Exploring the Importance of Players’ Characteristics and Performance on Serve and Return of Serve in Winning the Women’s Singles Grand Slam Tennis Tournaments

A W Zainan Abidin¹ and N A S Ruslan²

¹,²Center for Statistical & Decision Science Studies, Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, MALAYSIA

Email: ¹aida@tmsk.uitm.edu.my

Abstract. Sports analytics has become very popular in recent years. Many researchers have been using the tools and technique to study tennis player performance in various aspects during matches. However, much of the studies were performed on male players while much less attention were made on female players. This study explores the female players’ characteristics and serve performance that leads to winning matches in women’s singles 2018 Grand Slam tennis tournaments. Using data provided by the official website of the Women Tennis Association (WTA), 1016 single matches in all four 2018 Grand Slam Tournaments were analysed to evaluate winning traits in tennis specifically looking at the two most important strokes in tennis which are the serve and return of serve. Additionally, other personal player characteristics that would influence the match outcomes were also evaluated. To achieve the objectives of the study, odds ratio and chi-square test were used to compare between the match winner and match loser.

The study revealed that percentage of first serve return won (PFSRW) has the highest odds ratio value (OR=1.263) followed by the percentage of first serve won (PFSW) with odds ratio equal to 1.244. These showed that having a good first serve, plays an important factor in increasing the winning odds as compared to second serve either in serve or in return of serve. This study also reveals that a taller player had the advantage over a shorter player for faster serve speed (p-value<0.05).

1. Introduction

The advancement of the sports analytics has made it possible to analyze the performance of any matches in sports including tennis. According to a research done by [1], tennis is identified as the fifth most widely played sports in the world. With the advent and improvement of sports analytics, and the availability of the matches statistics online have provided an opportunity for sport analysts to study factors that would lead to winning a match or at least to play a better game in a match.

The main focus of this research is to identify important factors that lead to winning matches in women’s singles tennis match by exploring player characteristics and serve performance. Samples are taken from all four 2018 Women’s Singles Grand Slam Tournaments, namely US Open, Wimbledon Open, French Open and Australian Open. There were 1016 matches in all four 2018 Women’s Singles Grand Slam Tournaments in which all the match data were made available at the WTA official website. This is a cross sectional study in which data was gathered at one point of the time in 2018.
The most important strokes in tennis is the serve and the return of serve. Every tennis player would need to serve well in order to score a point, and would also need to return serve well in order to break the server dominance. Player often serves an aggressive first serve and serves second serve conservatively, if the first serve is faulted. Good serve can shorten the match and helped to set up the rest of the game. Because of these, serve is nominated as the most importance stroke in tennis. The odds ratio and chi-square test is used in analyzing the data to determine the factors which differentiate the match winner and the match loser.

2. Literature Review
Psychological factor is one of the important criterions for an excellent tennis player. Good psychological factor is an added value to the tennis player in helping them to score well in tennis match. Good physical factors also could provide higher chances of winning the tennis matches.

Taller players could reach higher contact point as they would have an extra reach at the ball that enable them to hit distant balls more easily compared to shorter player. Serve speed (first and second), is also related to a physical and morphological features in particular to the male professional players. Player height and body mass are among the main anthropometric determinant of serve speed in tennis player [2]. Using this as the baseline, this study explored the height and weight of the female tennis player in order to see whether they were one of the contributing factors to win the match or not.

Another factor being analysed in this research was the player’s type, i.e. right-handed or left-handed player. According to [3], left-handed player will benefit more when the opponent is right-handed player. Right-handed player must adjust their return stroke as the ball served by left-handed player would differ in their spin from the right-handed player. Often, this was a disadvantage over a right-handed player who were not familiar playing with left-handed player, hence had less experience with their style of play.

Age is also an important role in predicting the match outcome. Research done by [3] found that as player get older, they were generally in the more experience state but in poorer physical condition. In tennis, the high stamina and good muscle strength was important to help the player maintained the rally. Aging would result in the gradual loss in muscle function with type II muscle fibre being the most affected [4]. This is also supported by [3].

