Cricket Game Predictor

Term Paper

Salman Karim (p13-6032)

Muhammad Mohsin (p13-6016)

Qaisar Abbas (p12-6016)

Session 2014-2018

Submitted to

Dr. Muhammad Nauman

DEPARTMENT OF COMPUTER SCIENCE, NATIONAL UNIVERIST OF COMPUTER EMERGING SCIENCES,

PESHAWAR CAMPUS

(May, 2018)

ABSTRACT

Cricket is one of the popular sports played in the world. It is ranked as the second most watched game in the world. The Cricket game has produced multi million revenue for countries. There is a lot of interest in predicting the outcome of a cricket match in their corresponding formats like one-day-International, T20 and Test Match. Predicting the game of cricket has not been easy. Apart from the skills of players there are many other factors which play vital role in winning a cricket match including Natural factors like weather conditions.

Prediction systems for different sports have already been developed been but it's yet to be done for the game of cricket. We will develop a Cricket predictor which will predict the game before it's been played based on teams records.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Developing a predicting system for the game of cricket is of great importance in order to guide teams for strategy making. We will develop a system which will predict a cricket game before it is been played. Advances in technology has made it easy to store large sets of data on simple devises. Luckily, past match records for cricket games in all formats like One Day International (ODI), T20 and Test matches are also available. We will collect all the match records and find out different patterns which can lead a team towards victory. Although Different approaches have been used for predicting the game of cricket but we will develop a more consistent system which will predict a game of cricket more accurately as compared to the previous systems.

As winning a cricket game is not only based upon the skills of players but many other features influence the game results. keeping in view this reason our system will not only predict the game based on the skills of the players but will consider other features which influence match results like weather, venue and pitch type. All these different features will help us predict a game more accurately.

1.3 Aims and Objectives

The aim of our project is to construct a statistical approach to predict the outcome of a cricket Match before it's been played. Winning the cricket game depends on the strength of players as well as strategy making of the team. There are different systems which evaluate the strengths of players but the decision making procedures have yet to be explored.

The following are the aims of our project

- To identify different features that play vital role in predicting the match results.
- To predict a match before it's been played.
- To help team in strategy making from patterns found in previous match records.

1.4 The Cricket Game

Cricket is a game played between two teams which include eleven players each. It is based on a bat-and-ball. It is played in a circular ground which contains a 22-yard pitch in the center. Each team takes turn to bat and other fields till the decided overs. Over includes six balls which have to be bowled by the fielding team. The turn of team to bat is called an inning. The batsman tries to hit the ball away from the fielders to score while the fielders try to stop the ball or catch the ball while it's in air to take a wicket. There is a wicket behind the batsman which the bowler tries to hit. If bowler hits the wicket the batsman is out and next batsman will come in to bat. When all the batting side team is out or the decided overs finished, the teams exchange their roles. The one that was batting now bowls and the other that batted now fields. There is a target set for the team that is batting after the first team. The second team will try to reach the target within the given overs and wickets in hand.

The winner of the game is decided based on the runs scored or chased. If team playing inning two reaches the target within the given overs and wickets in hand, it wins else team playing inning one will win.

Cricket is one of the most popular games in the world including Asia countries like Pakistan, India, Sri Lanka,
Bangladesh and Afghanistan. There are total 12 international teams in cricket game which include Pakistan, India, Bangladesh, Australia, England, New Zeeland, Sri Lanka, Zimbabwe, Ireland, Afghanistan, West Indies and Netherland

In professional cricket the range of overs bowled in an inning ranges from 20 to overs per side to a test match which is played for five days with two innings per team. International Cricket Council defines laws for the game of cricket.

Cricket was introduced in England before 16th century. By the time it became the national sport of England. Due to the expansion of British Empire in the world, different nations started playing it. And in 19th century first International Match of Cricket was played.

There are there different formats in cricket namely One Day International (ODI), T20 and Test Match. In one day international each team has to bowl and bat for 50 overs. The team scoring maximum runs is considered as winner. In T20 internationals, each team plays 20 overs. Winning criteria is the leading scorer. In test marches, each team plays two innings and played in five consecutive days. The team scoring more is considered as winner.

