Live Counter-Factual Analysis in Women's Tennis using Automatic Key-Moment Detection

Conference Paper · March 2022

Robert Seidl
Technische Universität München

Live Counter-Factual Analysis in Women's Tennis using Automatic Key-Moment Detection

Robert Seidl & Patrick Lucey | Stats Perform

1. Introduction

A recent trend in machine learning is to utilize interpretable techniques such as counter-factual analysis to explain predictions of individual events. Such techniques are powerful in sports where it can be used to answer the impact of a play or event on the overall outcome of the match by framing it as a "what-if" question, where you conduct two predictions - the first based on one outcome, and the second on the alternative outcome – and then you compare the difference in prediction output. Even though this technique has only recently gained traction in the ML field to make predictions more interpretable, it actually has been used extensively in sports over the last decade or so, with the "4th Down Bot" which was based on Brian Burke's initial work in American football being the prime example [1, 2].

Such counter-factual analysis lends itself nicely to other sports, such as tennis, where key moments like a break point occur often and can often decide the winner of the match. However, a limitation of such an approach is relying on a pre-defined notion of a key-moment ***only*** being a break point or game, set or match point – it could include other points, and these points depend on the relative strengths of the players as well as what has occurred already during the match (e.g., if a player is much stronger than another player, an early break point to the weaker player may not be important.)

In this paper, we present a counter-factual method for women's tennis that first automatically highlights the key moments in a match using our "leverage", "clutch" and "momentum" metrics which are created by chaining counter-factual predictions which capture the importance of a point contributing to a player winning the set and/or match, or the likelihood of a comeback. Not only can our approach highlight important moments before they occur in an automatic fashion (and not based on a pre-canned notion of what are important points such as break points), it can also link player behaviors at a season level which shines a light on their tendencies in key moments.

An example of our approach is shown in Figure 1, where we use an example of the match between Yafan Wang vs Donna Vekic from the semi-final of the 2019 WTA 250 Acapulco Tournament. Seeded third in the tournament, bookmakers see Vekic as 2:1 favorite to advance to the final. In the match that Wang eventually won in three-sets (2-6, 6-3, 6-1), after dropping the first set and a close second set until 3-3, when Wang gained the momentum and won six games in a row. There were many key moments during the match that do not coincide with break points. The first one we highlight (A), was 2-2 in the first set with Vekic serving at 30-15 against Wang where we flagged a potential "momentum" swing. Later in the set with the game-score at 3-2 with Wang serving at 15-30 we flagged a "clutch point" (B). Similarly, our counter-factual approach flagged a key "comeback" moment (C) before it occurred late in the second set which triggered Wang's resurgence to win the match. Additionally, our approach can summarize the flow of the match at the end of games and sets (D), as well as highlight player's performance in key moments relative to other players at the season level (E).