Ranking of the player contributes to the seeding; which is a system in any professional tournaments that was used to separate top players in a draw so that they would not meet in the early rounds of the tournament. In the tennis tournament, the top seed was given to the player that the tournament committee deemed the strongest player in that particular tournament during that particular of time. The top seed and the second seed were placed at the opposite end of the draw so that, if they both kept winning, they would meet at the final round. The number of seeds was determined based on the size of the draw. Strong correlations exist between physical qualities and performance in all age and gender groups [5].

Research done by [6] found that women players had more service winners, double faults, return winners and return unforced errors in the Australian Open and US Open, implying a “fast-fast” serve strategy, and higher dominance ratio and better serving performance in Wimbledon. In Roland Garros, receiver had greater chance to break opponents serve in order to obtain the game points. They also added that women players won more first and second service points at an average of 66.78% (11.1%) and 46.44% (12.7%) respectively in Wimbledon, served more aces with a mean of 4.84% (5%) and committed more double faults averaged at 13.40% (8.3%) in Australian Open. It is also found out that the first and second most influential clutch statistics for both junior girls and professional women were the total points won and the first serve points won. [7]

3. Methodology
The aim of this research is to analyse the match performance of tennis player based on the players' characteristics and serve performance that could lead to winning matches in women's singles 2018 Grand Slam tennis tournaments. The input variables are Height, Weight, Age, Types, Ranking, PFSI, PFSW, FSS, PFSRW, PSSW, SSS, PSSRW, Aces, DF, Surface and Score. The target variable is Score
which is coded 1 for won and 0 for lose. The characteristics of the players are presented in Table 1. While, the statistics related to serve performance in the match is presented in Table 2.

### Table 1. Description of Player Characteristics

| Variables | Role  | Scale | Description            |
|-----------|-------|-------|------------------------|
| Height    | Input | Ratio | Height of the player   |
| Weight    | Input | Ratio | Weight of the player   |
| Age       | Input | Ratio | Age of the player      |
| Types     | Input | Nominal | Right-handed or Left-handed |
| Ranking   | Input | Ordinal | Ranking of player      |

### Table 2. Description of Serve Performance

| Variables | Role  | Scale | Description                          |
|-----------|-------|-------|--------------------------------------|
| PFSI      | Input | Interval | Percentage first serve in           |
| PFSW      | Input | Interval | Percentage first serve win           |
| PSSW      | Input | Interval | Percentage second serve win          |
| FSS       | Input | Interval | First serve speed                    |
| SSS       | Input | Interval | Second serve speed                   |
| DF        | Input | Interval | Number of double faults              |
| PFSRW     | Input | Interval | Percentage first serve return won    |
| PSSRW     | Input | Interval | Percentage second serve return won   |
| Aces      | Input | Interval | Number of aces                       |
| Surface   | Input | Nominal | Type of court surfaces:             |
|           |       |        | 1=Hard                               |
|           |       |        | 2=Grass                              |
|           |       |        | 3=Clay                               |
| Score     | Target | Binary | 1 = win                              |
|           |       |        | 0 = lose                             |

Apart from descriptive statistics to demonstrate the players’ characteristics using frequency table, count and percentage, the data was also analysed by interpreting the odds ratio and chi-square statistics between the match loser and match winner. In general, the odds ratio (OR) is a measure of association between an exposure of an outcome [8]. In addition, fourfold diagram was also used to visualise the significance of the odds ratio observed. According to [9], visualisation of the patterns of relationships using appropriate graphics would provide a better understanding of the relationships prior to any model building to describe the odds ratio.
In this study, The OR simply represents the odds or the chance of players with certain attributes to win a match.

The value of:
- Odds ratio = 1  The players attributes does not affect the odds of winning the match
- Odds ratio > 1  The players attributes would have higher odds of winning the match
- Odds ratio < 1  The players attributes would have lower odds of winning the match

The chi-square test was commonly used for testing the relationship between categorical variables. It is used to evaluate test of independence when using cross-tabulation. Cross-tabulation presents the distribution of two categorical variables simultaneously. With the intersection of the categories of the variables appear in the cells of the table. Calculating the chi-square statistic and comparing it against critical value from chi-square distribution allow researcher to assess whether the observed cell counts are significantly different from the expected cell counts [10]. In this study, chi-square test is used to examine if any characteristics of players performance has any association in winning or losing a match.

