International Journal of Trend in Research and Development, Volume 5(1), ISSN: 2394-9333
www.ijtrd.com

Live Cricket Score and Winning Prediction

[1]Rameshwari Lokhande and [2]P.M.Chawan,
[1]Student, [2]Associate Professor,
[1,2]Computer Engineering Department, Veermata Jijabai Technological Institute, Matunga, Mumbai, India

Abstract: Winning is the aim of any game. Nowadays Cricket is one of the most watched game after Soccer. Winning of Cricket match depends on various factors like home advantage, past performances, experience in the match, performance at the specific venue, performance against the specific team and the current form of the team and the player. But in this paper prediction will be done while match is in progress i.e. live prediction .In this work number of wicket fallen, venue of match, ranking of team, pitch report, home team advantages those factors will be considered. In past few years lot of work and research papers have been published which measure only player performance and their winning predictions. This paper briefs about the elements that cricket amusement relies upon and examine around couple of other research papers that anticipated the cricket score and wining. Information Mining and Machine learning in Sports Analytics, is a fresh out of the box new research field in Computer Science with a ton of test.

Keywords: *Linear Regression; Naïve Bayes; Reinforcement algorithm: Data Mining; Projected Score; Winning Probability.*

I. INTRODUCTION

Cricket is one of the most admired sports in the world, second only to soccer. The game begins in England in the 16th century. Today, it is a most watched game in many Countries. Various natural factors affecting the game, enormous media coverage, and a huge betting market have given strong incentives to model the game from various perspectives. However, the there are some complex rules that governing the game, the players ability and their performances on a particular match day, and various natural parameters plays an important role in affecting the final result of a cricket match. This presents significant challenges in predicting the accurate results of a game.

The session of cricket is played in three configurations - Test Matches, ODIs and T20s. There are two groups played on the field. Rocking the bowling alley group has 11 players on the field. Batting group must have two players on the field. The fundamental point of Cricket is to score more 'Runs' than the other group. In the wake of hitting the knocked down some pins batsman running between the wickets at that point runs will be expanded, or any defender hitting the ball over the limit, or umpire rebuffed bowlers if bowler rocked the bowling alley unlawful bowl. The Captain of the Bowling group picks a man to bowl from his group; the other 10 players will do the handling. The bowler's primary objective is to ball at a wicket, which is comprised of three stumps stuck into the earth, with two little safeguards adjusted on them. One of the defenders remaining behind the wicket is known as wicket attendant, he will get the ball if the bowler neglects to hit the wicket. Alternate defenders endeavor to take care of business the ball after the batsman has crushed it. The bowler keeps running toward his wicket, and dishes towards the batsman at the other wicket. Bowler does not toss the ball. Bowler bowls the ball overarm with a straight arm. An 'over' comprises of six balls meaning bowler bowls six times. At that point chief will offer knocking down some pins to another player for the following over, and bowls from the other way of past bowler, et cetera. Two back to back finisher's can't bowl by same bowler. The batsman is endeavoring to shield the wicket from taking care of business hit with the ball. Batsman does this with a bat. At the point when Batsman crushes the ball with his bat, he may keep running toward other wicket. Two batsmen must be keep running from their wicket to the next wicket for scoring runs, the greatest number of times as they can, however once in a while they can be run out if any defender hit the wicket before any batsman not coming to flip side. At the point when a batsman is getting out then another player goes onto the field to have his spot. Ten wickets are taken by the playing group then inning will get over.

After this, handling group will desire the batting and the other way around. The group batted in second inning needs to score a larger number of keeps running than inverse group. On the off chance that they do before ten wickets are fallen, they win generally other group wins. In a one-day diversion arrange, each group has opportunity to play fifty over's. In test cricket arranges each group has two innings, and there is no specific breaking point to the quantity of over's in an innings. This exploration paper concentrates on constrained over's arrangement i.e. ODI and T-20 organize. Presently a day the most famous arrangement of the cricket is T-20 design. The principle objective of this paper is to influence a model which to will anticipate the score and aftereffect of a cricket amusement while the match is in live broadcast. Fundamental objective of this paper is to consolidate past matches information and in-diversion information keeping in mind the end goal to plan a best prescient model. For outlining great prescient model: number of wicket fallen, setting of the match, positioning of the batting group, hurl and home group advantage these parameters will be considered. As no such research has improved the situation this organization of cricket, so this will be chosen to take the test as cricket matches are especially prominent now days.

II. PROPOSED SYSTEM

This paper will execute live cricket score and winning forecast. In this model there are two sections. In initial segment I will anticipate add up to score of group which is bat first. To anticipating the aggregate score I will consider different variables; hurl, number of wicket fallen, setting, home diversion advantage, day/night impact and positioning of group. In this framework first it will watch that who will win the hurl and which group bat first. At that point it will be dissect that what will be add up to score of batting initially group from past record. In second part model will indicates who will win the match by utilizing previously mentioned factors. The accompanying variables will be considered for expectation procedure incorporates:

Home Game Advantage:

This refers to whether the game is played on home grounds or in a different country. This factor is very important because home team game advantage have big impact on game.