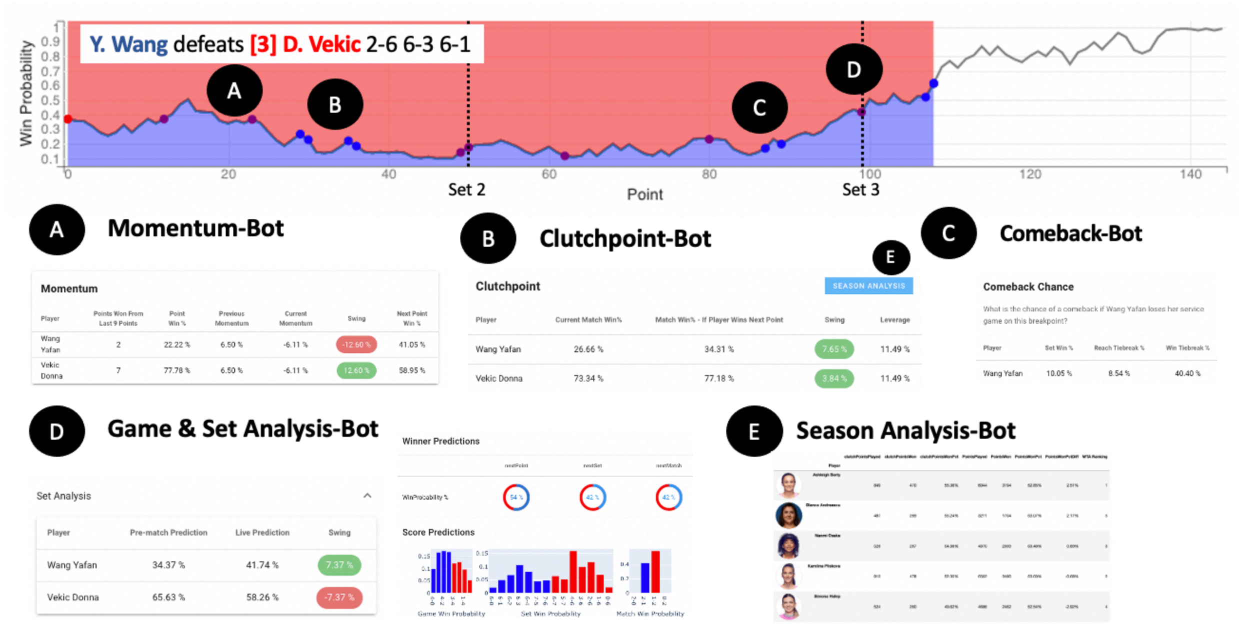


Figure 1: The game between Wang and Vekic highlights the utility of our counter-factual prediction approach to not only highlight the key moments ***before*** they occur, but also gives quantifiable measures of the importance and addresses important "what-if" examples. We highlight our (A) Momentum, (B) Clutch, (C) Comeback, (D) Game & Set Analysis, as well as our (E) Season Analysis. For full impact of our approach, please view the video at: link (password is: Sloan2022)).

To enable our live counter-factual prediction framework, we trained a chained ensemble model to predict not only discrete outcomes at various scales (i.e., next point, game, set, match) but also "alternative outcomes". We did this by arranging binary classifiers into a chain. Each model makes a prediction in the order specified by the chain using all features plus the prediction of the earlier model in the chain. A illustration of our approach is shown in Figure 2. Our dataset contains 1.5 Million points from the WTA Tour between 2012 and 2021. Our model picks the correct next point winner 55.8% of the time which is in line with current state-of-the-art [3-6], but our powerful approach gives us the optionality of measuring these alternative outcomes at a granular level which has not been done previously.

In the next section (Section 2), we describe how we created and evaluated the model. In Section 3, we specify how our "momentum" and "clutch" metrics were created from our counter-factual predictions and how they were validated. In Section 4, we show how our metrics can be used to frame individual player behaviors and tendencies in a season context. In Section 5, we summarize and highlight next potential steps of the work. In addition to the submitted paper, we also have a video showing our approach which can be accessed here - link (password is: Sloan2022).

2. Counter-Factual Analysis Framework

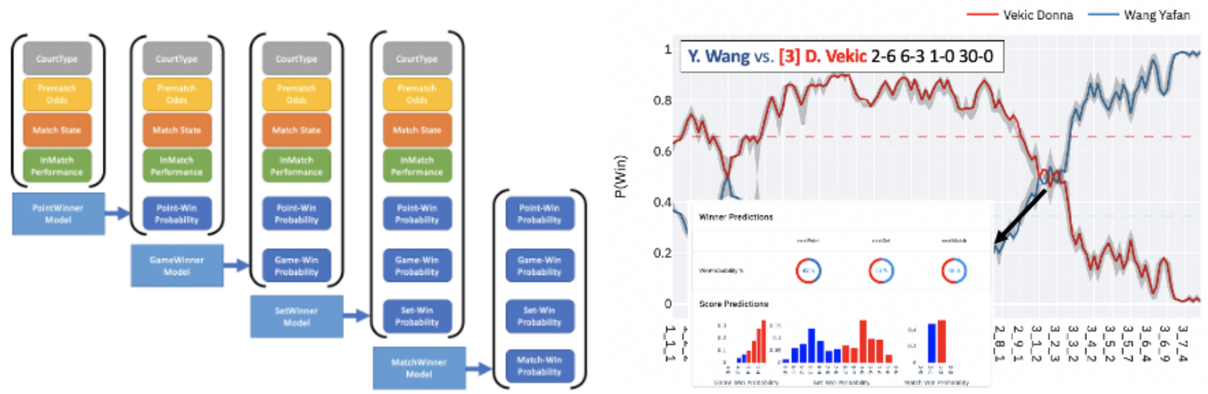


Figure 2: (**Left**) Given the hierarchical nature of tennis (i.e., point/game/set/match), our predictive model chains together outcomes at the different scales to enable self-consistency of predictions both for the actual and alternative outcomes. (**Right**) Using this approach, we can see the evolution of the match win probability and the more detailed predictions of winner and scores for the example of the Wang vs Vekic match.

