Cricket T20 International Match Prediction Report and Analysis

Antony Anuraj  
*Computer Science, University of Manitoba,* Winnipeg, Canada [anuraja@myumanitoba.ca](mailto:anuraja@myumanitoba.ca)

Ayush Patel  
 *Computer Science, University of Manitoba,* Winnipeg, Canada [patelad@myumanitoba.ca](mailto:patelad@myumanitoba.ca)

Darshan Pandhi   
*Computer Science, University of Manitoba,* Winnipeg, Canada  
[pandhid@myumanitoba.ca](mailto:pandhid@myumanitoba.ca)

Gurtej Boparai  
*Computer Science,* *University of Manitoba,* Winnipeg, Canada  
[boparai3@myumanitoba.ca](mailto:boparai3@myumanitoba.ca)

Ronak Vyas  
*Computer Science, University of*

*Manitoba,* Winnipeg, Canada

[vyasr@myumanitoba.ca](mailto:vyasr@myumanitoba.ca)

*Abstract* - Millions of people around the globe are fond of cricket. With such a big fan base, teams are sure to be competitive. This will lead to teams trying to get their players to play at their full potential. Cricket is a sport that has a large amount of data available per match. This data collected from matches could be used to provide evidence of why a specific match resulted in a certain outcome. Data related to team performance and individual player statistics can be retrieved from each match that is played no matter the format of cricket. This data could be analyzed and used to benefit a team to improve their player performances by focusing on a weakness to improve the skill in relation to it. However, sometimes factors like venue, toss winning, and ranking could be seen as data that could have an impact on the result of a match. In this research paper, we are going to implement sports data mining approaches to predict T20 international world cup winner. We will be using the dataset from cricsheet.org to examine/ analyze factors like venue, toss winning, team ranking, and matches won against a specific opponent to predict the winner of a given match.

Keywords: cricket, T20, world cup, data mining, sports mining, dataset, toss, venue, head-to-head, ranking.

# **Introduction**

Data mining is a process of extraction of useful information and patterns of data. In other words, it is also known as knowledge extraction or data and pattern analysis [11]. Data mining is also helpful when it comes to identifying trends and possible future insights about a specific topic based on the data provided. We currently live in a time where data is one of the key factors that could influence the direction of a field of study, business, sports, and other sectors of life. Understanding data and analyzing it could provide an added advantage that could help people excel in that area.

Sport is an area where each event contains many statistics that represent the performance of players/participants and teams. Sports mining is an area of data mining that can be seen as influential to sports-based activities. Especially those that have a competitive nature present in them.

Sports mining can provide a more in-depth analysis that could not have been seen in the past. The reason it is important to use sports mining is due to the reason that in the past coaches, managers and analysts used to use make decisions and predictions just based on their observation while watching the player/teams play [12]. These decisions made by these people weren’t supported with any data but just based on a trajectory that they thought the match would play out. These days data mining could help predict the outcomes of matches, provide in-depth analysis on a player based on past performances also coaches could use information collected from training sessions to identify areas that each player would need to improve to perform to their full potential.

Nowadays, the popularity of Cricket has grown very rapidly. “*To cater to potential future growth, global market research was commissioned by the International Cricket Council (ICC) which revealed that cricket has more than one billion fans worldwide, with the potential for significant growth*” [2]. Cricket is controlled by ICC “*The ICC is the global governing body for cricket. Representing 105 members, the ICC governs and administrates the match and works with our members to grow the sport. The ICC is also responsible for the staging of all ICC Events. The ICC presides over the ICC Code of Conduct, playing conditions, the Decision Review System and other ICC regulations*” [6]. As Cricket is growing rapidly people started to book their tickets early whether it is any kind of match.

In cricket, coaches could use stats retrieved from the data about past players’ performance to get insights on whether a player is likely to perform in a specific stadium or against a particular team. Coaches could also use stats like marginal wins, such stats could help model an ordinary team’s performance with or without a specific player, and this could help coaches perform necessary changes to the playing squad for a match [12].