In this factor work is going to be considered how many matches were won in home ground by a particular team.

Day / Night Effect*:*

This factor considers the effect of whether the match is played during the daytime or at night. If match is played in day/night format, team playing in day has advantage over team playing in night because of dew factor. Winning the toss is important in a day-night game.

Winning the Toss:

A coin is tossed before match is started and then captain will decide which team should bat first. Nobody can disagree with the importance of toss in a cricket match. Winning the toss is very important because some pitches are very good for batting and some pitches are bowling favorable. So for choosing batting or bowling toss is crucial factor.

Batting First:

This factor determines whether the concerned team bats first or bowls first in the given match. Some pitches are batting pitch so batting first is advantageous.

Venue:

The cricket ground where the match takes place. Home advantage may be replaced by use of a neutral venue. Alternatively neutral venue may be home stadium of another different stadium.

Using these factors, I will first apply different classification technique, to predict whether a given team will be winning the match or not. Based on these results, I will develop a model that can be used to predict the probability of victory and total score of any team which is played in first inning in a limited over's cricket match. These works will be developing for a particular sport with different algorithms and techniques describe as below.

Linear regression

If the data of a class and attributes are numeric then linear regression classifier is used for classification. This is a staple method in statistics. The concept is to get the expression of the class in terms of linear combination of the predetermined weights and attributes which is been given by the equation (i).

X=w0+w1.a1+w2.a2……..(i)

Where x is the class; a1, a2, ...av are the values of the attributes, and w0, w1, ... wv are weights. The weights can be computed from the training data. Here, the notation gets a little heavy, because a way of expressing the attribute values is needed for each training instance. In this paper, the Linear Regression Classifier have been implemented for the first innings datasets where the class attribute 'x' is the 'Score' and the input attribute 'ai'(i = 1,2,3…) are the current score and wickets fallen at 5-over period starting from 0-5 till 40-45 over's.

Naïve Bayes

Naive Bayes classifier is the based on the Bayes theorem . It is probabilistic classifier which depends upon the strong (naive) independence assumptions. This classifier assumes that presence or absence of a particular feature of a class is independent to the presence or absence of any other feature. It considers all of these features to independently contribute to find the probability even if the attributes depend on each other. Depending on the accuracy of the probability, Naïve Bayes classifier is used for efficiently training a supervised learning setting [14].It is based on the model of conditional probability in which a given instance has be classified is given by a vector x = (x1,…….,xn) which represents some n features (dependent variables) of the n attributes, A1, A2, . . . ,An respectively, it assigns the probabilities to each of k possible classes or outcomes. Using the Bayes' theorem, the conditional probability is given by the equation (ii).

$$P(Ci|X)= \frac{P(X|Ci)\ P(Ci)}{P(X)} \qquad (ii)$$

Where,P(Ci|X) is the posterior probability of the class, given predictor (attribute),P(Ci) is the prior probability of class, P(X|Ci) is the likelihood which is the probability of predictor given class and P(X) is the prior probability of predictor. The Equation (ii) can also be written of the form as depicted in equation (iii). The evidence, also termed as normalizing constant is equal to the sum of the posteriors.

$$\text{Posterior}=\frac{\text{prior}*\text{likelihood}}{\text{evidence}} \qquad (iii)$$

The evidence can be ignored as it is a positive constant. (Normal distributions are always positive)[15]. So, only calculate the numerator of the equation (ii) which is P(X|Ci)P(Ci).

Reinforced Learning Algorithm

- Reinforcement learning emphasizes learning feedback that evaluates the learner's performance without providing standards of correctness in the form of behavioral targets.
- Evaluative feedback:
  - tells the learner whether or not, and possibly by how much, its behavior has improved; or
  - provides a measure of the 'goodness' of the behavior; or
  - Just provides an indication of success or failure.

