1. Summary
These days, the phrase "Greatest Of All Time" (G.O.A.T) has been introduced among those fascinated by sports. It plays a vital role in the sports world or sports industry, indicating which athlete is renowned for the most outstanding performance in each sport. If someone asks you, "Who is the G.O.A.T of basketball?" You might quickly answer him, "Michael Jordan." However, finding the G.O.A.T is not easy as it seems because this topic has been arguing for a long time, but there still has not been any exact and precise answer yet. It depends on how people judge and define one's performance. Moreover, several factors affect finding the G.O.A.T, which are **the results of competitions, the Number and consistency of participating competitions, and the Z-score** of each participant. For this reason, our team aims to develop a model called "GFinder," which is the mathematical model for finding the G.O.A.T accurately. In order to find the correct G.O.A.T, our team must always keep in mind that there are some regulations of our model, which are athletes' **strategies and tactics, the point-counting method, fairness of judgment, and other factors, such as athlete's health and surroundings.**

In the GFinder model, there are two main formulas: the first one is subject to rank the players' rankings, while the second one is for the players who have been inactive for a long time. For the first one, we mainly focus on the point difference, points, the ratings of two individuals including winning and losing. If the rating of two players is enormously different, the points that receive or lose will be low, whereas if the rating of two players is just slightly different, the points that receive or lose will be high. If a player wins, he gets points. In contrast, if he loses, he will lose his points. Furthermore, if the point difference increases, the points that receive or lose will also increase. If the two players' point difference is very close, the two players' score will be high. But in tough contests, they will get a higher score than in easy competitions.
The second one; however, is different from the first one. The main concepts are: Firstly, in the case of rating score is above 1200 and if the duration of the competition is not high, it will cause the received score to decrease enormously. Nevertheless, for tennis, the scores will never be less than 1200, while the decrease rate of points will keep increasing.
Based on the above factors, we then explore the different types of values in our model by analyzing with datasets from the internet. Next, we input the data into our mathematical model written in a python program. We chose and improved our model by changing the values, until the result of our model is precise enough in terms of ranking and determining who is the G.O.A.T.
Our model is designed to suit an individual and can calculate the players' performance from the closeness of the points very well; however, there is still some weakness because of the insufficient data that we used.

Keywords: Elo rating system, Inactive player, Closeness, Dichotomous

2. Table of Contents

1. Summary 1
2. Table of Contents 2
3. Introduction
   3.1. Background 3
   3.2. Restatement of Problems 3
4. Regulations
   4.1. Strategies and Tactics 4
   4.2. Point-Counting Method 4
   4.3. Fairness 4
   4.4. Surroundings 4
   4.5. Physical and Mental Health of Athletes 4
   4.6. Inappropriate data 4
5. Factors
   5.1. The results of competitions 5
   5.2. The number and consistency of participated competitions 5
   5.3. Z-score of each participant 5
6. Analysis Data, Variables, Equations in GFinder Model
   6.1. Analysis Data used 5
   6.2. Equations and Variables in GFinder Model
      6.2.1. GFinder in tennis sport 5
      6.2.2. Inactive players' rating (previously mentioned in 4.2.3.) 10
      6.2.3. Applying with Individual Sports (Running) 12
      6.2.4. Applying with Any Individual Sports 16
      6.2.5. Applying with Team Sports 16
7. Strength 18
8. Weakness 18
9. Letter to the Director of Top Sport 19
10. References 20
11. Appendices 24

3. Introduction

3.1. Background

In recent years, the phrase “Greatest Of All Time" (G.O.A.T) has been introduced. It is a phrase that indicates which player has the greatest performance of each sport; it plays an important role in the sports world and is brought up when generational athletes accomplish the incredible (Rank, A., 2019). For example, Michael Jordan (Basketball) and Lionel Messi (Football/Soccer) (“The G.O.A.T In Every Sport,” 2020). The G.O.A.T is currently widely witnessed among those who are enthusiastic about sports.

In this context, we divided sports into two types—first, individual: the one in which an individual competes directly or indirectly against other individuals, such as tennis, golf, boxing, swimming, gymnastics, and skiing—second, team sports: athletes compete cooperatively in a group of at least two people, such as baseball, basketball, volleyball, handball, and water polo (Reference, 2020). Nevertheless, Individual sports tend to build self-reliant athletes, whereas team sports teach people the value of working together. In individual sports, an athlete can measure his progress against his record, while team sports may encourage a sense of teamwork among its participants.

Due to the difference in the rules and the playing methods between each kind of sport, various factors that affect the players’ performance must be taken into consideration to find out how strong the G.O.A.T is. Consequently, we aim to create a model that can be used in every kind of sport based on the 2018 Grand Slam Results database, consisting of the 2018 Wimbledon Championships, 2018 French Open, 2018 U.S. Open, and 2018 Australian Open.

3.2. Restatement of Problems

The Problems of finding G.O.A.T:

- The most talented athlete might take a rest from sports for a long time, leading to a less skilled G.O.A.T athlete.
- Team and Individual sports are enormously different.
- In individual sports, the methods of finding G.O.A.T in one-on-one competition and inanimate-standard competition (rankings) are different.

These Problems lead to several essential factors that affect the G.O.A.T determination.

4. Regulations

4.1. Strategies and Tactics

Our model *does not* focus on athletes' ages, their experiences, and how well they played during practice. We only focus on what they did at the moment that they participated in competitions.

4.2. Point-Counting Method

4.2.1. Retirement is *excluded* in this model's calculation. If it occurred, the winner would pass to the next round immediately.

4.2.2. Foul is *excluded* in this model's calculation because the competition has not finished yet.

4.2.3. Ranking comparison in any competition must take players who have been away from playing their sports for a long time and returned to play in that competition into account

4.2.4. The decrease in performance of those players that are mentioned in 4.2.3. must be concerned in the GFinder model.

4.3. Fairness

Our model assumes that every single competition was fair, every match was judged correctly, every athlete and coach, including people who worked behind every game, obeyed the rules.

4.4. Surroundings

We suppose that every match occurred at the same field's level and environment, such as the weather's temperature, humidity, etc.