The players are evaluated using statistical approaches finding their Run rates of bowling and batting and many other. These features determine the strength and skills of the players.

1.6 Motivation

Cricket is ranked as the second most revenue generating sports in the world. The international Cricket Council arranges different events like T20 world Cup, Champions Trophy and ICC world Cup. Among these ICC world cup is one of the largest sporting event in the world in terms of most watched or most revenue generating event.

Cricket has over 2-3 billion television viewers including Pakistan. Almost every individual is a fan of cricket and it’s one of the favorite games of the nation.

The system we are developing can benefit the Pakistan cricket board (PCB) in team’s strategy making and team selection according to their opponents. If the system predicts a win against a team, the players selected will play confidently in the ground and also help improve team strategies for captain. If it predicts a loose, the strategists, captain and team coach will make changes in team and decisions to make sure the teams victory.

CHAPTER 2

Literature Review

We Know Cricket is most popular game for entertainment specifically in Asian countries and some western countries. So there is lot of work have done on this game like movies, games, tool, sites for commentary. There are also some other prediction model which was introduced by some company and ICC (international cricket council). Basically prediction is more important in many fields like politics, games, and sports. The complete set of data sets for every game is on many sites or social media. For cricket there are many website where ball by ball data is available. Which we can use for analysis or prediction of this project. All cricket Data and records is accessible to the public through websites.

There are also some other prediction model which was introduced by some company and ICC (international cricket council).

1. The Duckworth–Lewis (D/L).
2. WASP (cricket calculation tool).

The Duckworth-Lewis (D/L)

The Duckworth–Lewis (D/L) method is a mathematical formulation and set to calculate the target score for the team batting second in a limited overs cricket match which are disturbed by weather or other circumstances. It is generally accepted to be the most accurate method of setting a target score.

WASP:

WASP is a model introduced by sky sports of New Zeeland in November 2012. It is a score predictor. WASP predict the outcome of two things:

I. It predicts the score for first inning based on past record of players and venue.

II. It predicts the outcome of match based on past record of team and venue.

Wasp model basically work on the basis of 2 main point. Wasp predict a person contribution in past played matches and total team performance on specifically venue. It use batting average and strike rate of team or batsman on this place or venue.

The prediction depends on factors like scoring on the day as per the match or pitch, climate and limit measure. For the Team1 batting to Firstly it gives the prediction of the last aggregate. For the team2 batting second it gives the probably of the chasing team2 winning.

CHAPTER 3

DATA COLLECTION AND IMPLEMENTATION

3.1 Data Collection:

Dataset was collected from cricsheet. Cricsheet is Retro sheet for Cricket. It is a website that provide ball-by-ball data for Men's and Women's Test Matches, One-day internationals, Twenty20 Internationals, some other international T20s, and all Indian Premier League seasons. We collected .yaml files for ODI matches and stored them in CSV files.

3.2 Data filtration: As it is mentioned above that cricsheet provides ball-by-ball data but we don't want to have ball-by-ball data for this we need to summarize this information. In order to summarize information we have used R Api called Yorker. The collected data is stored in CSV files.

Yorkr: R package is used to analyze performances of cricketers and different teams based on our dataset. Collected from cricksheet. Yorkr have ability to handle all type of matches' data from ODIs, IPL, and Tweenty-20s. It just require data to be in yaml format.

Using this R Api we make processing on yaml data to make database for matches. The database consists of following attributes that are important for system. Database would consist of following set of features data:

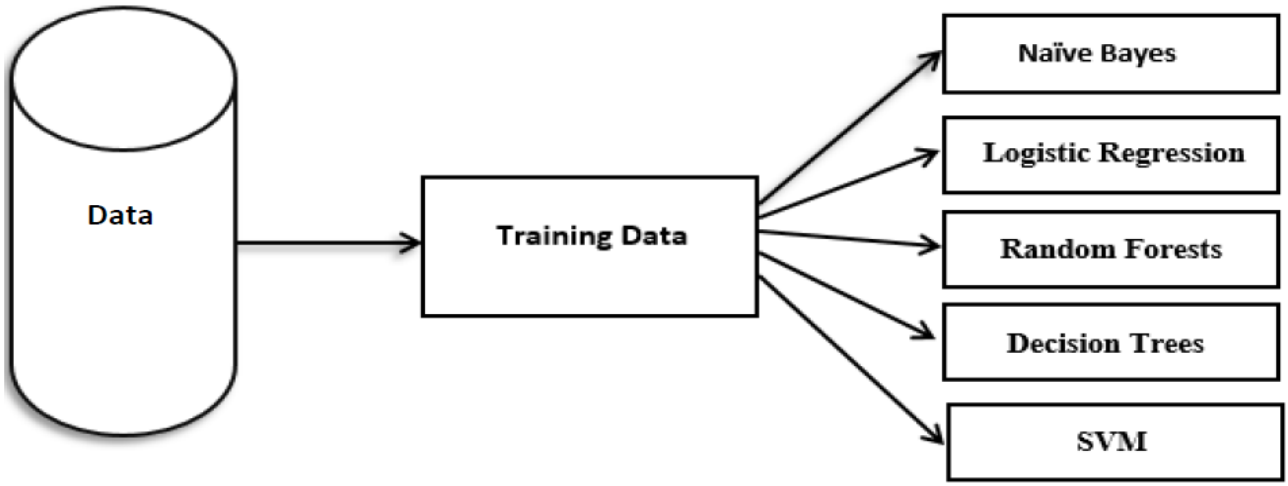
- Date
- City
- Team_1
- Team_2
- Toss Winner
- Toss Decision
- Result
- Winner
- Win-by-Runs
- Win-by-Wickets
- Player of the Match
- Venue

3.3 Implementation:

After storing the match records in CSV files, we implemented five different machine learning algorithms and selected the one which gave high accuracy. We used 70% of our data for training and 30% of the data for testing.

The list of implemented algorithms are as following.

- Naïve Bayes
- Logistic regression
- Random Forests
- Decision Trees
- SVM



3.4 Results:

The results we obtained after implanting the Algorithm are as following.

Algorithm	Accuracy
Naïve Bayes	47.581%
Logistic regression	33.871%
Random Forests	93.93 %
SVM	63.78%
Decision Tree Classifier	88.70%

We obtained good results from Random Forests so we selected it for prediction of matches.

3.5 Conclusion:

This analysis produced some of the energetic features that will rise to remarkable results. The previous work has targeted on perspective for predicating the match outcome on current statistics of match being played but this system uses historical data of teams to predict the result

REFERENCES:

[1] Khabir Uddin Mughal. Top 10 most popular sports in the world. 2015. URL http://sporteology.com/top-10popular-sports-world/8/

[2]. A. J. Lewis. Towards fairer measures of player performance in one-day cricket. URL http://www.jstor.org/stable/4102181

[3]. H. H. Lemmer. An analysis of players' performances in the first cricket Twenty20 World Cup series. URL http://academic.sun.ac.za/sajrsper/30.2/Lemmer%20665.pdf

[4]. P. E. Allsopp and S. R. Clarke. Rating teams and analyzing outcomes in one-day and test cricket. URL http://www.jstor.org/stable/3559882

[5]. D. Beaudoin. The best batsmen and bowlers in one-day cricket. PhD thesis, Simon Fraser University, 2003. URL https://www.stat.sfu.ca/content/dam/sfu/stat/alumnitheses/MiscellaniousTheses/Beaudoin.pdf

[6]Yorkr. URL https://gigadom.wordpress.com/2016/03/05/the-making-of-cricket-packag [5]. P. O. S. Vaz de Melo, V. A. F. Almeida, A. A. F. Loureiro, and C. Faloutsos. Forecasting in the NBA and other team sports: Network

[7] I. G. McHale and M. Asif. A modified Duckworth-Lewis method for adjusting targets in interrupted limited overs cricket. URL http://www.sciencedirect.com/science/article/pii/S0377221712007151

[8] Michael Bailey and Stephen R. Clarke. Predicting the match outcome in one day international cricket matches, while the game is in progress. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861745/

[9] Stylianos Kampakis and William Thomas. Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches URL http://arxiv.org/abs/1511.05837