Cricket is a sport that has a vast amount of historical and past game statistical data. We mine cricket data to have a better performance of the cricket players as the Sports management uses a tool to mine the data to select the players to have the best outcome of the results [8] as in the past the stat for cricket matches and the players have just been used for comparison purposes and were not explored greatly. However, with the rising importance of data in the world and advancements in technology. The data in cricket can be seen as a promising area waiting to be explored. With the increasing popularity of sports betting, cricket is seen as an ideal candidate for sports betting since it has a lot of statistical data based on each game played [3]. People are also increasingly playing on online fantasy sports platforms; people would be curious to know if their created team is going to when or if the players they are backing are going to perform according to their expectations. With huge leagues like IPL (Indian Premier League) having a massive fanbase, fans are driven towards online fantasy platforms providing prizes for people who win while playing on this online platform. The benefits of mining cricket-based data managers, coaches, and fans could keep track of player performances and be able to identify the trends in a player’s performance to predict the performance in the upcoming matches.

***1.1 CONTRIBUTION***

We used a new unique algorithm that we to implement for the purpose of this paper. It uses feature instead of player statistics that most of the other research paper have used. The feature that we considered for the purpose of this paper are:

* Ranking
* Head-to-Head matches
* Toss
* Venue (the stadium and the country that the match was played in)

This paper is organized as follows. The next section presents the background and related work. In Section 3 we introduce our Datasets. Section 4 represents the methodology. Analytical and Evaluation Results are shown in section 5. Finally, we present Future Work and Conclusion in Section 6.

# **Background And Related Work**

**2.1 BACKGROUND**

Cricket is one of the most popular team-based sports played around the world. Various cricket events take place throughout the year. These include tournaments between 3 or more teams, and series that usually involve two teams playing multiple matches against each other and the winner of the series being the team with the most number of matches won. Cricket matches are usually made up of two teams competing against each other. Each team consists of 11 players. There are 3 common formats that cricket is played in which are: Test matches, ODIs (One Day Internationals), and T20s.

In the 16th Century, the match of cricket was originated in East London and later it developed more and became the most popular match in the world [13]. The T20 match lasts for about 3 hours as this match was presented to attract the spectators at the stadium. As in the T20s format, 2 batsmen are open to start the innings. As these 2 batsmen start to open the match the other and remaining batsmen are ready and waiting in the dressing room for their turn to play the match [10] but one cannot predict the match result up to and until the last minute of the match [13].

For this paper, we will be only considering the T20 format of cricket. T20 matches are the most played format of cricket. It involves two teams playing against each other to see who can make the most runs (runs are a way of keeping the score of each team).

Each team will have a maximum of 20 over to bat and make runs while the other team bowls and vice versa.

* Each over consists of 6 balls, and 1 bowler usually bowls all 6 balls one at a time.

At the start of the match, a toss is done. The toss-winning team gets to decide whether to bat first or bowl first. The first team that bats will have a goal of trying to set a high score by trying to make the most amount of runs that they can.

The batting side will send out batsmen that will have to make runs for the batting side. Runs are counted per ball bowled.

In the 20 over match of the T20 cricket the first 6 overs that is 1-6 overs are the power play.

A power play is used in all cricket formats with the fielding rules which means that the two fielders are allowed outside the circle and the remaining fielders are inside the circle during the 1-6 overs. In that, the batsmen have more advantage to score more runs during the period. Once the 6th over is completed all the fielders are spread out in the circle.

* Runs range from 0-6 per ball (additional runs could be got from a penalty given to the bowling side)
* Runs are usually achieved by the two batsmen on the pitch if they successfully run to each other’s side without getting run out (both batsmen should run across the pitch towards each other and the opposite stumps and cross the line before the bowling team hits the stumps towards which the batsman is running). For each time they successfully run between the wickets (stumps), the batting side will achieve 1 run.
* Then 4 and 6 runs can also be earned per ball bowled by successfully hitting the ball to the boundary. 6 runs if the ball goes directly over the boundary, 4 runs if the ball bounces at least once inside the field before it crosses the boundary.

The goal of the bowling team would be to get the batting team’s players out or prevent the team from setting a large score. A bowler can take a batsman out by either:

1. Bowling at the batman and the ball hitting the wickets behind the batsman.
2. Batsman hitting the bowl and the ball being caught by the bowling team before it hits the ground (this requires the player catching the ball to be within the playing area and not beyond the boundary)
3. Hitting one of the two wickets (stumps) on either side of the pitch before the batsmen try to cross the line to achieve a run.

The team that bowled first will now have to try making more runs than their opponent within the allocated 20 overs or less, if possible, to win the match.

