point without the returner touching the ball.
Success: If the serve won the point by making returner miss (Forced Error).
Fault: A miss serve, out or stayed on the net, moves to the second serve if on the first serve.
Double Fault: If the server faults on the second serve, losing the point.

3.5.3 *K-Means Clustering.* Logical first choice due to simplicity. We used the Elbow method with inertia that represents the within-cluster sum of squares (WCSS) against the number of clusters to estimate the optimal number K of clusters, for a balance between fit to the data and model complexity. This way, the optimal number of clusters is identified where the rate of decrease in inertia slows down.

3.5.4 *Gaussian Mixture Model (GMM).* Probabilistic nature allows capturing more complex non-linear patterns and different cluster shapes with full covariance, also better grouping serves with overlapping features, Ex. same court side and direction but the speed, serve type and actual coordinates can be different. We define an interpretable model as one with a balance between parameter count and cluster quality, to find this balance, we used the Akaike Information Criterion (AIC) [Akaike 1974] and Bayesian Information Criterion (BIC) [Schwarz 1978] scores. AIC estimates the relative amount of information lost by a model, while BIC introduces a stricter penalty for the number of parameters in the model. As seen in Figures 1 and 2, to further access the cluster quality and consistency we also used the Silhouette [Rousseeuw 1987], it scores how similar an object is to its own cluster compared to other clusters, to estimate the K for the experiment.

3.6 Cluster Regression

Serve success of clusters is calculated by first filtering the context, serve numbers and breakpoints, then calculating mean success and rank points for each cluster, court side, and serve number. To avoid skewing, 0% success rates are removed from the plots. We used the R^2 statistic to assess the proportion of variance in serve success explained by the model (ranking), higher R^2 indicates better fit. If the null hypothesis is true, the associated P-value for the F-test shows the likelihood of witnessing the data. So a low P-value indicates that at least one predictor is substantially influencing the model's capacity to explain the variance in the dependent variable, indicating that the model has explanatory power, as demonstrated by [Aiken et al. 1991]. In this manner, we searched for relations that can be statistically explained by the ranking.

4. RESULTS

This section presents the evaluation of the optimal number of clusters for K-Means and GMM, and court visualizations for the found cluster centers in Section 4.1. The statistical summary of each cluster in general contexts in Section 4.2, and pressure contexts experiments with the ranking influence in Section 4.3.

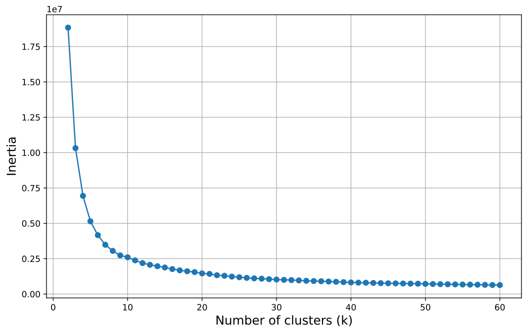

Fig. 1. Elbow Method for K-means Clustering.

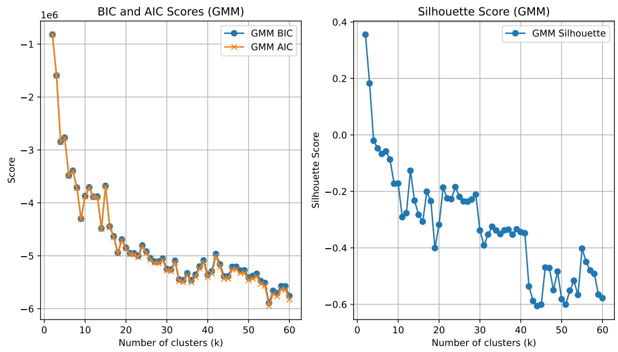

Fig. 2. AIC, BIC and silhouette scores for GMM Clustering and fit time.

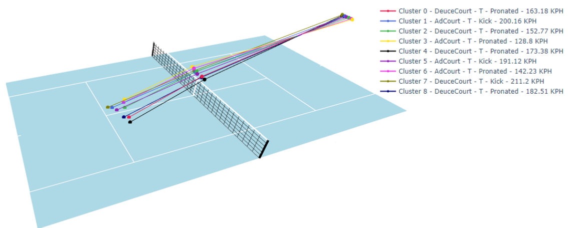

Fig. 3. Cluster Centers on Court K-Means.