Tennis is a hierarchical game where winning points lead to winning games, winning games lead to winning sets and winning sets to winning the match. We formulate the problem of predicting the winner of the next point, game, set and match as a multi-output model. Intuitively, the four models are chained as the predicted point winner will affect the game winner which affects the set winner which finally affects the match winner as illustrated in left-side of Figure 2. Each model makes a prediction in the order specified by the chain using all features plus the prediction of the earlier model in the chain. Furthermore, we train a second suite of models that predict the final game, set and match score.

Given these chained models, counterfactual analysis is then made possible by adapting the current features to an alternative future outcome (i.e., what is the likelihood of winning the next point, game, set and match if the player of interest wins or loses the current point). The approach is similar in spirit to the "4th Down Bot" [1,2] but we trigger a host of predictions at every point, which requires both very fast feature generation and model inference, and subsequent logic.

As a tennis match evolves as a sequence of points, this allows us to use our model to perform a counterfactual analysis by looking one step into the future of a match. Instead of just predicting the winner at the current point, we predict the most likely winner in the case that the next point ended in a win or lose of a player by updating the current feature vectors accordingly and rerunning our predictive models. There is no uncertainty involved in updating the feature vector after the point as in other sports like soccer or basketball.

With the score and winner predictions of the two next match states we are ready to answer what-if questions before the actual point is played out. For example, before a break point is played we can already tell how likely it will be that the player conceding the break will be able to win the set, reach a tiebreak or win a tiebreak which enables predictive analysis before the point of interest occurs.

As depicted in the left-hand side of Figure 2, the core input feature are: court-type, current match-state (i.e., current game and set score), in-match statistics (difference in points won, games won and breaks won in the current game, set and match as well as other KPI's as serve/return percentages), as well as pre-game odds as a key input feature.

These core features are concatenated with the other point/game/set and match predictions into our chain of models which take the form of a gradient boosting tree model. We trained our model on 1.5 Million points played on the WTA Tour between 2012 and 2020, and used the target labels as the next point, game, set and match winner for our winner model and the final scores for our score model. We ran a grid-search to determine the best parameters of the model.

On the right of Figure 2, we show the evolution of win probabilities of both players in the match from the introduction. The dashed lines show the pre-match odds and the shaded areas around the predictions show how winning or losing this point will affect the win probability. At each point in the match our models allow us to predict the game, set and match winner and the related scores. Each model picks other features to be important for its prediction (i.e., the point winner model focuses on who is serving, pre-match odds, if the match is in tiebreak state and the difference in points won by each player). In contrast the match winner model focuses on the set difference of the players and the chained probability that a player will win the set. Our model picks the correct next point winner at 55.8% accuracy which is in line with current state-of-the-art [3-6] and betting markets. In Figure 3, we show our comparison to the Bet365 in-play market odds where we had access for comparison.

Given that we use the pre-game market data as a key input feature, and our predictions correlate it is fair to ask the question – ***"why not just use the betting market information as the prediction output?"*** This highlights the benefit our approach, as even though a predictive market, if efficient, is the best prediction that can be obtained by virtue of the efficient market hypothesis [7], it is severely limited in its application utility as it can only provide a prediction of the current game-state. The power of our data-driven approach is that it allows literally for infinite number of hypothetical predictions to occur without the need to create a market to get people to place a bet – and it is the only way that this type of analysis can occur.