The result of a match can be affected by a lot of factors such as the venue at which the match was played, the team that won the toss, the strength of the team in comparison of the other team (refers to team ranking), and other factors. The reason factors like toss are important is because it allows the toss winners to play according to the conditions of the field and pitch, they are about to play on.

**2.2 RELATED WORK**

Cricket is one of the sports that are extremely rich in statistics. Statistics of cricket are available for individual players and the whole teams as well. For this paper, we are focussing on team-based statistics. Often, these statistics are used to predict the winner of the upcoming matches between two teams. A similar approach called Sabermetrics was implemented to use the statistics meaningfully in baseball. [5] Sabermetrics was the analysis of statistics to determine which metrics contributed towards a team win. Like baseball, a bat and a ball are needed to play the match of cricket. But other than that, both matches are totally different from each other in the way they are played. Cricket is a more diverse sport than baseball and it is played in different formats across different countries. Inspired from Sabermetrics, we have tried to find out which factors (toss, venue, ranking, previous matches) contribute the most to a team’s win.

There is a lot of research done to find out which factors contribute the most to a team’s win. And most of the research papers have found that factors like toss result and venue of the match have a significant advantage for one of the two teams. However, the research paper, *Winning the Coin Toss, and the Home Team Advantage in One-day International Cricket Matches [4],* concludes the following:

1. Winning the coin toss provides no competitive advantage.
2. Playing at a home ground is an advantage for the home team.

On contrary, we analyzed that the coin toss does play an important role in a team’s win. Based on the 1134 matches played among 87 countries, the probability that a team will win a match if they win the toss is about 60%. We will talk more about toss evaluation in the methodology section. However, we completely agree with the 2nd point, and it will be discussed further.

K.Kapadia et.al. used T20 matches data to make a machine learning model to predict the winner of a match. The prediction is based on the historical match data considering the factors like home advantage and toss decision of a team. Their Naïve Bayes model predictions have 57% accuracy and 60.5% precision. [7]

Other than the interest in international T20 matches, many researchers have worked on the data from the Indian Premiere League (IPL), that is a cricket league based in India. Since many players from across the globe take part in this league, IPL is considered as one of the best cricket leagues in the world. Similarly, Praveen Banasode et.al did research to predict results of IPL T20 matches.[1] The goal of their paper is to use past matches data and analysis them to find the most useful detail to predict the outcome of a match. They focused on which players play in a match, which team wins the toss and other information about the match. Their research claimed over 95% accuracy and mentioned that the prediction can be done by considering individual player’s performance and the performance of the whole team.

As of now, we have mentioned about research that were done to predict the match winner before the match started. Interestingly, a lot of interest has been seen in predicting the outcome of a match while it is in progress. Nimmagadda Akhil et.al developed a model to predict the result of a cricket match while it is being played.[9] They used data of previous matched played between the two teams to design their model using Multiple Variable Linear Regression. Using this model, they predict innings score and the winner of the match at regular intervals.

There are a lot of similarities between these related works and our paper. Such as similar interest to predict the match winner in the shortest form of cricket, that is T20. The craze for the shortest form is motivating researchers to continue doing research in this field. Moreover, crucial factors like toss result and venue of the match in addition to historical match data are considered in almost all the papers. These factors are foundations of our algorithm too. The major difference here is that most of these research papers predict result of a match using individual players data. In contrast, we use whole team data instead of individual players data. The reason behind this is the uncertainty in the combination of 11 players playing in a match. Nowadays, the playing 11 of a team is not known until the toss is done. The uncertainties simply come out of player’s unavailability and injuries. Hence, if a team plays their best combination on any given day, the team’s performance represents their individual players performances.

# **Datasets**

As this paper focuses on the cricket T20 world cup prediction we would provide information on datasets and some background-related work. The Dataset used in the paper is from the cricsheet[[1]](#footnote-2).org. The T20 international world cup matches are played between the years 17th February 2005 – 16th October 2021. We have trimmed the ongoing world cup matches from after 16th October 2021. From the Cricsheet website, we have selected the T20 internationals gender men to type new so it will contain the CSV zip file which will have all the data. There are a total of 1134 matches played from the year 17th February 2005 – 16th October 2021. There is a total of 87 teams/countries playing in the T20 world cup are shown in figure1.