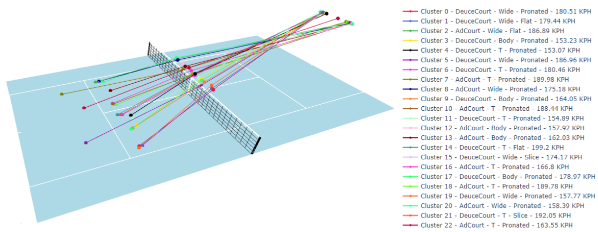

Fig. 4. Cluster Centers on Court GMM.

4.1 Optimal K evaluation and K-Means x GMM Clusters

The number of K clusters for K-Means was set at 9 based on the Elbow method's inflection point, while for GMM, 23 components were chosen based on the AIC and BIC, following the methodology outlined in Sections 3.5.3 and 3.5.4, respectively. As seen in Figures 1 and 2, the complexity increased before the AIC and BIC scores could greatly penalize them, thus we identified the range of components K where the scores began to marginally decrease, and AIC scores gradually gets a bit smaller than BIC indicating possible overfit, among this range of low AIC and BIC scores we evaluated the optimal Silhouette.

The Cluster Centers for K-Means and GMM seen in Figures 3 and 4 show a fundamental difference between the means of the models, the means for K-Means got skewed towards the center T of the serve boxes because K-Means tends to balance out the cluster centers to minimize the overall Euclidean distance variance with equal shaped spherical clusters, Table I confirms that the overall the central T region has more data points. Meanwhile, the GMM cluster centers are more distributed, our assumption that the underlying distribution is overlapping is reflected in the GMM clustering better representing the serve strategies

Tables II, III, show that the K-Means Clusters have less variation between first and second serves, the order of the 6 clusters with the highest ace percentages stay the same for both first and second serves, while the GMM clusters have more percentage and order variety, for instance while first serves with the most aces are from Cluster 10 (AdCourt, T-Pronated, Higher than average speed) with 14.12% ace, on second serves it is only the 4th best with 4.40% ace. This further contributes that the GMM clusters captured better the underlying distribution of serves, while the K-Means clusters with less variation provided less knowledge discovery. Additionally, as demonstrated by the correlation matrices in Figures 5 and 6, K-Means shows stronger correlations to the Cluster column, indicating that it oversimplifies and generalizes the strategies, while GMM more subtle correlations indicates that each cluster captured the complex relationship, before overfitting as seen in Figure 2.

Table II. K-Means Cluster Centers Statistics.

Cluster	Court Side	1st Serves	2nd Serves	1st Success	2nd Success	1st Ace	2nd Ace	1st Fault	2nd Fault	1st BP Won	2nd BP Won	1st Avg Spd	1st Max Spd	1st Min Spd	2nd Avg Spd	2nd Max Spd	2nd Min Spd
0	Deuce	8414	9300	22.58%	9.65%	5.19%	0.76%	10.24%	5.82%	56.11%	55.09%	163.92 KPH	168.57 KPH	157.43 KPH	162.50 KPH	168.07 KPH	157.72 KPH
1	Ad	21982	184	27.50%	33.70%	11.40%	11.96%	26.94%	11.96%	56.39%	66.67%	200.17 KPH	205.71 KPH	195.38 KPH	199.72 KPH	205 KPH	196 KPH
2	Deuce	3427	13753	16.46%	6.10%	2.28%	0.23%	9.80%	4.12%	56.77%	49.38%	153.54 KPH	158.49 KPH	147.21 KPH	152.57 KPH	158.44 KPH	147.72 KPH
3	Ad	478	5154	14.85%	2.23%	2.72%	0.08%	8.58%	1.79%	42.50%	51.27%	124.94 KPH	135.15 KPH	97 KPH	129.15 KPH	135.18 KPH	98 KPH
4	Deuce	15295	3570	24.71%	16.72%	6.59%	1.43%	12.36%	7.39%	60.22%	55.13%	173.59 KPH	178.24 KPH	168 KPH	172.46 KPH	178 KPH	168.13 KPH
5	Ad	23556	412	27.02%	28.40%	9.36%	7.04%	17.85%	10.92%	60.66%	57.58%	191.14 KPH	195.94 KPH	186.22 KPH	190.29 KPH	195 KPH	187 KPH
6	Ad	1347	12665	12.99%	3.97%	1.26%	0.10%	10.17%	2.44%	48.92%	50.71%	142.70 KPH	147.52 KPH	135.64 KPH	142.18 KPH	147.56 KPH	135.71 KPH
7	Deuce	11561	95	25.39%	40%	12.72%	17.89%	43.09%	25.26%	50.44%	57.14%	211.19 KPH	243 KPH	205.58 KPH	212.48 KPH	251 KPH	206 KPH
8	Deuce	20408	1160	25.17%	20.86%	7.61%	3.79%	14.77%	9.14%	59.19%	59.26%	182.58 KPH	187.48 KPH	177.47 KPH	181.29 KPH	187 KPH	177.85 KPH