4. Results and Discussion

4.1. Player Characteristics
The distribution of type of players in the sample is shown in Figure 1 below.

As indicated in Figure 1 above, most of the female tennis players are playing with their right hand. It is represented by 873 right-handed players (91.51%), while only 81 players (8.49%) are left-handed players. The information of the players characteristics are summarized in the Table 3.

The dominance age of players who take part in the matches came from age ranging between 21 to 30 years old. These players are moderate age group where they are not too young and not too old. They might be in the state of great stamina with some experience as tennis players. As for the height group, most of the players who take part in the game have height between 171cm to 180 cm. They are categorized as taller players. Meanwhile, most of the female players have the dominance weight of 61kg to 70 kg.
Table 3. Summary for Player Characteristics

| Variable | Group | Description     | Won         | Lose         |
|----------|-------|-----------------|-------------|--------------|
| Age      | 1     | <= 20 years old | 6 (1.21%)   | 10 (2.04%)   |
|          | 2     | 21 – 30 years old | 359 (72.53%) | 381 (77.60%) |
|          | 3     | >= 31 years old  | 130 (26.26%) | 100 (20.37%) |
| Height   | 1     | <= 160 cm       | 7 (1.44%)   | 9 (1.99%)    |
|          | 2     | 161 – 170 cm    | 145 (29.87%) | 155 (34.29%) |
|          | 3     | 171 – 180 cm    | 270 (55.67%) | 225 (49.78%) |
|          | 4     | >= 181 cm       | 63 (12.99%) | 63 (13.94%)  |
| Weight   | 1     | <= 60 kg        | 110 (24.02%) | 99 (25.38%)  |
|          | 2     | 61 – 70 kg      | 289 (63.10%) | 251 (64.36%) |
|          | 3     | >= 71 kg        | 59 (12.88%) | 40 (10.26%)  |

4.2. Summary Statistics for Serve and Return of Serve

Table 4 shows that the mean for the percentage of first serve in (PFSI) is 61.86% for all the four Grand Slam tournaments. The average double fault is 4 with the maximum value for the double fault is 17. Double fault is the phase in which player faulted both their serve and subsequently lose a point to the opponent. As for ace, the minimum value was 0 and the maximum aces is 18, giving the mean value of 3 for all the matches. Ace is a serve from the server that successfully lands in the service box and failed to be touched by the receiving player’s racquet.

The mean percentage of first serve won (PFSW) is higher than percentage of second serve won (PSSW) with the values of 63.93% and 45.56% respectively. This shows that server has full controlled over the ball during a faster first serve. Another aspect of serve that is important during serve is the return of serve in which players are able to return the serve from the server well. Players are more likely to score well on return of serve and won more points during the second serve with the mean values of 40.57% for the percentage of second serve return won (PSSRW) and 35.92% for the percentage of first serve return won (PFSRW). As the server serve the ball more conservative, the opponent can attack the server by making them to lose control over the ball during the slower second serve.

Table 4. Summary Statistics for Serve Performance

| Variable | Mean   | Standard deviation | Median | Maximum | Minimum | Skewness | Kurtosis |
|----------|--------|--------------------|--------|---------|---------|----------|----------|
| DF       | 3.72   | 2.66               | 3      | 17      | 0       | 1.036    | 1.36     |
| PFSW     | 63.93  | 10.94              | 64     | 100     | 16      | -0.29    | 0.64     |
| Aces     | 2.88   | 2.86               | 2      | 18      | 0       | 1.68     | 3.70     |
| PFSI     | 61.86  | 7.72               | 62     | 100     | 30      | -0.01    | 0.91     |
| PFSRW    | 35.92  | 10.89              | 36     | 83      | 0       | 0.28     | 0.63     |
| PSSRW    | 40.57  | 11.51              | 41     | 83      | 0       | -0.02    | 0.53     |
| PSSW     | 45.56  | 12.13              | 45     | 100     | 0       | 0.20     | 0.78     |
| FSS      | 156.95 | 9.00               | 157    | 180     | 120     | -0.40    | 0.37     |
| SSS      | 132.02 | 8.68               | 132    | 158     | 105     | 0.06     | -0.28    |
The mean of first serve speed (FSS) is 156.95 km/h and the mean for the second serve speed (SSS) is 132.02 km/h. This shows that players would serve first aggressively to take full controlled over the ball and more conservative on the second serve. When the server served aggressively, the opponent would have to play defensive and only able to return the ball to the server without being able to attack the server back. The server would then be in the position that enable them to attack the opponent with faster ball speed provided no unforced error was made. On the other hand, if the first serve was faulted, the server tend to serve the second serve conservatively for not risking of losing the point by making sure the ball continue to the rally. However, this gives an opportunity for the opponent to attack the server back. In addition, the skewness and kurtosis values also indicate that the distribution of the variables are normally distributed [11].