|  |  |  |
| --- | --- | --- |
| Spain | Rwanda | United Arab Emirates |
| Uganda | Sri Lanka | Bangladesh |
| Portugal | Finland | Germany |
| England | Malta | Bulgaria |
| Namibia | South Africa | New Zealand |
| Australia | Luxembourg | Guernsey |
| Malaysia | Singapore | Bermuda |
| Papua New Guinea | Kenya | Hong Kong |
| Mexico | Scotland | Romania |
| Ireland | Denmark | Zimbabwe |
| Nepal | India | Czech Republic |
| Pakistan | Thailand | Oman |
| Jersey | Italy | Ghana |
| United States of America | Netherlands | Canada |
| Lesotho | Seychelles | Malawi |
| Panama | Isle of Man | Nigeria |
| Hungary | Gibraltar | Norway |
| France | Belgium | Serbia |
| Greece | Austria | Kuwait |
| Qatar | Maldives | Saudi Arabia |
| Bahrain | West Indies | Tanzania |
| Vanuatu | Cayman Islands | Botswana |
| Philippines | Afghanistan | Eswatini |
| China | Turkey | Indonesia |
| Myanmar | Chile | Brazil |
| Saint Helena | Bhutan | Mozambique |
| South Korea | Argentina | Costa Rica |
| Japan | Samoa | Sweden |
| Fiji | Peru | Belize |

Table 1: Countries Playing World Cup T20.

In that CSV file, we have 2 types of data one is the match id.csv and the other is the info.csv in the Match id.csv that file has the match id of the two teams and in the info.csv the file contains the details of the match that is being played.

There are different Venues where the T20 matches had been played as shown in Figure 1.



Figure 1: showing two Countries Playing World Cup T20 with their Match Id’s.

Also, there are different venues for each team where the t20 matches are being played.

# **Methodology**

This section focuses on the system of methods used to predict the match outcomes depending on a few of the chosen factors from many that have small or large influences in cricket matches between two teams. These factors have been carefully analyzed to help us figure out what data would be required to be collected from datasets. Also, analyzing these factors helps us figure out how to use the data collected to make predictions of a particular match between two teams in the match of cricket.

To be able to predict something, past actions are considered to better tell the future (predict) based on previous experiences. Similarly, the datasets as mentioned above were obtained from “Cricsheet[[2]](#footnote-3)” Which provided us with data from all the relative cricket matches played in the past. Specifically, the 20 overs match format that we are focusing on.

The chosen factors that were considered to help us predict matches are namely: a) Team rankings, b) Head-to-head Matches, c) Venue Evaluation, and d) Toss Evaluation

The technique used with these features for a match prediction is that the prediction outcome is based on examining each feature individually.

We use the team rankings to get a prediction only if the given two teams have never played a match, and we do not have any data of them playing against each other. We will move on to use the other features for prediction only if a user provides us with certain information such as which team won the toss (Toss Evaluation) and where the game is played (Venue Evaluation).

If the follow up prediction, which are Toss Evaluation and Venue Evaluation, is not a significant advantage for the team that won the toss or is playing at a certain venue, we rely on the results from Head-to-head matches and use the result to get the final winner. Hence, the toss and venue evaluation predictions mostly help us try to get a better accurate prediction if a team carries the advantage of winning a toss or playing at a certain venue. Therefore, Head-to-head matches would be the most important part of our prediction technique that we rely on for most of the matches to be predicted.

So far, we have seen the technique behind what feature is used and when to give a match prediction, now we will focus on why and how each feature is used to come up with a prediction percentage of win.

## **Team Rankings**

Team rankings is a necessary aspect in most of the sports especially cricket in that it gives a glimpse of how a particular team has been doing over a period. The International Cricket Council (ICC) is the world governing body of cricket. ICC does the necessary calculations and keeps a track of the team rankings to determine where these teams are in terms of performance.

The t20 format world cup matches are held after every two years. Therefore, world cup matches alone would not be enough to rank teams hence every series of t20 matches that are played between any two teams between these two years are taken into consideration to calculate their rankings. Rankings are updated in two ways:

1. **Updated after every match played**

After every match played between any two teams, their rankings are updated within 24 or 48 hours accordingly on whether the team won or lost.

1. **Updated annually**

Annual ranking works by assessing the teams’ performances over a period of three years at the end of every year. Therefore, at the end of one year, the previous three years are considered including the year that has just ended. Thus, this eliminates the oldest year with addition to the newest year that just ended. This way all the teams’ performances are looked at over the period of three years and their rankings are recalculated.