4.5. Physical and Mental Health of Athletes

In all competitions that occurred, we assume that all participants were in good physical and mental health; they could present their full potential. Hence, their health *did not* affect their winning.

4.6. Inappropriate data

Unknown and unsaved competitions' results are not used in calculations. Moreover, the data that we used for analyzing tennis' data and cycling's data must only refer from W.T.A Tour ("Current Elo ratings for the WTA tour," 2021) and Running (Elite Running Analysis, 2019), respectively.

4.7. Skill level of players change over time

We know the player if inactive they skill (performance, strength, etc.), the value of ranking will drop by the time, and the ranking equal but not in the same era the old will be smaller than too.

5. Factors

Our team has done some research and found several factors that influence the G.O.A.T analysis, but two stand out:

5.1. The results of competitions: Win-loss record of each player in each kind of sports

5.1.1 Winner-Loser

5.1.2 Score difference

5.1.3 Loser score

5.1.4 Tiebreaker match

5.2. The number and consistency of participated competitions :

How often a player participates in competitions, including muscle masses (Ogasawara, R., et al., 2012), playing intimacy which may lead to a decrease in muscle memory and performance.

5.3. Z-score of each participant

6. Analysis Data, Variables, Equations in GFinder Model

6.1. Analysis Data used:

1. W.T.A.'s ELO Rating ("Current Elo ratings for the WTA tour," 2021)
2. Elite Running Analysis ("Elite Running Analysis," 2019)

6.2. Equations and Variables in GFinder Model

6.2.1. GFinder in tennis sport (Task 1.a.)

ELO Ranking is a system that is based on the scores earned by the winner of a match. It is determined by the probability of defeating that he had before the match. (Mittal, 2020; "Elo Rating System," n.d.)

However, The Rating Value that is used in the GFinder model is applied from the ELO Rating System; it is the value that is made up for the strong consideration of each contestant in a competition, and then find the possibility that an athlete will lose or win in that competition. When an athlete meets an opponent with very low chances of winning; if he/she wins, he/she will deserve a more stamina change than winning an easier-to-compete opponent or an opponent with a lower RV.

GFinder highly focuses on the match's quality per game, leading to the interest in scoring in every match. Consequently, GFinder can tell not only the chances of winning but also the performance (or quality) of the match. Hence, our GFinder can rank the players' ranking accurately; for example, Mr.A, a talented player, competed 50 times and won 30 times, whereas Mr.B, another talented player, competed for 100 and won 60 times.

Even though Mr.B (60 points in total)has more total points than A (30 points in total), our GFinder model will present that they get equal rankings because their performances are practically the same.

Finally, let us emphasize that our GFinder believes that the players who are away from the competition for a long time, regardless of accidents or any other reasons, these players will be in lower-ranking, to make them suit for the next match and make the method of GFinder accurate.

Given:
ΔP = Point difference; P_f - P_i

$$\Delta P = k\,(w - w_e) \tag{1}$$

P_i = Initial Points (the score that an athlete receives when he wins any competition)
P_f = Final Points (the final score that is taken into consideration)
k = The maximum possible adjustment per game*;* how stiff or easy competition is, obtained from:

$$k = f \cdot x \tag{2}$$

Where:
x = the value retrieved from the normal distribution of the rank to find out how strong an athlete's opponent is, compared to other rivals to find out how much k factor of two competitors with various ranks should be.

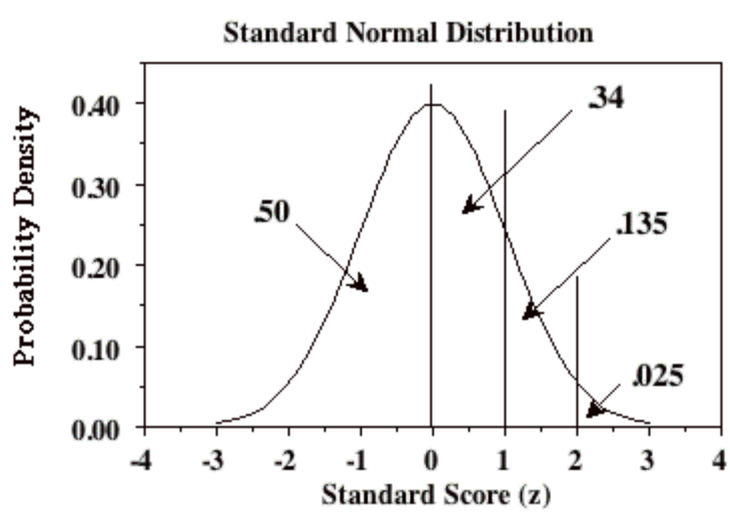


Fig.1: an example of the bell curve

From http://psych.fullerton.edu/mbirnbaum/psych101/grading_methods.htm

We have rated each player on the curve based on where a player stands compared to other players in the same competition, spreading across the entire competition of the match. Note that the curve is a bell curve that arises from each player's score comparing to each player's z-score

$$Z_{x(i)} = \frac{x\,(i) - \overline{x}\,(i)}{s_x}$$

...z-score of each player

Given that:

$Z_{x(i)} \geq 2$	; x=25	at 2.5% of competitors
$2 > Z_{x(i)} \geq 1$	; x=20	at 13.5% of competitors
$1 > Z_{x(i)} \geq 0$	; x=15	at 34% of competitors
$0 > Z_{x(i)} \geq -1$	; x=10	at 34% of competitors
$-1 > Z_{x(i)}$	; x=5	at 16% of competitors

After finding the x value,

In the game, the winner used the x value of the loser to fairness on calibrating, and the loser used the x value of the winner to fairness on calibrating too.

Why can the loser use the x value from the winner in $Z_{x(i)}$ if the winner is bigger than the loser?

The loser has a chance to win but the winner is so big, so this model can switch the x value if it has a battle.

This system's advantage is that every score is used to calculate the mean and the standard deviation. It compares each player to the average of other players in the same competition, causing this system to be the system that every single player must put an effort to try to earn one's point because every player wants to be the best player.