Table III. GMM Cluster Centers Statistics.

Cluster	Court Side	1st Serves	2nd Serves	1st Success	2nd Success	1st Ace	2nd Ace	1st Fault	2nd Fault	1st BP Won	2nd BP Won	1st Avg Spd	1st Max Spd	1st Min Spd	2nd Avg Spd	2nd Max Spd	2nd Min Spd
0	Deuce	2428	494	27.68%	11.94%	10.54%	3.24%	22.94%	4.86%	50%	54.55%	186.74 KPH	228.19 KPH	126 KPH	150.08 KPH	201 KPH	115 KPH
1	Deuce	3185	666	23.08%	8.11%	6.19%	0.45%	17.52%	8.41%	55.88%	51.72%	184.19 KPH	227.89 KPH	141 KPH	156.66 KPH	209 KPH	117 KPH
2	Ad	13142	1918	25.92%	6.26%	10.09%	0.57%	24.26%	7.35%	59.26%	45.26%	192.68 KPH	232.60 KPH	118 KPH	147.02 KPH	213 KPH	101 KPH
3	Deuce	184	5037	9.78%	4.55%	0%	0.02%	5.98%	2.54%	28.57%	48.37%	161.85 KPH	197.48 KPH	115 KPH	152.72 KPH	182 KPH	106 KPH
4	Deuce	594	9147	12.63%	5.79%	0.17%	0.11%	8.08%	2.51%	73.33%	50.29%	165.68 KPH	192.33 KPH	108 KPH	152.15 KPH	202 KPH	98 KPH
5	Deuce	989	121	0.40%	0%	0%	0%	98.58%	99.17%	20.51%	0%	190.68 KPH	231.32 KPH	132 KPH	156.90 KPH	192 KPH	131 KPH
6	Deuce	5371	1497	24.54%	5.68%	7.95%	0.40%	16.72%	2.61%	59.79%	52.63%	189.65 KPH	243 KPH	121 KPH	147.64 KPH	223.45 KPH	108 KPH
7	Ad	1075	113	0.37%	0%	0%	0%	98.23%	98.23%	18.33%	0%	193.18 KPH	232.76 KPH	130 KPH	157.71 KPH	196 KPH	115 KPH
8	Ad	5660	1137	25.58%	8.88%	8.20%	0.70%	21.87%	5.45%	56.50%	49.18%	180.25 KPH	229.43 KPH	115 KPH	149.98 KPH	211 KPH	109 KPH
9	Deuce	727	783	16.64%	4.34%	0.14%	0%	16.78%	3.96%	45.83%	59.09%	181.92 KPH	233.27 KPH	102 KPH	147.43 KPH	186 KPH	107 KPH
10	Ad	2889	532	29.73%	13.53%	14.12%	4.14%	18.83%	5.64%	58.99%	65.62%	193.28 KPH	239.29 KPH	125 KPH	162.29 KPH	231 KPH	124 KPH
11	Deuce	195	325	19.49%	6.15%	2.05%	0%	2.05%	1.23%	33.33%	10%	170.69 KPH	191.46 KPH	131 KPH	144.88 KPH	183 KPH	103 KPH
12	Ad	3182	7431	15.49%	5.06%	0.09%	0.03%	10.25%	2.10%	53.40%	51.92%	184.12 KPH	230 KPH	106 KPH	146.70 KPH	221 KPH	98 KPH
13	Ad	812	1110	15.76%	5.41%	0.12%	0%	9.85%	1.80%	57.14%	54.55%	178.31 KPH	228 KPH	125 KPH	150.11 KPH	213 KPH	104 KPH
14	Deuce	15662	198	27.71%	21.21%	12.85%	6.06%	22.37%	9.60%	55.79%	87.50%	199.38 KPH	235 KPH	129 KPH	185.47 KPH	223 KPH	154 KPH
15	Deuce	18265	2953	30.23%	15.27%	9.92%	2.00%	14.41%	3.83%	64.77%	66.09%	176.11 KPH	213 KPH	135 KPH	161.68 KPH	199 KPH	131 KPH
16	Ad	2345	3578	12.58%	7.83%	3.07%	1.03%	44.18%	3.75%	44.15%	51.01%	181.28 KPH	229.37 KPH	104 KPH	154.95 KPH	208 KPH	99 KPH
17	Deuce	3208	1216	16.02%	9.38%	0.03%	0%	15.59%	6.74%	46.30%	46.94%	186.20 KPH	230 KPH	109 KPH	164.46 KPH	211 KPH	101 KPH
18	Ad	15384	970	28.14%	21.65%	10.18%	4.12%	12.47%	4.43%	60.85%	61.46%	190.59 KPH	237.17 KPH	154 KPH	177.68 KPH	228 KPH	154 KPH
19	Deuce	1013	1165	11.35%	7.12%	1.78%	1.12%	55.58%	8.93%	39.39%	55.77%	166.11 KPH	230 KPH	99 KPH	147.50 KPH	200 KPH	99 KPH
20	Ad	5484	5511	23.74%	7.35%	4.69%	0.31%	12.87%	4.59%	59.53%	52.43%	171.57 KPH	235 KPH	112 KPH	145.28 KPH	208 KPH	99 KPH
21	Deuce	4242	281	28.38%	28.47%	10.56%	9.25%	14.73%	7.83%	58.60%	37.50%	192.64 KPH	229.75 KPH	142 KPH	185.31 KPH	219 KPH	158 KPH
22	Ad	432	110	7.18%	4.55%	3.01%	0%	69.44%	43.64%	31.11%	15.79%	167.49 KPH	232 KPH	97 KPH	147.70 KPH	251 KPH	99.94 KPH