Since these rankings determine which team has been at their peak and which team might be slacking off, it acts as a great factor for predicting a match outcome. Through this, we can tell which team has a higher chance at winning against another. We have assembled a file which carries a list of team rankings in an ascending order by collecting and mining t20 rankings data from our datasets. From this file, we can compare rankings between any two teams and calculate a prediction percentage of win for both teams.

The prediction percentage of win is calculated using the formula below:

**For Team A:**

**For Team B:**

This provides a prediction percentage of win for both teams by relying on their rankings as of the day they would be playing against another team.

## **Head-to-head Matches**

Data is gathered every match played between two specific teams. Among this data, includes head-to-head matches played in the past which helps keep a record of wins and loses a team has against another. Such data is usually recorded for the team to be able to assess their performance over the years and learn from it as they move on to face the same team in the future.

Similarly, head-to-head matches provides data for us to predict a future match between two certain teams. From our dataset, we can pull out a list of all Match IDs that represents matches played by two certain teams and analyze the number of wins and losses against each other. The information obtained from these would help us determine which team has had a greater number of wins than the other thereby increasing its chances of winning in a future match.

The head-to-head algorithm works by taking a list of Match IDs that represent matches played by the two certain teams and determining the number of times both teams won against each other. Thereafter, a calculation is performed for each team to get the percentage of win against each other.

Eventually, the prediction percentage of win is calculated using the formula below:

**For Team A:**

**For Team B:**

This provides a prediction percentage of win for both teams which would show the chances of either teams winning when playing against each other reflecting on their previous encounters.

## **Venue Evaluation**

The venue of the sport is another important aspect in predicting a cricket match. In a world cup event, all the matches are played in one chosen country.

A team that plays in its country’s particular ground often, carries the advantages of knowing the pitch well. They have trained on them for a long time just enough to know how well the pitch might be suitable for either bowling or batting or even both. An instance of this would be such that the team players, specifically bowlers, are able to adapt on how to bowl and work their way out to get the opposite team batsmen out. The same applies to the team’s batsmen who can know what shots to play and what not, to hit as many runs as they can.

This is also true for teams playing at their home ground who would carry an advantage of knowing their grounds well.

This might seem to place a huge advantage for the team playing often in a certain venue, however, it is not the case always. The opposing team playing might just be strong enough to even out the advantages and pose an equal or more chance of winning the match.

The venue algorithm works by collecting all the venue data from the datasets. We have created a file that contains the list of all Match IDs (every match played to date of this cricket format) and the country (with the specific venue) the match was played in. Furthermore, a whole list of Match IDs of a particular team is obtained and matched with the venues it has played in to determine the total number of matches that were played at a certain cricket stadium.

Eventually, we investigate each Match IDs of the team to see how many matches they won at a certain venue. The prediction percentage of a team winning if it plays at that certain venue is achieved by the formula below:

For matches played at a certain venue:

This provides a prediction percentage of win of a team depending on the team’s encounters on a particular venue

## **Toss Evaluation**

Toss evaluation involves a coin toss where the team that wins the toss gets to choose whether to bat or bowl first. This is one of the main factors chosen for predicting a match because winning a toss is not just about choosing what to do first, it can be a team’s strategic way to win a match. This is because the pitch they play on is analyzed and can be determined if its condition is favorable for batting or bowling. Depending on this, the toss-winning captain of the team can choose wisely whether to bowl or bat first. Another reason could be evaluating and knowing the opposite team’s strength whether they are good in either bowling or batting, the captain would then have to choose accordingly to their advantage.

We collected toss data from datasets to evaluate a team’s experience depending on whether they won or lost a toss. An algorithm is used to obtain a list of all Match IDs (each match played by the team to date of this type of cricket format). Thereafter, we obtain a list of all toss results of the matches that pertains to this certain team. The number of times the team has won a toss is accounted for. Finally, we also account for the number of times a team has won after winning a toss.

Eventually, the prediction percentage of a team winning is calculated by using the formula below:

This provides a prediction percentage of win either low or high indicating the chances of a team winning the match if they win the toss based on their previous toss results.