Moreover, we invented *x* value to give a fair and high-quality assessment of the changes in points; for example, if a player with a very high score wins (no matter how overwhelming the winning is), that winning should not receive as many points as a high score over an opponent with a close score to this player. Similarly, if a low-scored player loses a high-scored player, his score should not get deducted that much.

***f* = closeness variable is defined as:**

$$f = \left(\frac{d}{s}\right)^{\frac{1}{t}} \tag{3}$$

Where:

***d* = point difference**

d = 0 ; if there's no battle or any accident occurred.

d = 1.8 ; if a game won by 2 games ,

d = 2 ; if a game won by 3 games, and

d = 2 + 0.1N ; if a game won by N > 3 games.

***s* = tie-break;** Do they fight until a tie break?

s = 1.5 ; if there is tie-break round

s = 1 ; otherwise

***t* = loser's point**

t = loser's point - 10 ; if loser's point ≥ 12

t = 1 ; otherwise

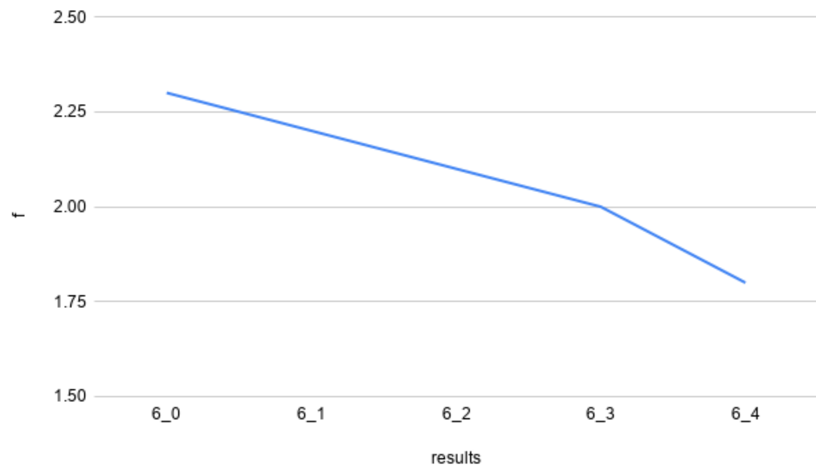
***w* = the result of the game** *(win = +1, loss = 0)*

w_e = winning expectancy from the following formula:

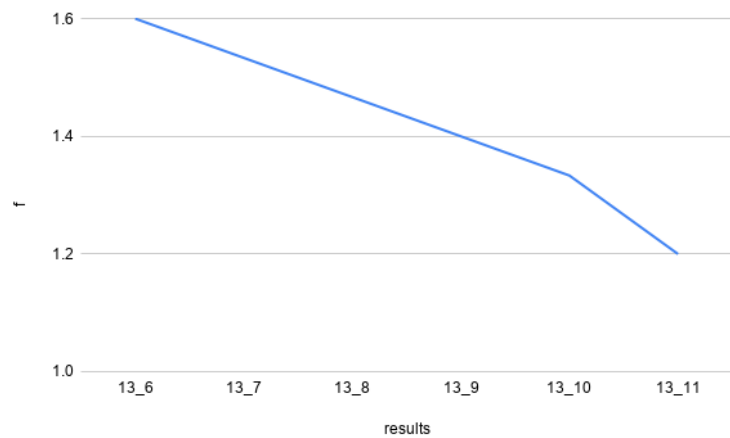
$$w_e = \frac{1}{1 + 10^{\frac{dr}{600}}} \tag{4}$$

where ***dr* = the difference in ratings**

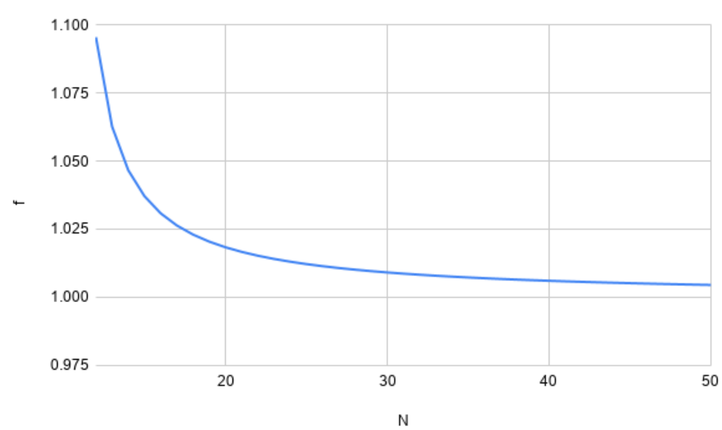
We applied the hockey scoring system, which defines its d-value from the score difference to our model's *d*-value. In our model, the *k*-value has a linear correlation with the *d*-value. Due to having Tiebreaker in tennis tournaments, it was known that both players had a close match. Therefore, the performances of them are roughly equal, meaning that k should be decreased. So the s value is defined as above to calculate the closeness of both players' performances. The reason why t value is defined and used in the f value as above is that f value must have the lower bound to prevent ideal match outcome, such as t value approach infinity, which equation (3) satisfies the condition.



Graph 1: indicates the value of *f* when the winner's point is 6, *created on Google Sheet*



Graph 2: indicates the value of *f* when the winner's point is 13, *created on Google Sheet*



Graph 3: indicates the value of *f* when the loser's point is

$N = \{1.2^{\frac{1}{N-10}}, N \in I^{+} | 12 \leq N \leq 50\}$, *created on Google Sheet.*

6.2.2. Inactive players' rating *(previously mentioned in 4.2.3.)*

When a player has no competition for more than 4 weeks, the ranking will start to decrease, because there will be new talented players appeared. It is significantly reduced as a regression function.

Considering the simple regression equation:

$$y = a + bx + e$$

Where :

x = Independent variable

y = Dependent variable

a = Constant

b = Coefficient

e = Residual

We saw that rating in the future will increase its value. So, GFinder should have a formula that represents the drop of rating. We know that each athlete should have a different rate of decrease in rating; all athletes' ratings will decrease until the rating is 1200, which is the default rating. Moreover, y-value is a variable that can either decrease or not in the current period. The logistic function is one of the functions that have the following conditions. So, GFinder will use the logistic function to represent inactive players' ratings.