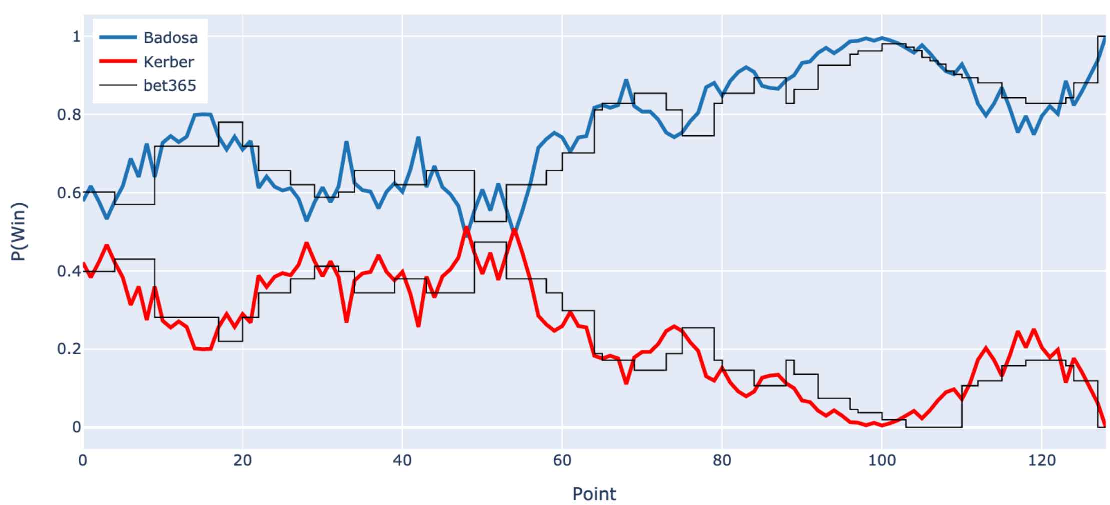


Figure 3: Comparison to marked odds: P. Bardosa vs. A. Kerber (6-4, 7-5), Semifinal, Indian Wells 2021

3. “Leverage”, “Momentum” and “Clutch” Metrics

A current short coming in tennis analytics (and sports analytics generally), is that most of the metrics are static and only related to pre-defined key moments. If a key insight is delivered which falls outside of the pre-defined templates, it is often reported in hindsight. For example, after 10 points or so, it will be reported that a player has won say 7 out of the last 10. Based on this criterium this will be construed at gaining “momentum”. Although it is fine as a coarse barometer of short-term match dominance, it also will miss a lot of important moments as it is missing match (and player specific) contextual information.

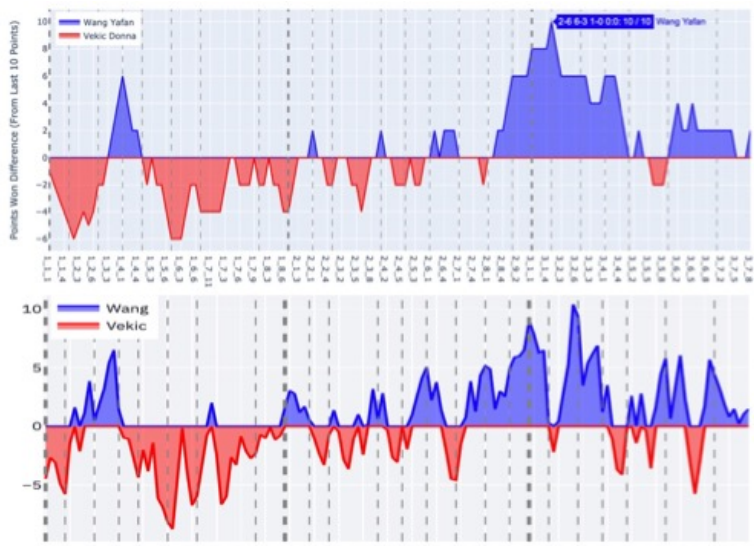


Figure 4: Estimating Momentum – (**Top**) Shows counting the consecutive points won, (**Bottom**) Shows our contextually weighted estimate using our counter-factual predictions.

For example, if a player is down 0-5 and wins 7 out of the last 10 points, the set-score would still be 1-5, and deuce in the current game which is not that interesting. Or a player might struggle to win her serve but then breaks the others serve to 15. None of this is captured in simply counting points won. At the top of Figure 4, we show the difference of the last ten points won by each player in the example game we have been using between Wang and Vekic. A positive value indicates that Wang won more than five out of the last ten points and a negative value favors Vekic. Highlighted is the point when Wang is on a run and wins 10 points in a row to win the second set 6-3 and is up 1-0 in the final third set. If we want to highlight only important changes in a match there is no simple threshold to choose as the importance of each point is not considered. We solve this problem basing our definition of momentum on leverage gained which puts an importance weight on each point.