# **Evaluation And Results**

Chart, bar chart

Description automatically generated

Figure 2: Graph showing the percentage of matches won by a team that also won the toss

Chart, bar chart

Description automatically generated

Figure 3: Graph showing teams have significant advantage when they play at home

**5.1 Toss**

Toss result was as we expected. The percentage of matches won when a team wins the toss comes out to be 57.59%.

Percentage of matches won by a team that won the toss: 57.59 % (average of 14 teams so) – will be utilizing all the teams to come up with a better average.

However, since it is greater than 50%, we can conclude that winning the toss does give you an advantage. And that is exactly as expected. Afterall, the lucky captain gets to decide whether they want to bat or bowl first. And they can make the call according to the weather, pitch, team’s strengths, opponents’ strength/weakness, or their whims.

**5.2 Algorithm Testing**

For the T20 Cricket World Cup there are 3 stages to win for the world cup. For the First Stage also known as First Round where there is a total number of 8 teams being split into the two groups, so each group contains 4 teams named as Group A and Group B. The Next Stage is the advance stage also known as the Group stage that is Group 1 and Group 2 where only 4 teams qualify from the First Round that is from Group A 2 and from Group B 2 and they play with advanced teams.so in this stage 4 teams qualify plus with 8 advanced teams so total of 12 teams each has 6 teams each in 2 groups. Once they qualify for the Group stage the teams move towards to Semi Finals, so in this two of the topmost teams from Group 1 and two topmost teams from Group 2 play with each other to fight against for the final round. For the Final Round only 1 team gets selected from Group 1 and 1 team gets selected for Group 2 and the winner is awarded for the winning team.

As For the T20 World cup prediction, First, we have taken the current t20 world cup matches for prediction that is from 17th October 2021 until 14th November 2021 so for the first round before the super 12 or the Group Stage so there are 8 teams from 2 groups that is based on current matches Group A contains Sri Lanka, Namibia, Ireland and Netherlands and Group B contains Scotland, Bangladesh, Oman, and Papua New Guinea. From these Groups only top 2 teams would get selected in the Group stage. As from the official cricket T20 world cup standings from the Group A Sri Lanka and Namibia got selected for the group stage whereas for the Group B Scotland and Bangladesh got selected. Once the teams get selected for the Group stage these 4 teams need to play with the fixed advanced teams. So, a total of 12 teams will now be split into two groups called Group 1 and Group 2. Group 1 contains based on current T20 matches England, Australia, South Africa, West Indies these are advanced teams plus the Sri Lanka and Bangladesh.

Group 2 contains Pakistan, New Zealand, India, and Afghanistan as these are advanced teams plus Namibia and Scotland. Once these all teams have been played so the Topmost 2 teams from the Group 1 and Group 2 would get selected for the next Stage which is the semi–Final Round. In these based on current T20 match from 2 Groups each has 2 teams in each Group that is from Group 1 England and from Group 2 New Zealand these two teams play with each other, and the winner team goes directly to the Final match which is England and from other Group 1 Australia and from Group 2 Pakistan and the winning team among both is the Australia so there is a Final match between the England and Australia. Based on the Actual Winner of the T20 world cup the winner team is Australia.

Based on our Prediction Algorithm for the First Round we got for the Group A that contains Sri Lanka, Netherlands and for Group B we got Bangladesh, Scotland. As we see that for Actual teams in Group A that got selected is Sri Lanka and Namibia as Namibia has lost some matches based on our prediction algorithm. For Group B based on our prediction algorithm we got Bangladesh and Scotland. Moving into next stage which is the group stage the 2 teams that qualify from both the Groups were Sri Lanka, Netherlands and Bangladesh, Scotland to play against the advanced teams. So, In Group 1 based on prediction algorithm we got England, Australia, South Africa, West Indies plus Sri Lanka, and Bangladesh for Group 2 stage based on our prediction algorithm we got Pakistan, India, New Zealand, Afghanistan plus Netherlands, and Scotland. But apparently only 2 teams got selected from Group 1 For the semifinal round based on our algorithm the team which is Australia and Pakistan played together and Pakistan is the winner, and it moves directly to final match whereas for the other Group which is the West Indies and New Zealand, and New Zealand moves to final to battle against Pakistan so there were 4 teams in the semifinal round. And finally, our Prediction Algorithm predicts with the two teams and Pakistan is the Winner for the T20 World Cup shown in Figure 3