When we calculated the regression equation when the players did not attend the competition as an equation as follows:

$$y = \frac{\varepsilon}{1 + 1.001^{x-i}} + m \tag{5}$$

Where:

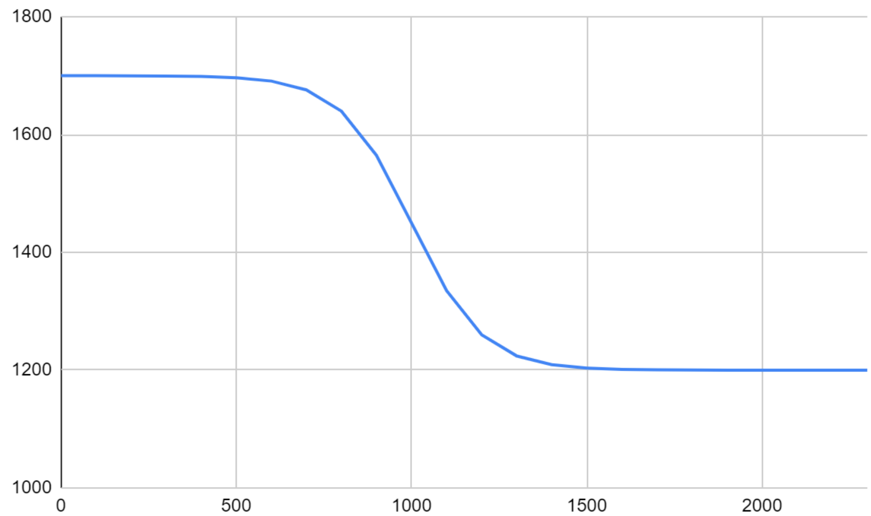
ε = (Ranking's score at the moment) - 1200

m = 1200

i = 1000, which is the value that causes the ranking to be constant for all 4 weeks before the regression.

y = New score, when the time passed x.

If y < 1200; the score is not increased, GFinder will not re-rank it, because it assumes that the score is in the standard range.

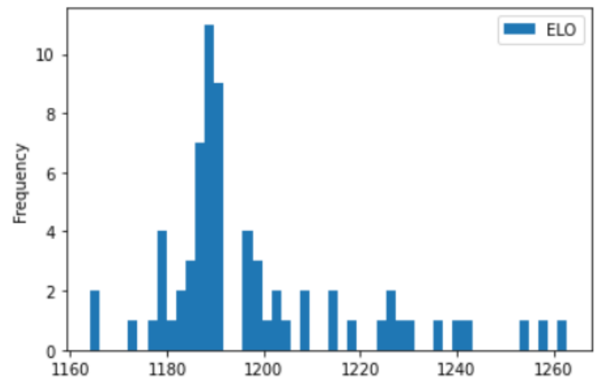


Graph 5: Our team's logistic function, *created on Google sheet.*

After creating GFinder for finding G.O.A.T of women's tennis players, we use GFinder to analyze tennis data from W.T.A.'s ELO Rating. The results are shown in Graph 6.

Results (Task 1b)

```
                 Player Name          ELO
1            Serena Williams  1262.889003
2           Angelique Kerber  1257.500083
3               Simona Halep  1254.375068
4        Caroline Wozniacki  1242.224992
5              Madison Keys  1239.365038
6               Naomi Osaka  1235.588106
7       Carla Suarez Navarro  1229.365096
8       Anastasija Sevastova  1229.040461
9            Sloane Stephens  1226.352943
10        Marketa Vondrousova  1226.132877
11        Alison Van Uytvanck  1224.726176
12           Garbine Muguruza  1217.907234
13              Julia Goerges  1215.043765
14           Yulia Putintseva  1213.763376
15  Barbora Zahlavova Strycova  1208.391308
Number of Players: 68
mean: 1198.2842415110297
STD: 21.086753210955703
```



Graph 6: Data Analysis of GrandSlam 2018, *created on Google Colab*

By analyzing GrandSlam 2018 the Fourth Round to the Final Round, the results show that the player that has the most rating (the G.O.A.T.) is Serena William

Analysis

From ranking in GFinder, we found that Serena William is the G.O.A.T of Grand Slam 2018, and if we take a look at her previous achievements, she had participated in the final round in the 2018 Wimbledon Championship and 2018 US Open. While the second place from GFinder is Angelique Kerber, she is the winner of the 2018 Wimbledon Championship. But in three other Grand Slam 2018 tournaments, Her ranking is not as high as Serena's. So, we will see that Serena William really had the most outstanding overall performance, and truly deserves the G.O.A.T.

6.2.3. Applying with Individual Sports (Running) (Task 2.a.)

We choose "running" as our sports proposal because it's an individual racing against more other players, not one-on-one. The winner is the one who can achieve the indicated target which is time measurement. That is different from the other individual sport like Tennis which is to achieve the highest scores over the other opponent by the restricted rules or called one-on-one competition.

Our model is to analyze the G.O.A.T of the racing sport by using the same methodology as the Tennis competition in 2018. By doing so, we specify the starting score of one player (or runner) at 1,200 scores. Then, after the competition, we will compare the actual final scores of each player; for example, selecting the one who gets the best scores among the 2 players. To gain or to lost scores from the competition will be as according to our initiated rules which are based on Elo-rating, as follow:

Given:

ΔP = Point difference; P_f - P_i

$$\Delta P = k\,(w - w_e) \tag{1}$$

P_i = Initial Points (the score that an athlete receives when he wins any competition)

P_f = Final Points (the final score that is taken into consideration)

k = The maximum possible adjustment per game; how stiff or easy competition is, obtained from equation 2:

$$k = f \cdot x \tag{2}$$

Where:

x = the value retrieved from the normal distribution of the rank to find out how strong an athlete's opponent is, compared to other rivals to find out how much k factor of two competitors with various ranks should be.