4.3. Test of Association for Players Characteristics

Table 5 gives the chi-square values for the categorical group in determining the match status of tennis match.

| Variable       | Chi-square | Df | p-value   |
|----------------|------------|----|-----------|
| Seeded group   | 56.2611    | 1  | <0.0001   |
| Weight group   | 5.9263     | 2  | 0.0517    |
| Age group      | 4.0457     | 2  | 0.1323    |
| Type           | 0.6577     | 1  | 0.4174    |
| Height group   | 0.4576     | 2  | 0.7955    |

Based on the above, only one variable that was significant which is seeded group with the p-values of <0.0001. Therefore, it could be concluded that there is an association between seeded group and winning in tennis match in the Grand Slam Tournaments. Seeded player tends to win in the tennis match as compared to non-seeded player.

Table 6 shows the contribution of player height and the speed of their serves in winning a match. It can be seen that players with height greater than 171cm (taller players) are significantly more likely to win the match with faster first serve speed (p-value <0.05).

| Height of players (cm) | Match Status | Mean (kmh) |                  |                  |
|------------------------|--------------|------------|------------------|------------------|
|                        |              | First serve speed | Second serve speed |
|                        | Lose         | Win        | Lose             | Win              |
| <=160                  |              |            |                  |                  |
|                        | Lose         | Win        |                  |                  |
| 161-170                | Lose         | Win        |                  |                  |
| 171-180*               | Lose         | Win        |                  |                  |
| >=181                  | Lose         | Win        |                  |                  |
4.4. Comparing of Serve Performance based on Match Status

Table 7 summarizes the serve performance based on winning or losing status.

| Table 7. Descriptive Statistics of Serve Performance based on Match Status |
|-----------------|-----------------|-----------------|
| Variable        | Match Status    | Mean            | Std. Deviation |
| PFSI            | Won             | 62.50 (n=507)   | 7.424          |
|                 | Lose            | 61.23 (n=507)   | 7.957          |
| PFSW            | Won             | 70.10 (n=507)   | 8.555          |
|                 | Lose            | 57.74 (n=507)   | 9.497          |
| PFSRW           | Won             | 42.14 (n=507)   | 9.429          |
|                 | Lose            | 29.70 (n=507)   | 8.410          |
| PSSW            | Won             | 51.28 (n=507)   | 11.314         |
|                 | Lose            | 39.85 (n=507)   | 10.056         |
| PSSRW           | Won             | 44.93 (n=507)   | 10.950         |
|                 | Lose            | 36.21 (n=507)   | 10.346         |
| DF              | Won             | 3.35 (n=507)    | 2.621          |
|                 | Lose            | 4.10 (n=507)    | 2.652          |
| Aces            | Won             | 3.44 (n=507)    | 3.193          |
|                 | Lose            | 2.32 (n=507)    | 2.366          |
| FSS             | Won             | 157.85 (n=347)  | 0.459          |
|                 | Lose            | 156.06 (n=347)  | 0.503          |
| SSS             | Won             | 131.60 (n=346)  | 0.463          |
|                 | Lose            | 132.45 (n=346)  | 0.469          |

The first variable being measured was percentage of first serve in, winner had higher mean of percentage of first serve in as compared to loser. This indicated that excellent players always makes sure that their first serve in to have a better control of the point. Winners also had higher mean of percentage of first serve won and percentage of first serve return won as compared to losers. The same patterns were shown in second serve performance, either in serve or serve return. The percentage was higher in winners than the losers.