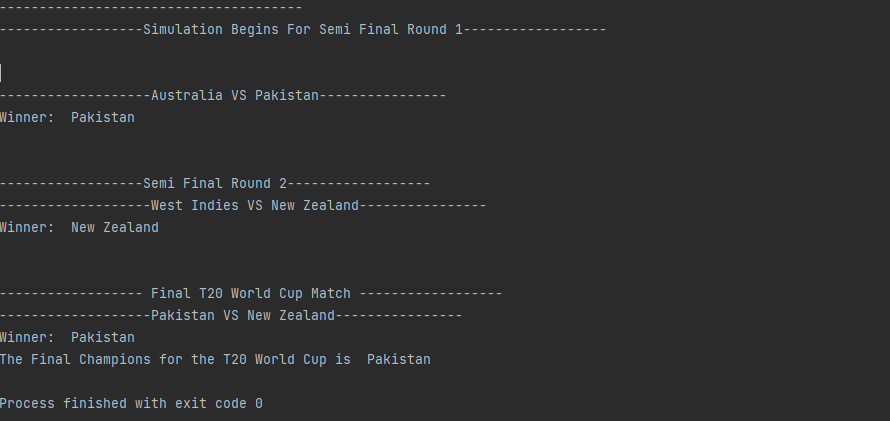
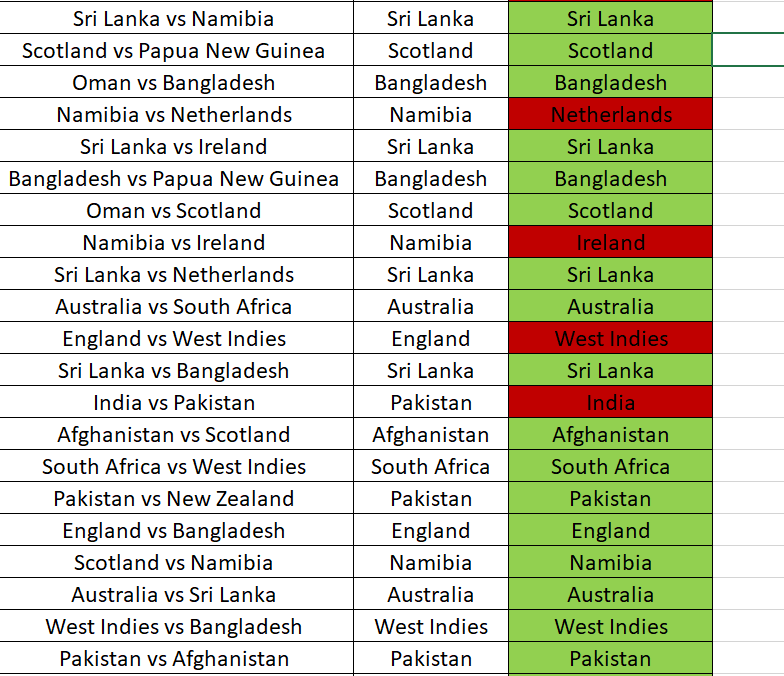


Figure 4: Showing the Cricket T20 World Cup Winner.



*Figure 5: Showing the Cricket T20 World Cup 2021 simulation*

# **conclusion and Future work**

For this paper, as we plan to predict the Cricket T20 international men’s but for future work, we would also add more features as our current features for this paper is:

* Team Rankings
* Toss Evaluation
* Venue
* Head-to-head matches

But for future work, we would like to add more features such as:

* Weather Conditions
* Most Runs scored on the Power play
* Pitch Conditions
* Crowd
* Highest Runs
* Most No of Catches
* Most No of Wickets Taken
* Most Sixes and Four’s
* Batting Strike rate and Average
* Bowling economy rate

Additionally, we would also be likely to further increase our datasets as it is up to 16th October 2021 so that we can further analyze and predict based on our data. Furthermore, we would also like to predict this technique based on different Cricket Format’s such as:

* ODI (One Day International) Cricket Match.
* Test Match.
* IPL (Indian Premier League).
* ICC Champions Trophy.
* PSL (Pakistan Super League).

Although we are predicting for men’s T20 international in this research paper but in the future, we would also be likely to have a prediction of the women’s T20 international and women’s ICC Champions Trophy for further analyzing.