$Z_{x(i)} \geq 2$	; x=5	at 2.5% of competitors
$1 > Z_{x(i)} \geq 1$	; x=4	at 13.5% of competitors
$1 > Zx(i) \geq 0$	; x=3	at 34% of competitors
$0 > Zx(i) \geq -1$	; x=2	at 34% of competitors
$-1 > Zx(i)$	; x=1	at 16% of competitors

f = closeness variable is defined as:

$$f = e^{\left(\frac{1}{t} + \frac{1}{d}\right)} \tag{6}$$

Where:

d = time difference

t = time sum

w **= the result of the game** *(win = +1, loss = 0, draw = 0.5)*

w_e **= winning expectancy** from equation 4:

$$w_e = \frac{1}{1 + 10^{\frac{dr}{600}}} \tag{4}$$

where ***dr*** **= the difference in ratings**

The reason why *d* looks like the above equation is that we referred from equations in the Ice Hockey ratings, where *k*-value is linearly related to *d* over the *t*-value ("World Ice Hockey Elo Ratings," n.d.). Due to the low sum of time duration, their abilities are very close to each other. Therefore, we decided to use the above equation in the *k*-value configuration.

We choose "the power of $\frac{1}{t}$and $\frac{1}{d}$" because we want the graph with a low bound that competitors' scores approach infinity in the "ideal" case. Moreover, the graph where the relationships between *t, d,* and *f* in equation 6 satisfy all conditions. Thus, we use equation 6 because of the simplicity.

Finally, for the resting of athletes, the points often have a probability approach to the mean, causing the reduction of their ranking points. From equation (5):

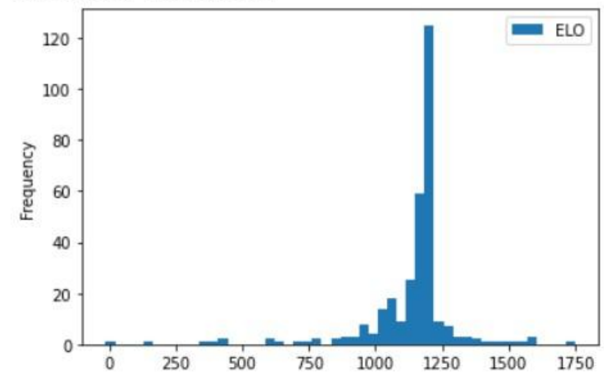
$$y = \frac{\varepsilon}{1 + 1.001^{x-i}} + m \tag{5}$$

For running sport ("Elite Running Analysis," 2019):
i = 700
m = 1200

After creating GFinder for finding G.O.A.T of 100-metre running players, we use GFinder to analyze tennis data from W.T.A. 's ELO Rating, the results are shown in Graph 7.

Results

```
             Player Name          ELO
1            Aaron BROWN  1747.795080
2           Femi OGUNODE  1603.383608
3    Henricho BRUINTJIES  1601.374829
4        Churandy MARTINA  1588.590601
5           Mosito LEHATA  1561.915804
6             Jeff PORTER  1508.067724
7           Peimeng ZHANG  1468.385486
8    Christophe LEMAITRE  1453.847930
9              Zhenye XIE  1413.115649
10            Yohan BLAKE  1391.682181
11       Johnathan CABRAL  1376.516031
12           Akeem HAYNES  1356.673393
13         Keston BLEDMAN  1345.183436
14           Luke BEZZINA  1335.250971
15       Gabriel MVUMVURE  1319.982266
Number of Players: 316
mean: 1141.0943304619261
STD: 175.57471014951219
```



Graph 7: Data Analysis of Summer Olympics 2004-2016, *created on Google Colab*

By analyzing Summer Olympics 2004-2016, the results show that the player that has the most rating (the G.O.A.T.) is Aaron BROWN.

Analysis

From ranking in GFinder, we found that Aaron Brown is the G.O.A.T of the above result, and if we take a look at his previous achievements at the Summer Olympics, he got 31st place on Men's 100m Rio Olympics 2016. While the second place from GFinder is Femi Ogunode, he got 37th place on Men's 100m Rio Olympics 2016. But if we consider Usain Bolt, he deserved the first G.O.A.T. from GFinder because he received numerous outstanding results from many athletics tournaments, such as Gold Medal in Rio Olympics 2016. Due to the lack of a creditable data source, we conclude that the dataset we used in this analysis contains incomplete data, which leads to an, inaccurate analysis.

6.2.4. Applying with Any Individual Sports (Task 2.b.)
We can find x by considering the scores' distribution of players in the bell curve in **Fig. 1,** After that, we score players of each position. This score can change into various ranges along the bell curve. By considering the positions of that player, compared to others on the same curve in the same part, distributed in all parts of the competition, compared to the z-score of each player.

The score criteria are evaluated depending on the competition rules as well that the players will consider the value of x at various percentage points.

In order to adjust to any kind of individual sports, the model must do methodology as follows:

1. In the case of individual **one-on-one** competition:
   - The mentioned model in 6.2.1. is required, but adjust other variables, such as d,s,t,x, we, to suit that sport.
   - For example, in some sports that include ties *(i.e., when both loser and winner get identical scores)*, GFinder must increase the w by +0.5 when the tie occurs.
2. In the case of the competition **against an inanimate standard**:
   - The mentioned model in 6.2.3. is required because it competes in the same way as swimming, but adjusts other variables, such as d,s,t,x, we, to suit that sport.
   - For example, in some sports with no tie-break round, there might not have an s variable in the k variable.

In our opinion, the ability of an athlete has an identical probability is decreasing, no matter ***what kind of sport*** is follows the given equation (5), but changes exactly the value of all variables to specify in that sport:

$$y = \frac{\varepsilon}{1 + 1.001^{x-i}} + m \tag{5}$$

6.2.5. Applying with Team Sports (Task 3)
To consider the model of team sports, we cannot use the "quality" of competition because there are two ways of playing team sports: first, everyone in the team plays the same role, such as double badminton, doubles tennis, and paddle; second, everyone in the team

plays a different role, such as football, baseball, and petanque. Hence, in the GFinder Model, we must consider the quality of the competition by focusing on the players' role in team sports.

