Thesis Proposals at the Chair of Sports Analytics @Saarland University

Chair of Sports Analytics, Saarland University

April 7, 2026

1 Match Charting Project Reliability on Tennis (ID: Match_Chart_Reliability)

Detailed point-by-point data from professional tennis matches can support many research goals, including deriving technical insights [4], modeling in-match win probabilities [2, 5], and investigating psychological effects on performance [3, 1]. This type of data is available on Jeff Sackmann's website[1] and GitHub repository[2], which currently provide point-by-point data for more than $10k$ matches.

What makes this dataset particularly interesting is not only its size, but also how it is created. The data is crowd-sourced through Sackmann's platform, where tennis enthusiasts can contribute by watching match recordings and annotating them according to a standardized coding system, provided in an Excel file on GitHub.[3] In brief, each point is annotated with the current match state (point, game, and set), the server, and the point winner. In addition, rallies are coded using a categorical framework that captures shot type (e.g., serve, forehand, backhand, volley), shot direction/location (e.g., T, wide, cross, long), and point-ending outcome (e.g., winner, double fault, forced error, unforced error).

Although, this crowd-sourcing approach is a cost-effective way to build a large public tennis data base, the reliability of non-expert annotations remains an open question. Therefore, the task in this course is to use Jeff Sackmann's annotation system to source recorded tennis matches available from open sources (e.g. youtube) and evaluate the reliability of the resulting data using statistical methods.

Supervisor: Luis Holzhauer

Scope: Collaborative Bachelor Thesis (2 Students)

[1] https://www.tennisabstract.com/charting/meta.html

[2] https://github.com/JeffSackmann/tennis_MatchChartingProject

[3] https://github.com/JeffSackmann/tennis_MatchChartingProject/blob/master/MatchChart%200.3.2.xlsm

References

[1] Julio González-Díaz, Olivier Gossner, and Brian W. Rogers. "Performing best when it matters most: Evidence from professional tennis". In: *Journal of Economic Behavior & Organization* 84.3 (2012), pp. 767–781. DOI: 10.1016/j.jebo.2012.09.021. (Visited on 05/30/2025) (cit. on p. 1).

[2] Luis K. H. Holzhauer and Pascal Bauer. "Game, set, match–A finite stochastic Markov chain approach for in-match win predictions in professional tennis". In: *Proceeding of the 15th International Symposium of Computer Science in Sport*. Accepted. Tokyo: Springer, 2026 (cit. on p. 1).

[3] Stephanie Kovalchik and Machar Reid. "Measuring clutch performance in professional tennis". en. In: *Statistica Applicata - Italian Journal of Applied Statistics* 2 (2018), pp. 255–268. ISSN: 2038-5587. DOI: 10.26398/IJAS.0030-011. (Visited on 12/23/2024) (cit. on p. 1).

[4] A. James O'Malley. "Probability formulas and statistical analysis in tennis". en. In: *Journal of Quantitative Analysis in Sports* 4.2 (2008). ISSN: 1559-0410. DOI: 10.2202/1559-0410.1100. (Visited on 06/24/2025) (cit. on p. 1).

[5] Min Kyu Sim and Dong Gu Choi. "The winning probability of a game and the importance of points in tennis matches". In: *Research Quarterly for Exercise and Sport* 91.3 (2020), pp. 361–372. ISSN: 0270-1367. DOI: 10.1080/02701367.2019.1666203. (Visited on 01/28/2025) (cit. on p. 1).