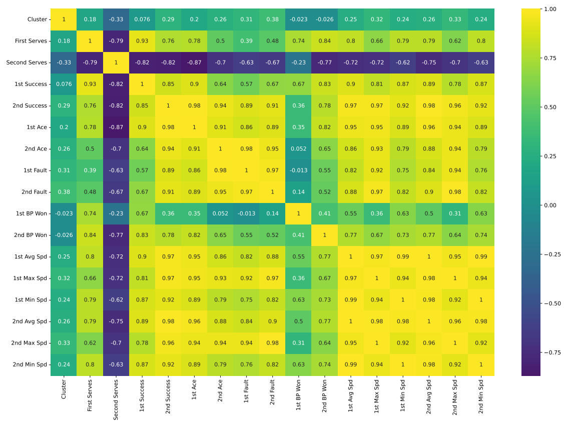

Fig. 5. Correlation Matrix of K-Means Cluster Statistics.

Fig. 6. Correlation Matrix of GMM Cluster Statistics.

4.2 Global context situations

The statistics from the GMM means in Table III illustrate how the players manage the trade-off between risk x reward when the context changes, for instance between first and second serves, the statistics show that first serves, as seen by clusters like Cluster 10 (AdCourt, T-Pronated) and Cluster 14 (DeuceCourt, T-Pronated), frequently favor high success rates with acceptable fault rates. These clusters indicate a more aggressive strategy looking for opponent forced errors and aces, with success rates of about 25% to 30% with a moderate fault percentage. Meanwhile, in second serves is seen more balanced clusters to reduce errors, while maintaining median success rates, such as Cluster 17 (DeuceCourt, T-Pronated) and Cluster 21 (AdCourt, T-Slice) which are also further away from the max global serve speed. These occurrences and correlations from Figure 6 may indicate that the high success of serves are not only accompanied by the serve speed, but also by the serve type and location.