Given that:

1. Every team member plays a different role from each other.
2. There are *n* members, plus *m* substitutes. Hence, there are *n*+*m* people per team.
3. For *n* members team, GFinder allocates their roles into *k* roles:
4. In calculations, we use equation (1): $\Delta P = k\left(w - w_e\right)$

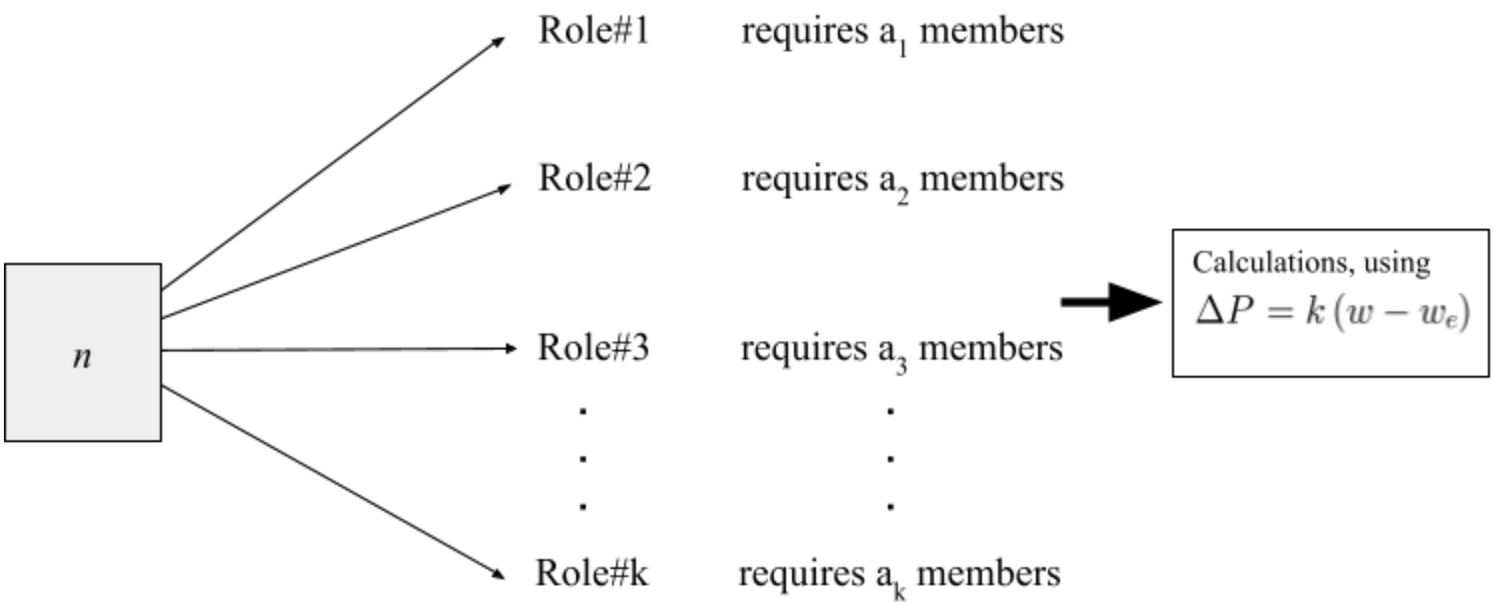


Fig.2 Diagram presenting the overview of GFinder for Team Sports, *created on Google Docs*

Hence, we can directly apply equation (5) to every kind of team sport because performances depend on an athlete not a team.

7. Strength

- ✓ GFinder can calculate the players' performance from "closeness," a factor other than winning-and-losing and the point difference. Closeness is measured from the scores or time duration that players did in a competition.
- ✓ To calibrate the rank of a player, GFinder focuses on how well an individual does his/her role or completes his/her goal.
- ✓ GFinder is designed to suit an individual. Hence, when applying to team sports, the model can find the G.O.A.T effectively by using the same methodology as used in individual sports.
- ✓ GFinder has two equations for calibrating the rank during the competitions and the absence of competition, to rebalance the score that could have a stronger effect in the future.
- ✓ This model isn't specific in gender (Both male and females can use this model).

8. Weakness

Winning-and-losing results not only depend on the athlete's skill but also other factors in the competition, such as the physical and mental health of athletes, surrounding, and so on, leading to the changes in the results.

However, some formulas used in GFinder are not directly calculated from the data, but they are obtained from analyzing the graph's trend to be close to the desired graph. So, those formulas may not be the best formulas to find the G.O.A.T.

Due to the change of skillset over time, our model contains an ELO decay part that will reduce the ELO by the inactivity. Therefore, our model can not precisely compare new players' skill level to old players' skill levels. GFinder will be inaccurate when used to find G.O.A.T on a wide range of years.

9. Letter to the Director of *Top Sport*.

Dear Sir,

Thank you for giving us an opportunity to design and propose a mathematical model as a solution for your program. We understand that, as a Director of Top Sport, you would be very passionate about sports. Then, you might have heard one keyword, "the Greatest of All Time (G.O.A.T)." It is the phrase that is so popular among those who are interested in sports. It means an athlete who is renowned for having the all-time most outstanding performance.

Accordingly, we are proudly presenting you our "GFinder", the model that we created over the past few days, which aims to find the G.O.A.T of any sport. Our team has done some research and found several factors that influence the G.O.A.T analysis, but three stand out: the results of competitions, the consistency of participating competitions, and the z-score. Furthermore, we've seen problems that need to be fixed in your G.O.A.T finding model, which are:

- The most talented athlete might take a rest from sports for a long time, leading to a less skilled G.O.A.T athlete.
- Team and Individual sports are enormously different.
- In individual sports, the methods of finding G.O.A.T in one-on-one competition and inanimate-standard competition (rankings) are different.

These Problems lead to several essential factors that affect the G.O.A.T determination.

To begin with, our GFinder is created by analyzing data on computer programming, based on some mathematics equations about statistics and data analysis, using the database from W.T.A.'s ELO Rating and Elite Running Analysis. Due to the difference between various kinds of sports; for instance, the difference in point-counting method the individual and team sports, we use GFinder to calculate the ELO Rating of the athletes that can be used in every kind of sport. We already test our GFinder, which has undergone trial and error to optimize results.

The attached file is the program that we used for calculating the ELO Ratings. You may use it by following the users' manual that we've prepared for you.