Defining Leverage

In this paper, we define “***leverage” as measuring the magnitude of how the win probability will change as a result of either winning or losing the point***. Using our counterfactual prediction framework, we do this by using the current game-context, and then adding the future state of a player winning or losing the next point and measuring that difference. This allows us to objectively measure the importance of a point which takes into consideration the current match context but also relative player strength. Based on our leverage metric, we can then estimate the current momentum of both players. A similar concept has been used in baseball to measure the important moments during a match [8].

Momentum

Player	Points Won From Last 9 Points	Point Win %	Previous Momentum	Current Momentum	Swing	Next Point Win %
Wang Yafan	2	22.22 %	6.50 %	-6.11 %	-12.60 %	41.05 %
Vekic Donna	7	77.78 %	6.50 %	-6.11 %	12.60 %	58.95 %

Figure 5: Shows the key momentum swing at a critical point during the match.

Defining Momentum

We then define momentum ***as an exponentially weighted moving average of the leverage gained by a player:***

$$y_t = \frac{x_t + (1-\alpha)x_{t-1} + (1-\alpha)^2 x_{t-2} + \cdots + (1-\alpha)^t x_0}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^t}$$

Where $[x_0, x_1, \dots, x_t]$ are the gained leverages of the last t points with smoothing factor α and y_t the momentum at point t. In this work we utilize $\alpha = 0.33$. At a high-level, this means a player is attributed the leverage/importance of a point when she wins it and it is reduced when she loses it and the latest points are weighted more heavily in the momentum equation. In the bottom of Figure 4, we show our match momentum plot. In Figure 5, we highlight a point in the match which had a strong momentum swing. Specifically, we detect that the momentum changes looking back 8-12 points. If a threshold of 3% was reached, and there was a zero crossing and the threshold was reached for the other player, we say that the momentum changed. Using this trigger, we can also report on other interesting insights, such as the points won metric in this interval and also what our model thinks who will win the next point.

Another problem with current tennis metrics is that they are very simple and rigid. Break points are by definition important as you can't win a set without a break. The more break points you win the more dominant you are given that you hold serve. Game points on your own serve are considered less important than break points but more important than all other points because you are only one point away to win a game. But here the importance of winning this point for winning the match in the end is ignored. A break at the beginning of a set is worth less than at 5-4. Also, all points that Brad Gilbert called setup points in "Winning Ugly" [9], where you are two points away from winning a game are ignored. To summarize, key points cannot be pre-defined and are heavily contextual so we need a metric which can capture the in-game context to highlight the key points before those points actually are played. We call these points "clutch points" and our definition of clutch is given below.

Defining Clutch

In this paper, we define ***"clutch" as an important point where winning or losing it has a significant effect on the current match win probability. More formally we say that a point is clutch when its leverage exceeds 10%.***

Clutchpoint (2-3 30:30) SEASON ANALYSIS

Player	Current Match Win%	Match Win% - If Player Wins Next Point	Swing	Leverage
Wang Yafan	26.66 %	34.31 %	7.65 %	11.49 %
Vekic Donna	73.34 %	77.18 %	3.84 %	11.49 %

Clutchpoint (2-3 30:40) SEASON ANALYSIS

Player	Current Match Win%	Match Win% - If Player Wins Next Point	Swing	Leverage
Wang Yafan	22.82 %	26.66 %	3.84 %	12.27 %
Vekic Donna	77.18 %	85.61 %	8.43 %	12.27 %

Comeback Chance

What is the chance of a comeback if Wang Yafan loses her service game on this breakpoint?

Player	Set Win %	Reach Tiebreak %	Win Tiebreak %
Wang Yafan	10.05 %	8.54 %	40.40 %

Figure 6: (**Left**) A clutch point in the mid of the first set that is not a break point. (**Right**) The following break point with an additional analysis of a comeback for Wang if she concedes the break.