In conclusion, we believe that our algorithm has a lot of potential since it predicted the actual winner for most (approx. 79%) of the fixtures from the recent world cup. However, it failed to predict the correct winner of the world cup. But we do believe that Pakistan were definitely the favorites to be the champions this year because of their performance in recent years. Apart from their performance, the venue (United Arab Emirates) of the world cup was an added advantage to their campaign since they play most of their matches at this venue. Having said that, we thrive to improve our algorithm and make it more accurate. We plan on considering factors like weather and pitch conditions in the future work because it was one of the deciding factors in this world cup.

##### **Acknowledgment**

I would like to thank my instructor Carson Leung for teaching this course and TAs for helping students throughout the course. I would also like to thank the authors of references that we used for this paper.

##### **References**

[1] Banasode, P., Patil, M., & Verma, S. Analysis and Predicting Results of IPL T20 Matches <https://iopscience.iop.org/article/10.1088/1757-899X/1065/1/012040/pdf>

[2] Basit, A., Bux, A.M., Majdah, A., Memon, K.H., Shah, R.A., & Hassan, J.F. (2020). ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques. 10.13140/RG.2.2.31021.72163.

[3] CA Editor (2021). How Big is the Cricket Betting Market Globally?. Yahoo!sport. 19 August 2021 <https://uk.sports.yahoo.com/news/big-cricket-betting-market-globally-123400027.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAABOwSKX7YyV_NxqN2C9Wk7kcPHxhaYWgvgzqfHlliL13ZH3BdRWqCTj7W4xywUKaMInNNQJbjJaEI12tgwwtDrvT4N4fUHgjtpQJpWSfAyIjZInYcZqJLnJ5CSPtfhEZ0S-nhXIjMT09g3SmWZLIH8xhK3D8lc0TPYaubQ_9kG9Q>

[4] De Silva, B.M., & Swartz, T.B *Winning the Coin Toss and the Home Team Advantage in One-day International Cricket Matches* <https://www.sfu.ca/~tswartz/papers/cricket.pdf>

[5] Gregory, P.A., Herath, D.H., Karunasekera, D.S., Deegalla, D.S., & Bandaranayake, A.U CricketSabremetrics- A Data Mining Analysis of Cricket [**http://www.nastec.gov.lk/files/ysf\_proceeding/ysf\_proceeding\_6th/8\_CRICKET\_SABREMETRICS\_A\_DATA\_MINING\_ANALYSIS\_OF\_CRICKET.pdf**](http://www.nastec.gov.lk/files/ysf_proceeding/ysf_proceeding_6th/8_CRICKET_SABREMETRICS_A_DATA_MINING_ANALYSIS_OF_CRICKET.pdf)

[6] Gulati, A., & Mutigwe, C. ‘Are Today’s Test Cricket Batsmen Better Than the Greats of Yesteryears? A Comparative Analysis’. 1 Jan. 2021: 139 – 153.

[7] Hasanika, D., Liyanage, D., & Dilhara, R. Data Mining System for selecting a Winning Cricket Team <https://cepdnaclk.github.io/e15-4yp-cricket-analysis/>

[8] Kumar, S Data Mining for Performance Analysis in Cricket <https://analyticsindiamag.com/data-mining-performance-analysis-cricket/>

[9] Nimmagadda, A., Kalyan, N.V., Venkatesh, M., Teja, N., & Raju, C.G. Cricket score and winning prediction using data mining <https://www.ijarnd.com/manuscripts/v3i3/V3I3-1230.pdf>

[10] Patel, Dr B.,& Bhathawala,Dr P. (2014). ICC WORLD TWENTY20-2014 (WORLD CUP-2014) International Journal of Mathematical Sciences and Applications Vol.5 No 1. January-June 2015.

[11] Ramageri, B.M (2010). Data mining techniques and applications. Indian Journal of Computer Science and Engineering Vol.1 No. 4 301-305

[12] Schumaker, R.P., Solieman, O. K., & Chen, H. (2010). Sports Data Mining (1st ed. 2010.). Springer US. <https://doi.org/10.1007/978-1-4419-6730-5>

[13] Shah, S., Hazarika, P.J., & Hazarika, J. (2019). A Study on Performance of Cricket Players using Factor Analysis Approach. International Journal of Advanced Research Vol 8 No.3 March-April 2017 656 - 660.

1. Cricsheet: <https://cricsheet.org/downloads/> [↑](#footnote-ref-2)
2. Cricsheet: <https://cricsheet.org/downloads/> [↑](#footnote-ref-3)