However, winning-and-losing results not only depend on the athlete's skills but also other factors in the competition, such as the physical and mental health of athletes, surrounding, and so on, leading to the changes in the results.

This may help you to better find the pertinent candidates for your program as it diminishes some pain points in your existing model. It would be highly appreciated if you could give us feedback on the outcome of using our program.

Best Regards,
GFinder's Team

10. References

About World Football Elo Ratings. (n.d.). Retrieved Mar 30, 2021, from eloratings.net: https://www.eloratings.net/about

Ace (tennis). (n.d.). Retrieved Mar 28, 2021, from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/wiki/Ace_(tennis)

Ashleigh Barty Stats. (2021). Retrieved Mar 28, 2021, from WTATour: https://www.wtatennis.com/players/318033/ashleigh-barty/stats

Atkinson, B. (2019, Feb 26). *Rating Sports Teams — Elo vs. Win-Loss.* Retrieved Mar 28, 2021, from Towards Data Science: https://towardsdatascience.com/rating-sports-teams-elo-vs-win-loss-d46ee57c1314

ATP Competition's Calendar. (n.d.). Retrieved Mar 30, 2021, from MSN: https://www.msn.com/th-th/sports/tennis/tour-calendar

Baback Vaziri, Shaunak Dabadghao, Yuehwern Yih, Thomas L. Morin. (2018). Properties of sports ranking methods. *Journal of the Operational Research Society, 69*(5), 776-787. doi:10.1057/s41274-017-0266-8

Building a Multiplayer Elo Rating System. (2012, Dec 16). Retrieved from gautamnarula.com: https://www.gautamnarula.com/rating/

Current Elo ratings for the WTA tour. (2021, Mar 22). Retrieved Mar 28, 2021, from tennisabstract.com: http://tennisabstract.com/reports/wta_elo_ratings.html

Daniel Barrow, Ian Drayer, Peter Elliott, Garren Gaut, Braxton Osting. (2013). Ranking rankings: an empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports, 9*(2), 187-202. doi:10.1515/jqas-2013-0013

David J. Irons, Stephen Buckley, and Tim Paulden. (2014). Developing an improved tennis ranking system. *Journal of Quantitative Analysis in Sports, 10*(2), 109-118. doi:10.1515/jqas-2013-0101

Elite Running Analysis. (2019, Aug 17). Retrieved Mar 30, 2021, from Github: https://github.com/mattjezza/ds-proj1-t2-elite-athletics

Elo rating system. (n.d.). Retrieved Mar 28, 2021, from Wikipedia, The Free Encyclopedia: https://en.m.wikipedia.org/wiki/Elo_rating_system

Elo rating system for runners? (2017, Jan 25). Retrieved Mar 30, 2021, from letsrun.com: https://www.letsrun.com/forum/flat_read.php?thread=8028181

Exploring WTA Players. (2018). Retrieved Mar 28, 2021, from Kaggle.com: https://www.kaggle.com/residentmario/exploring-wta-players

How long can we take a break from running?. (2020, May 4). Retrieved Mar 30, 2021, from runnology: https://runnology.com/

Jeff. (2016, Aug 15). *How Elo Solves the Olympics Ranking Points Conundrum.* Retrieved Mar 28, 2021, from Heavy Topspin, The TennisAbstract blog: http://www.tennisabstract.com/blog/2016/08/15/how-elo-solves-the-olympics-ranking-points-conundrum/

Kovalchik, S. (2017, June 14). *Why women's tennis rankings need an overhaul.* Retrieved Mar 28, 2021, from The Conversation: https://theconversation.com/why-womens-tennis-rankings-need-an-overhaul-78389

Leighton Vaughan Williams & Chunping Liu & Hannah Gerrard. (2019). How well do Elo-based ratings predict professional tennis matches? *NBS Discussion Papers in Economics, 03*. Retrieved from ideas.repec.org.

Marveldoss, R. E. (2019). An Elo-Based Approach to Model Team Players and Predict the Outcome of Games.

Michael Jordan. (2021, Mar 25). Retrieved Mar 28, 2021, from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/wiki/Michael_Jordan

Miller, C. (2019, Jul 30). *Grades Aren't Normal.* Retrieved Mar 30, 2021, from CURTIS MILLER'S PERSONAL WEBSITE: https://ntguardian.wordpress.com/2019/07/30/grades-arent-normal/

Mittal, R. (2020, Sep 11). *What is an ELO Rating?* Retrieved Mar 28, 2021, from Medium: https://medium.com/purple-theory/what-is-elo-rating-c4eb7a9061e0

Mirshams Shahshahani, P. (2018, May 24). *Downloaded IAAF Sprint Results in all Heats for 2004 - 2016 Olympics for both Men and Women.* Retrieved Mar 30, 2021, from Deep Blue Data: https://deepblue.lib.umich.edu/data/concern/data_sets/cr56n184r?locale=en

Ogasawara, R., Yasuda, T., Ishii, N. et al. Comparison of muscle hypertrophy following 6-month of continuous and periodic strength training. Eur J Appl Physiol 113, 975–985 (2013). https://doi.org/10.1007/s00421-012-2511-9

Pairwise comparison. (n.d.). Retrieved Mar 28, 2021, from Wikipedia, The Free Encyclopedia: https://en.m.wikipedia.org/wiki/Pairwise_comparison

Rank, A. (2019, Feb 18). *G.O.A.T of G.O.A.Ts: Ranking the best of the best in sports.* Retrieved Mar 28, 2021, from NFL.com: https://www.nfl.com/news/g-o-a-t-of-g-o-a-t-s-ranking-the-best-of-the-best-in-sports-0ap3000001018234

Saengja, T. (2020). *Large-Scale Network: A Scalable Learning Algorithm and Visualization.* Massachusetts: Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Sasala, M. (2020, May 30). *Greatest of All-Time: Maureen Connolly?* Retrieved Mar 30, 2021, from lastwordonsports.com: https://lastwordonsports.com/tennis/2020/05/30/greatest-of-all-time-maureen-connolly/

SoftwareEngenius. (2020, Aug 28). *Learn in 5 Minutes: Rating Systems (ELO, DWZ, TrueSkill).* Retrieved Mar 28, 2021, from YouTube.com: https://www.youtube.com/watch?v=nFKyRDkb60Y

Stanislav Dadelo, Zenonas Turskis, Edmundas Kazimieras Zavadskas, Ruta Dadeliene. (2014). Multi-criteria assessment and ranking system of sport team formation based on objective-measured values of criteria set. *Expert Systems with Applications, 41*(14), 6106-6113. doi:10.1016/j.eswa.2014.03.036.