Temporal Difference Learning

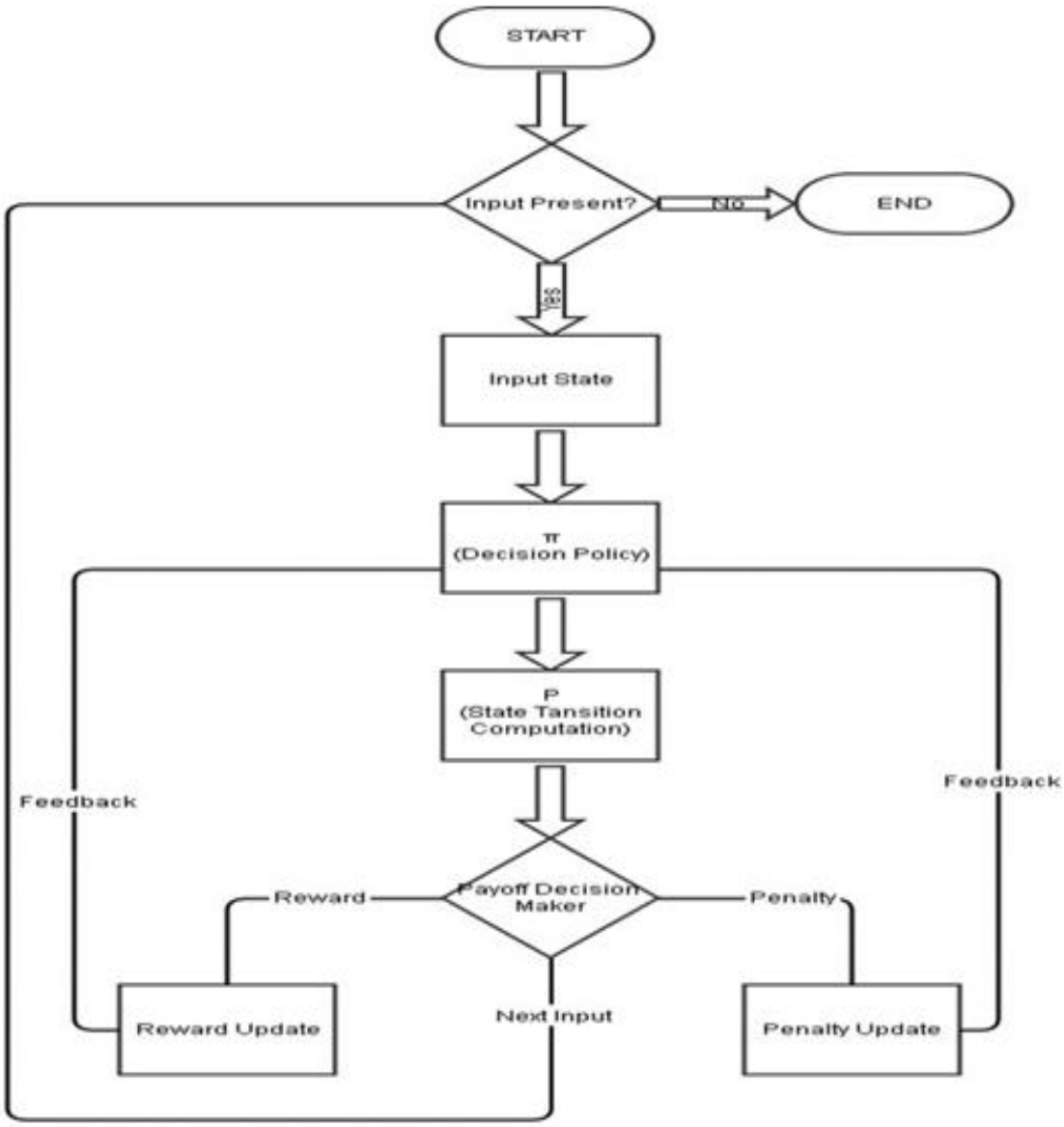
- TD algorithms are often used in reinforcement learning to predict a measure of the total amount of reward expected over the future, but they can be used to predict other quantities as well.
- $R_t = r1 + Yr_2 + Y^2r_3+….+ Y^nr_{n+1}+….$
- R_t is the reward at t time step.

The aim of these policies is to balance the trade-off between exploitation and exploration, by not always exploiting what has been learnt so far.

Greedy - most of the time the action with the highest estimated reward is chosen, called the greediest action. Every once in a while, say with a small probability, an action is selected at random. The action is selected uniformly, independent of the action-value estimates. This method ensures that if enough trials are done, each action will be tried an infinite number of times, thus ensuring optimal actions are discovered.

Soft - very similar to greedy. The best action is selected with probability 1 and the rest of the time a random action is chosen uniformly.

Flow Diagram



Q Learning Algorithm

- We use this algorithm in the result and evaluation phase.
- At each step s, choose the action a which maximizes the function Q(s,a)

Q is the estimated utility function**, its tells us how good an action is given a certain state**.

- Q(s,a) = immediate reward for making an action +best utility(Q)
- $Q(s\ ,a)=r\ (s\ ,a)+\gamma max_{a'}(Q(s\ '\ ,a'\ ))$
- $r\ (s\ ,a)$=Immediate reward
- γ =relative value of delayed vs. immediate rewards (0 to 1)
- $s\ '$=the new state after action a
- $a\ ,a'\ :$ actions in states s and $s\ '$, respectively
- Selected action:
  $\pi(s)=argmax_a Q(s\ ,a)$

CONCLUSION

The main purpose of this paper is to make a model for predicting the final score of the first innings and estimating the outcome of the match in the second innings for the limited overs cricket match. Factors like the toss, the ODI ranking of the teams and the home team advantage will be considered in the predictions. Two separate models, one for the first innings and other for the second innings using the Linear Regression classifier and Naive Bayes classifier respectively on the past matches have been proposed. Reinforcement algorithm will be used instead of linear regression.

References

[1] Laws of cricket. http://www.lords.org/mcc/laws-ofcricket/Accessed 2 January 2015. Machine
[2] 2. Learning Group at the University of Waikato. Weka 3:Data Mining Software in Java.
[3] Naïve Bayes Classifier. http://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906 /aulas/naive-bayes-classifier.pdf Accessed12 March 2015
[4] Tejinder Singh, Vishal Singla, Parteek Bhatia, "Score and Winning Prediction in Cricket through Data Mining", International Conference on Soft Computing Techniques and Implementations, 2015.