4.3 Ranking Influence

Given the importance of serves in winning matches [O'Donoghue and Brown 2008], and that the players' ability to win matches reflect on their rankings, we assumed that higher ranked players are

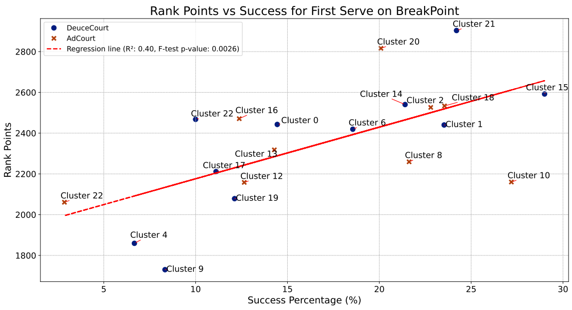

Fig. 7. Cluster Scatter Rank Point x Serve success (Aces or Forces Opponent Errors) % for First Serves per Court Side on Break Points, R^2: 0.40, p-value: 0.0026.

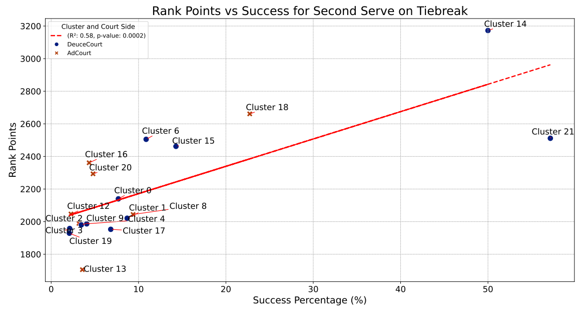

Fig. 8. Cluster Scatter Rank Point x Serve success (Aces or Forces Opponent Errors) % for First Serves per Court Side on TieBreaks, R^2: 0.58, p-value: 0.0002.

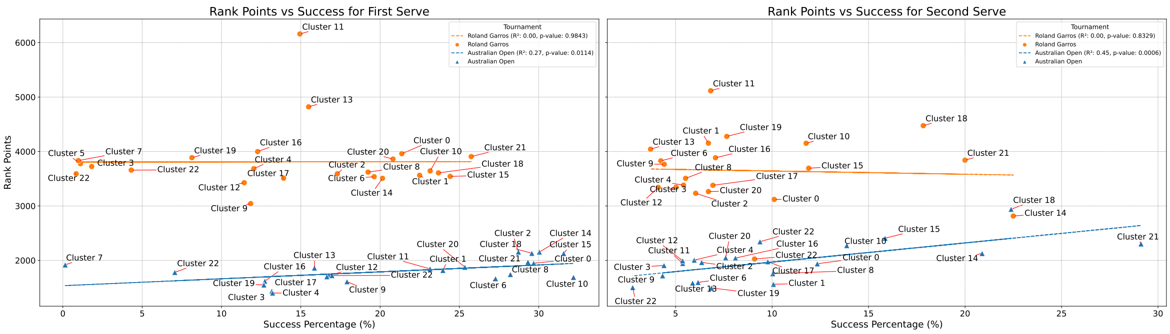

Fig. 9. Cluster Scatter Rank Point x Serve success (Aces or Forces Opponent Errors) % for First Serves per Court Side on Australian Open x Roland Garros.

better serves than lower ranked players. In this manner, the rank of the players can describe success of cluster strategy choices, so under a certain condition, the best serves should also follow what the best players are choosing. Therefore, in Figures 7 and 8 we visualized the relation between the serve strategies found by GMM clustering and the server's ranking points, we checked that higher ranked players are producing serves in the clusters with higher percentage of success than the lower ranked players in high-pressure situations. Specifically in Breakpoints (Figure 7) with R^2: 40% of the variance in the rank points and P-value: 0.0026, and in Tiebreaks (Figure 8) with R^2: 58% of the variance in the rank points and P-value: 0.0002, we can consider these relations statistically significant.

4.3.1 *Tournament Surface Influence.* As observed in Figure 9 there are notable differences between tournaments with different surfaces, in the Australian Open hard courts, the ranking explains the increase in success rates, while at the slower clay courts in Roland Garros the overall success of first and second serves are lower, and the ranking is not significant in predicting the serve success there, while in the Australian Open it is, highlighting how in slower surfaces other aspects of the players game may be more important in relating to performance than winning the point right away with the serve, since when looking at general serve speed from Table I, and cluster speeds from Table III, the serves at Roland Garros are not served at lower speeds, in fact, the top success clusters in both are similar, thus have similar speed, this suggests that the notion of the surface shifts success is relevant.