STATS HUB. (2020). Retrieved Mar 28, 2021, from WTATour: https://www.wtatennis.com/stats/2020

Teeguarden, P. (1951, Apr 17). *Tour-Level Seasons.* Retrieved Mar 30, 2021, from tennisabstract: http://www.tennisabstract.com/cgi-bin/wplayer.cgi?p=PamTeeguarden

Tennis' Rules. (2014). Retrieved Mar 28, 2021, from Beaverpur Open Tennis: https://sites.google.com/site/beaverpuropentennis/home/tournament-rules

THE NEW ISL RATINGS SYSTEM. (2020, Sep 23). Retrieved Mar 30, 2021, from swimswam.com: https://swimswam.com/the-new-isl-ratings-system/

Three Systems for Grading. (n.d.). Retrieved Mar 30, 2021, from psych.fullerton.edu: http://psych.fullerton.edu/mbirnbaum/psych101/grading_methods.htm

World Ice Hockey Elo Ratings. (n.d.). Retrieved Mar 28, 2021, from EloRatings.net: https://www.eloratings.net/icehockey/about

Writer, R. S. (2020, Mar 26). *What Are the Differences Between Individual Games and Team Sports?* Retrieved Mar 28, 21, from Reference: https://www.reference.com/world-view/differences-between-individual-games-team-sports-1269ff584af52b2d

Yang, L. (2020, Apr 26). *Why Sigmoid: A Probabilistic Perspective*. Retrieved Mar 28, 2021, from Towards Data Science: https://towardsdatascience.com/why-sigmoid-a-probabilistic-perspective-42751d82686

[ZtarOZ]. (2020, Sep 4). *The GOAT In Every Sport.* Retrieved Mar 28, 2021, from YouTube.com: https://www.youtube.com/watch?v=asuIN7gUFAI

11. Appendices

Figure 1: Show the header code that gets the data to test from : [https://github.com/JeffSackmann/tennis_wta/raw/master/wta_matches_2018.csv]

```
import numpy as np
import pandas as pd
from scipy import optimize
import scipy.stats as stats
import matplotlib.pyplot as plt
from lmfit import Model
from datetime import date
import os
```

```
wta_match_2018 = pd.read_csv('https://github.com/JeffSackmann/tennis_wta/raw/master/wta_matches_2018.csv')
```

Figure 2: The table that shows a sample of the 2018 Grand Slam game (This is a data set is a sam in Github form Figure 1)

Figure 3: Sample of GFinder of tennis calculating code

```python
def elo_calc(winner_name, loser_name, score_str, date_now, dict_elo, dict_prev):
    who_won = 1

    ## Set new ELOs and change ELOs due to inactivity
    if winner_name not in dict_elo:
      dict_elo[winner_name] = 1200
      dict_prev[winner_name] = date_now
    elif dict_elo[winner_name] > 1200:
      date_prev = str(dict_prev[winner_name])
      dict_prev[winner_name]=date_now
      date_now = str(date_now)
      d0=date(int(date_prev[:4]), int(date_prev[4:6]), int(date_prev[6:8]))
      d1=date(int(date_now[:4]), int(date_now[4:6]), int(date_now[6:8]))
      delta = d1 - d0
      day_change = delta.days

      if day_change > 28:
        dict_elo[winner_name] = (dict_elo[winner_name] - 1200)/(1 + 1.001 ** (day_change/7)) +1200

    date_now =str(date_now)

    if loser_name not in dict_elo:
      dict_elo[loser_name] = 1200
      dict_prev[loser_name] = date_now
    elif dict_elo[loser_name] > 1200:
      date_prev = str(dict_prev[loser_name])
      dict_prev[loser_name]=date_now
      date_now = str(date_now)
      d0=date(int(date_prev[:4]), int(date_prev[4:6]), int(date_prev[6:8]))
      d1=date(int(date_now[:4]), int(date_now[4:6]), int(date_now[6:8]))
      delta = d1 - d0
      day_change = delta.days

      if day_change > 28:
        dict_elo[loser_name] = (dict_elo[loser_name] - 1200)/(1 + 1.001 ** (day_change/7)) +1200
```

Figure 4: Sample of Program that gives to the Director of TopSport code : [https://colab.research.google.com/drive/1AwIFdYDh5Q0nOMYpIiGvIJKiwFl4lHWZ?usp=sharing]

```python
print('*************** S T A R T - G F i n d e r ***************')

global letter_elo
letter_elo = {}
global letter_prev
letter_prev = {}

i=1
while 1 :
  print('Match %d :' %i)
  print('Enter match date : (YYYYMMDD)')
  tour_date=input()

  print('Enter winner name : ')
  winner = input()


  print('Enter loser name : ')
  loser = input()

  print('Does this match has a Tiebreaker? : (Y/N)')
  tie=str(input())

  if tie =='Y':
    print('Enter loser\'s tiebreaker score : ')
    st2 = int(input())
    score_mix = str(s1) + '-' + str(s2) +'(' + str(st2) + ')'
  else:
    print('Enter winner\'s score : ')
    s1 = int(input())
    print('Enter loser\'s score : ')
    s2 = int(input())
    score_mix = str(s1) + '-' + str(s2)

  elo_calc(winner_name=winner,loser_name=loser,score_str=score_mix, date_now=tour_date, dict_elo=letter_elo,dict_prev=letter_prev)

  i=i+1

  print('ELO Rating Updated Completed')
  print('Continue the Program? : (Y/N)')
  con = str(input())
  if con == 'N':
    break

print('ELO Ranking')
display(letter_elo)
```

```
What's your gender ? : (Male / Female)
What's your name ? :
```

Credits:

This document was created on Google Docs.