Most of the loser always makes fault during the first and second serve called double faults that would give points to the opponent. This mistake could make most of them to lose points and lose in the match. This was proven as the mean of double fault for loser is 4.10 as compared to winner of only 3.35. The number of aces score was also high in winner than loser given by the mean values of 3.44 and 2.32 respectively. Lastly, winner has high mean first serve speed, but lower mean of second serve speed. As overall comparison, winner always excel in all the variables with minimum number of double faults occurred but for loser, most of them always make faults during first and second serve and did not conquered all the other serve performance.

4.5. Summary of Odds Ratio for Serve Performance

Table 8 summarized and interpret the value of odds ratio calculated for serve performance variables. Only the top five variables with the highest odds ratio are displayed. Out of these five variables, two odds ratio are significant at 5% level.
Table 8. Odds Ratio of Significant Variables

| Variable | Odds ratio | Interpretation |
|----------|------------|----------------|
| PFSI     | 1.047      | Players with higher percentage of first serve in are 1.047 times more likely to win the tennis match. |
| PFSRW    | 1.263      | Player with higher percentage of first serve return won are 1.263 times more likely to win the tennis match. |
| PFSW     | 1.244      | Player with higher percentage of first serve won are 1.244 times more likely to win the tennis match. |
| PSSRW    | 1.078      | Player with higher percentage of second serve return won are 1.078 times more likely to win the tennis match. |
| PSSW     | 1.124      | Player with higher percentage of second serve won are 1.124 times more likely to win the tennis match. |

In addition, the odds of certain attributes of the players are also examined to explore the chances of the players to advance to at least Quarter Finals of the tournaments. The odds ratio of certain serve performance were examine to examine if they might contribute to higher chance of winning a match are shown in Figure 2 to Figure 4.

![Figure 2. Fourfold Plot for Odds Ratio PFSW and Final Round](image2.png)

![Figure 3. Fourfold Plot for Odds Ratio Seeded and Final Round](image3.png)
Figure 2 shows the fourfold plot of PFSW towards the Final Round in which when the percentage of first serve won is greater than 80%, player will have 1.59 likely chance to be in the final rounds. Figure 3 shows the advantage of being a seeded players. The odds i is 3.57 for these players to win a match compared to unseeded players. Lastly, Figure 4 indicates that players who can take advantage of the opponent first serve by winning the point will have slightly higher chance to be in the next round of the tournament.

4.6. Discussion
Based on the findings, it could be concluded that being a taller player has an advantage for winning a women’s singles tennis match provided if the player can improve on their first serve speed. Player should maintain a high percentage of getting a point during the first serve return to take control of the game. At the same time, it is also important to hold their own serve by assuring higher percentage of first and second serve won. Players who can overcome their opponent’s first serve would benefit a chance of winning the point. Being a seeded player is a huge advantage. These results support by [12] findings that player won significantly more points on their first serve than the non-seeded player (45.6% > 43.1%). Hence, a seeded player distinguishes themselves from a non-seeded player by having a better first serve. Figure 5 shows the summary of the variable importance.
Based on the findings of the importance of the serve and return serve, the following component are ranked according to their importance:

1) Percentage of first serve return won  
2) Percentage of first serve won   
3) Percentage of second serve won  
4) Percentage of second serve return won  

These result supported the findings by [13] when comparing two groups of players who won more matches (G_1) with the players who lost more matches (G_2). They found out that in 2010, G_1 had won 72.7% of first serve points and 52.4% of second serve-points compared to only 68.6% and 48.7% respectively for G_2. In addition, another research by [14] also showed same results where winning players had significantly higher values for aces, percentage of points won on first serve, points won on second serve, and percentage of points won on second serve.

Apart from the skills, personal attributes such as height could also help the players to play a better game against their opponent. The result is consistent with the findings by [15] where most female players at the top of the world tennis ranking list were 185 cm and above.