5. CONCLUSION

This study analyzed tennis serve strategies using ball tracking data from the multi-camera Hawkeye system to find serve strategies clusters: generalized K-Means cluster centers and GMM clusters that captured the overlapping nature of our tennis serve features, then we evaluated the found cluster strategies and analyzed these strategies to find the best serves supported by game context and outcomes. We found that in high-pressure contexts, breakpoints and tiebreaks, the success rate in the clusters increases with the player ranking points. The Australian Open's hard courts show higher success rates due to ranking, while Roland Garros' slower clay courts show lower first and second serve success. In this slower courts, rankings are not significant in predicting serve success, suggesting other aspects of the game may be more relevant. As a result of our study, we hope to demonstrate the benefits of open data, encouraging stakeholders to make the valuable Hawkeye resource more accessible to ultimately bridge the gap in data accessibility and utilization in tennis.

5.1 Future Work and Limitations

Future work would benefit from overcoming the limitations of this work: we did not have access to the full Hawkeye dataset. The unrestricted access to Hawkeye can greatly improve the quality, granularity, and amount of the data for a more detailed analysis while avoiding missing data due to redundancy, it may also reduce biases, particularly due to missing faulty serves. Future work could also benefit from using data not exclusive to serves, serve return strategies and rally shot strategies. Explore women's, doubles data and different tournaments. Furthermore, assigning average points based on observed tournament periods, may not account for changes in player performance over the course of a game, using previous points played in a match as pre-serve context could improve success analysis of serves and players. Additionally, future cluster validation with other indexes should be considered to enhance cluster selection accuracy, along with, state-of-the-art clustering methods like HDBSCAN, which provides a hierarchical structure of clusters along with outlier detection, should be explored, as they could potentially capture more nuanced player patterns, especially due to overlapping features.

REFERENCES

Aiken, L. S., West, S. G., and Reno, R. R. *Multiple regression: Testing and interpreting interactions.* sage, 1991.

Akaike, H. A new look at the statistical model identification. *IEEE transactions on automatic control* 19 (6): 716–723, 1974.

Association of Tennis Professionals. Atp tour singles rankings, 2024.

ATP Tour. ATP Rankings FAQ. https://www.atptour.com/en/rankings/rankings-faq.

BBC Sport. Hingis praises mauresmo display, 2006.

Dworkin, M., Barker, E., Nechvatal, J., Foti, J., Bassham, L., Roback, E., and Dray, J. Advanced encryption standard (aes). Federal Inf. Process. Stds. (NIST FIPS), 2001. [online].

Fangraphs. Fangraphs. https://www.fangraphs.com/, 2024.

Hawk-Eye Innovations. Hawk-eye innovations, 2024.

Infosys. Infosys atp tennis platform, 2024.

Innovations, H.-E. Hawk-eye's accuracy reliability line calling. *Hawk-Eye Innovations: Basingstoke, UK*, 2015.

Lewis, M. *Moneyball: The Art of Winning an Unfair Game.* W. W. Norton & Company, 2003.

Mecheri, S., Rioult, F., Mantel, B., Kauffmann, F., and Benguigui, N. The serve impact in tennis: First large-scale study of big hawk-eye data. *Journal of Sports Sciences* 34 (1): 29–37, 2016.

O'Donoghue, G. P. and Brown, E. The importance of service in grand slam singles tennis. *International Journal of Performance Analysis in Sport* 8 (3): 70–78, 2008.

Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* vol. 20, pp. 53–65, 1987.

Schwarz, G. E. Estimating the dimension of a model. *The annals of statistics*, 1978.

Society for American Baseball Research. The research collection. sabr.org/the-research-collection/, 2024.

Tea, P. and Swartz, T. B. The analysis of serve decisions in tennis using bayesian hierarchical models. *Annals of Operations Research* vol. 325, pp. 633–648, 2023.

Wei, X., Lucey, P., Morgan, S., Carr, P., and Matthews, I. Predicting serves in tennis using style priors. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

Whiteside, D. and Reid, M. Spatial characteristics of professional tennis serves with implications for serving aces: A machine learning approach. *Journal of Sports Sciences* 34 (22): 2073–2081, 2016.