For example, Donna Vekic is 3-2 down in the first set (highlighted in Figure 1(B) in Section 1) and it is 30-30 at Vekic's serve. Taking into account pre-match predictions and current performance of both players our win predictor identifies this point a clutch point (11.49%), left-side in Figure 6. By losing this point Vekic will need to win at least three more points to win the game played over advantage and Wang will only need to win the next point to serve at 4-2. In fact, Wang wins the setup point and her break point, right-side in Figure 6, has an even higher leverage of 12.27%. At this point our counterfactual approach allows us to already answer the question of a likely comeback of Wang if she will concede the break before it actually happens. We can see that being 4-2 down with a break she will only be able to win the set 10% of the times and she is also very unlikely to reach a tiebreak even though she would be able to win the tiebreak two out of five times.

4. Season-Level Performance of Players

In the previous sections, we have highlighted what our counter-factual prediction framework is, how it can be used to create dynamic metrics which can then be used to enhance the story-telling aspects of a game within a single tennis match. In this section, we highlight how we can capture the season-long behaviors of players which can describe their emerging behaviors. Given our leverage, momentum and clutch metrics on a point-level, we can aggregate them and get insights into the season-level performance of players. Often player performance is measured in how many points a player was able to win, how many break points a player created or conceded and how many of these she converted. Again, as break points are all weighted equally this does not tell the whole story. We can use our definition of clutch points to find the player who handled these important points the best. Figure 7 shows the individual performance of the top 5 ranked players at the end of 2019 in terms of points won and clutch points won 2019.

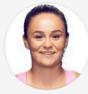
Player	clutchPointsPlayed	clutchPointsWon	clutchPointsWonPct	PointsPlayed	PointsWon	PointsWonPct	PointsWonPctDiff	WTA Ranking
 Ashleigh Barty	849	470	55.36%	6044	3194	52.85%	2.51%	1
 Bianca Andreescu	487	269	55.24%	3211	1704	53.07%	2.17%	5
 Naomi Osaka	528	287	54.36%	4375	2339	53.46%	0.89%	3
 Karolina Pliskova	913	478	52.35%	6562	3480	53.03%	-0.68%	2
 Simona Halep	524	260	49.62%	4686	2462	52.54%	-2.92%	4

Figure 7: The top 5 players in the 2019 WTA season who had the most clutch points.

In this Figure, we see that world no.1 Ashleigh Barty did not have the highest points won percentage (52.85%) but she won 55.36% (+2.51%) of all her clutch points, which made her the strongest clutch player in 2019. In contrast, Simona Halep performed equally well in terms of points won percentage (52.54%) but only won 49.62% (-2.92%) of her clutch points.

One thing to keep in mind for this analysis is that clutch points are highly individual to a match and there might be only few when a match ended 6-2 6-2 but a lot when a match is close and goes into a tiebreak. Thus, seeing only few clutch points is not necessarily bad for a player. We can also compare the performance of all players and the top 10[1] on points and specifically on clutch points in 2019. Figure 8 shows the performance of all players. Dashed lines indicate the means of the players on points won percentage and clutch points won percentage. We see that all top players in 2019 won significant more points than the average player. On clutch points this is only true for some player like Ashleigh Barty but false for Simona Halep.

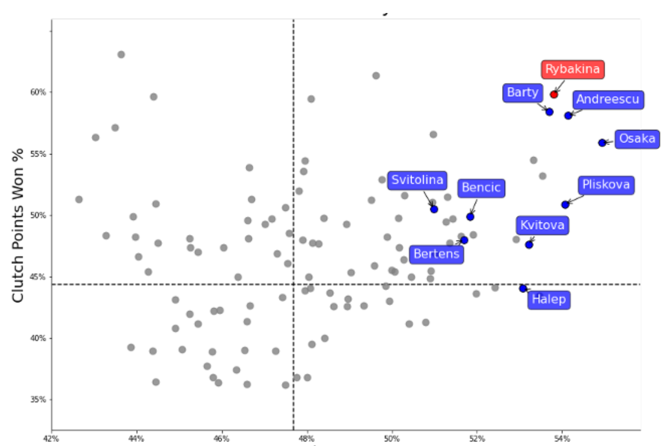


Figure 8: Cluster analysis of all the players in the 2019 WTA season, where we see all the top players grouped together. A benefit of this approach is that we can see players within this cluster which are not in this group which could give an indication of future emerging stars.

