

Bangladesh ODI Cricket Match Prediction

Abstract- The main goal of any game is to win. Today cricket is the most popular sport in the world. There are a lot of factors work behind the win or defeat in a match like home field advantage, venue, match condition, opponent team, toss, pitch condition, opening batsman, all-rounder, total spinner & fast bowler, team strength, captaincy & players etc. In the past many work & researches have been published about cricket prediction. But In this article a model is proposed for Bangladesh national team to predict win or loss in a match using data mining and machine learning approach & will be discuss about few others research paper. This model is just for one day international (ODI) cricket match. This model is applied using Linear Regression Classifier and Naïve Bays Classifier, Random forest etc.

Keywords: Cricket, Attributes for cricket winning, Data Mining, Linear Regression, Random forest, ODI Prediction, Logistic Regression .

Introduction

Recently cricket is the most popular game in Bangladesh. People of Bangladesh are much more interested in cricket. It has now merged with our culture. Not only in Bangladesh, cricket is now the most popular game in the whole world. Cricket is held in two teams, and each team has 11 players for 50 over's in ODI formats. Each player has different roles, such as batsman, bowler, All-rounder, wicket-keeper, captain etc. But everyone's role is very important to win a match. Cricket is divided into three format: ODI, Test and T20. Here we are mainly discuss about (One Day International) ODI cricket match. Now-a-days Bangladesh cricket has improved so much. Before any match or series, Bangladeshi people lot of concern about the national team of Bangladesh. Bangladesh played first ODI cricket in 31/03/1986 against Pakistan [1]. Bangladesh played 370 ODI matches till the last Cricket World Cup 19 [2].

Data

Variables like opponent team, venue, match condition, toss, bat-first, opening batsman, total all-rounder, fast-bowler, spinner, captaincy, Due-factor etc are important for prediction of winning percentage

Modeling

We considered all possible factors affecting the outcome at the start of a match. Team strength, ground effect and home field advantage were found to be important. For team strength, we have used win percentage differences and the ICC rating differences and find that the rating differences have better explanatory power. This may be because the ICC rating takes account of result (win, draw, loss), along with the win margin, wickets and opponent rating. Winning the toss was also considered in the model fitting but was found to be unimportant. The playing conditions vary from ground to ground and country to country. For example,

playing conditions for Bangladesh in England are quite different than in Bangladesh. Results of fitting the logistic regression model to 1169 match outcomes for various sets of predictors. Covariates here are win percentage difference, the ICC rating difference, home factor, ground effect. We can calculate the probability of win and loss given the position at the start of a match. This will help team captains and management to consider their batting and bowling strategy for the match. Matches with no results or tie results were deleted. Logistic regression was used to model the outcome of the cricket match.

Factors

Home-field Advantage:

Cricket is usually played at home and in other countries. This factor determines whether the match is on home ground or in another country. This factor is important because the home team game advantage has a lot of impact on the game. How many matches a particular team has won at home ground is considered by this factor. Throughout the year Bangladesh team plays a lot of series or matches at home or in other countries. So far, 169 matches have been played at home ground against different countries and Bangladesh team won only 72 matches. But a few years of data show that the Bangladesh team has made great progress in ODI format. (21-Nov-2014 to 14-Dec-2018) Bangladesh won 25 matches from 33 matches in home ground.

Specific venue:

Venue is very important and has a big impact on a match. Given below is a chart for Bangladesh team specific venue advantage:

Ground	City	Matches Played	Won by Home side	Won by Touring side
Sheikh Abu Naser Stadium	Khulna, Bangladesh	4	4(100.00%)	0(0.00%)
Bangabandhu National Stadium	Dhaka, Bangladesh	58	3(5.17%)	29(50.00%)
MA Aziz Stadium	Chittagong, Bangladesh	10	1(10.00%)	8(80.00%)
Shahid Chandu Stadium	Bogra, Bangladesh	5	4(80.00%)	1(20.00%)
Khan Shaheb Osman Ali Stadium	Fatullah, Bangladesh	10	3(30.00%)	4(40.00%)
Shere Bangla National Stadium	Mirpur, Bangladesh	108	44(40.74%)	46(42.59%)
Zahur Ahmed Chowdhury Stadium	Chittagong, Bangladesh	19	12(63.16%)	7(36.84%)

Here a few other optional sides are skipped, such as (neutral side, tied, no result etc.)

Day/Night Match:

This factor has a lot of influence when playing day or night match. In day/night format the team plays in day gets more advantage than the team plays in night because of dew factor. In a day/night game winning the toss is very important.

Algorithms

Linear Regression: Linear regression classifier is used for classifier of numeric data model and attributes . In statistics, linear regression is a foremost method. The idea is to get the reflex of the class in terms of linear abbreviation of the default weights and attributes which is been given by the equation (1).

$$X=w_0+w_1.a_1+w_2.a_2+w_3.a_3 \dots\dots\dots(1)$$

Where the class is X, the attribute values are (a₁,a₂,a₃..... a_v) and weights are (w₀,w₁,w₂,w₃ w_v) . From the training dataset weights are calculated. In this paper Linear regression classifier is used for training dataset , where the class attribute is 'X' .

Conclusion:

In this study, a comparative analysis of the predictions generated by 2 different supervised classification models was performed for the same input dataset. The proposed approaches are better than the statistical approach as unlike statistics which uses mathematical equations to formalize the relationships between variables, these approaches require no prior assumptions regarding the data variables and their underlying relationships. During training phase, data needs to be fed in and the algorithm after processing the data discovers patterns and finally makes predictions for freshly generating data.

References

[1]

https://en.wikipedia.org/wiki/List_of_Bangladesh_One_Day_International_cricket_records

[2]

http://www.howstat.com/cricket/Statistics/Matches/MatchListCountry_ODI.asp?A=BAN&B=XXX&C=XXX#odis