5. Conclusion  
Numerous methodologies had been applied in studying the performance of tennis match in the Grand Slam Tournaments but only a few analyse the women’s singles tennis performance as compared to men. Hopefully this research paper could close the gap by analysing the tennis performance of the female singles match that mainly focusing on serve performance for the four main Grand Slam tournaments. The findings would certainly benefit the players specifically and the coaches as it can be used as the guidance towards the planning of the training. On the other hand, players should spend more time training on their first serve skills and also the first serve return skills. No doubt, serve and return of serve are two most important strokes that need to be focused on in order to have higher chance of winning a match. If not, then certainly it would help players to play a better match in any of the tournaments.

Acknowledgement  
We would like to thank Universiti Teknologi MARA Shah Alam and Faculty of Computer & Mathematical Sciences, UiTM for supporting this research.

References  
[1] Dhinesh R, Preejith S P, and Sivaprakasam M 2018 Tennis Serve Correction Using a Performance Improvement Platform in Proceedings of IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH) Vienna pp. 1-7  
[2] Bonato M, Maggioni M A, Rossi C, Rampichini S, La Torre A, and Merati G 2015 Relationship Between Anthropometric or Functional Characteristics and Maximal Serve Velocity in Professional Tennis Players The Journal of Sports Medicine and Physical Fitness vol 55 no. 10 pp. 1157–1165  
[3] Ma S M, Liu C C, Tan Y, and Ma S C 2013 Winning Matches in Grand Slam Men’s Singles: An Analysis of Player Performance - Related Variables from 1991 to 2008 Journal of Sports Sciences vol 31 no. 11 pp. 1147–1155  
[4] Kirkendall D T, and W E 1998 Garrett The Effects of Aging and Training on Skeletal Muscle The American Journal of Sports Medicine vol 26 no. 4 pp. 598–602  
[5] Ulbricht A, Fernandez-Fernandez J, Mendez-Villanueva A, and Ferrauti A 2016 Impact of Fitness Characteristics on Tennis Performance in Elite Junior Tennis Players Journal of Strength & Conditioning Research vol 30 no. 4 pp. 989-998
[6] Cui Y, Gómez M A, Gonçalves B, and Sampaio J 2018 Performance Profiles of Professional Female Tennis Players in Grand Slams PLoS ONE vol 13 no. 7 pp. 1–11
[7] Kovalchik SA, and Reid M 2017 Comparing Matchplay Characteristics and Physical Demands of Junior and Professional Tennis Athletes in the Era of Big Data Journal of Sports Science and Medicine vol 16 pp. 489–497
[8] Szumilas M 2010 Explaining Odds Ratios Journal of Canadian Academic Children Adolescent Psychiatry vol 3 no.8 pp. 227–229
[9] Rasimah and Zainan Abidin A W 2008 Using SAS Graphics To Explore Employer Preference When Hiring New Graduates In Proceedings of the 22nd Annual SAS Malaysia Forum
[10] Wackerly D D, Mendenhall W, and Scheaffer R L 1996 Mathematical Statistics with Applications 5th Edition
[11] Kim H Y 2013 Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) using Skewness and Kurtosis Journal of Restorative Dentistry & Endodontics vol 38 no. 1(1) pp. 52
[12] Magnus J R, and Klaassen F J G M 1999 On the Advantage of Serving First in a Tennis Set : Four Years at Wimbledon The Statistician Part 2 pp. 247–256
[13] Filipcici A, Zecic M, Reid M, Crespo M, Panjan A, and Nejc S 2015 Differences in Performance Indicators of elite Tennis Palyers in the Period 1991 - 2010 Journal of Physical Education and Sport vol 15 no. 4 pp. 671-677
[14] Sanchez-Pay A, Torres-Luque G, Fernandez-Garcia AI, Sanz-Rivas D, and Palao JM 2017 Differences in Game Statistics between Winning and Losing for Male Wheelchair Tennis Palyers in Paralympics Games Motriz: rev. educ. fis. vol 23 no 3 pp. 1-6
[15] Söğüt M 2018 Stature: Does it Really Make a Difference in Match-play Outcomes among Professional Tennis Players? International Journal of Performance Analysis in Sport vol 18 no. 2 pp. 255–261