The analysis could also give an indication of future emerging stars like Elena Rybakina who played clutch points well and finished 2019 as world no. 37 and as no. 14 in 2021. Additionally, having a low number of clutch points in a match or season can be a strong indication of dominance. Conversely, if you have a match with many clutch points, that is a strong indicator of an exciting match which can

[1] The criteria to be on the list was that a player had played at least 10 matches and 100 clutch points played. As Serena Williams did not play at least 10 matches and have 100 clutch points, she was not on this list in 2019.

be used as a barometer to draw viewers attention during the game live, or highlight as a must re-watch game. The top 5 most exciting matches in the 2019 WTA season are shown in Table 1.

tournament	date	HomeTeam	ForeignTeam	sum_leverage	num_clutchpoints	Result
Bucharest Open	2019-07-16	Sevastova Anastasija	Bogdan Ana	1,578.68	91	[1]A. Sevastova d A. Bogdan 5-7,7-6(4),7-5
Sydney International	2019-01-08	Stephens Sloane	Alexandrova Ekaterina	1,570.86	86	[4]S. Stephens d E. Alexandrova 0-6,7-6(3),7-6(3)
Japan Women's Tennis Open	2019-09-11	Hsieh Su-Wei	Hon Priscilla	1,476.58	83	[1]S. Hsieh d P. Hon 1-6,7-6(2),7-5
BGL BNP Paribas Luxembourg Open	2019-10-14	Allertova Denisa	Minella Mandy	1,413.42	85	D. Allertova d M. Minella 6-7(7),7-6(5),7-6(6)
BGL BNP Paribas Luxembourg Open	2019-10-17	Puig Monica	Pliskova Kristyna	1,341.47	72	M. Puig d K. Pliskova 7-6(5),3-6,7-6(6)

Table 1: Top 5 most exciting matches during the 2019 WTA season according to our clutch point indicators.

5. Summary

In this paper we presented an method which utilized counter-factual predictions as a method of providing interpretable machine learning outputs to analyze tennis matches. The power of this approach is that it is dynamic and identifies key moments based on the relative strengths of players and what has occurred during the match 9and does not rely on the notion of pre-defined key moments such as break points. To do this, we have created the metrics of "leverage", "momentum" and "clutch" which are created by chaining counter-factual predictions which capture the importance of a point contributing to a player winning the set and/or match, or the likelihood of a comeback. Not only can our approach highlight important moments before they occur in an automatic fashion (and not based on a pre-canned notion of what are important points such as break points), it can also link player behaviors at a season level which shines a light on their tendencies in key moments. We showcased the utility of our approach on many examples within a match, as well at a season level. The power of our approach is that it can address any "what-if" hypothetical question which can drive potentially infinite insights to be generated. In terms of future work, the next logical step is to leverage player and ball tracking data and include finer-grain predictions (such as most likely serve location, or next shot location) which would enable not only a better user experience but would also be valuable to players as they prepare for future opponents.

References

[1] K. Clark, "Go for It: The Story Behind the NFL's Fourth-Down Conversion", Nov 13, 2019.
[2] B. Burke, S. Carter, T. Giratikanon and K. Quealy, "The New York Times, 4th Down Bot", 2013.
[3] X. Wei, P. Lucey, S. Morgan, M. Reid, & S. Sridharan, "The Thin Edge of the Wedge: Accurately Predicting Shot Outcomes in Tennis using Style and Context Priors", in MIT SSAC, 2016.
[4] C. Floyd, M. Hoffman, & E. Fokoue, "Shot-by-shot stochastic modeling of individual tennis points", Journal of Quantitative Analysis in Sports, 16(1), 57-71, 2020.
[5] S. Kovalchik & M. Reid, "A calibration method with dynamic updates for within-match forecasting of wins in tennis", International Journal of Forecasting, 35(2), 756–766, 2019.
[6] T. Barnett & S. Clarke, "Combining player statistics to predict outcomes of tennis matches", IMA Journal of Management Mathematics, 16(2), 113-120, 2005.
[7] Efficient Market Hypothesis: http://en.wikipedia.org/wiki/Efficient-market_hypothesis
[8] N. Allen, J. Templon, P. McNally, L. Birnbaum & K. Hammond, "StatsMonkey: A Data-Driven Sports Narrative Writer", Computational Models of Narrative, Papers from the AAAI Fall Symposium, 2010
[9] Brad Gilbert & Steve Jamison, "Winning Ugly: Mental Warfare in Tennis – Lessons from a Master", Simon and Schuster, 31.05.1994