

Match Prediction And 1st Inning Score Prediction

Syeda Roshni Ahmed¹, Sanjay Kumar Mahto², Ritish³, Shivaraj A Nandyal⁴

Assistant Professor, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India¹

Student, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India^{2, 3, 4}

Abstract - Applying machine learning for analysing cricket sports by considering historical game information, players performance, natural parameters, pre-game conditions. In an intensive cricket format like T20, where the situation in a game changes on every ball, it becomes challenging to predict the match outcome. For predicting the ultimate outcome of a T20 match, we've investigated machine learning technology for the chance of rising the prediction rate of the results of matches. We have developed the matter in 2 situations, named for the foremost potent options, first of all the house Team options set and second Toss Winner call options set.

Machine learning could slightly improve predicting the results supported pre-game conditions however at this stage it can not be an appropriate answer thanks to lack of variables within the dataset, which may be thought of jointly of this research's limitations. In order for machine learning techniques to be productive, additional knowledge as well as live knowledge streaming and statistics of players are required. Considering the dynamics of the tournament, team players' information and statistics square measure needed. It would be advantageous to predict the final score of the innings by analysing the run rate per over and also checking the likelihood of winning for every team counting on the particular run rate and therefore the needed run rate within the second innings. Similar models can be built for other cricket formats, i.e. test cricket and one day international (ODI) series.

1. INTRODUCTION

Cricket is the second most popular sport in the world with billions of fans across India, UK, Pakistan, Africa, Australia, etc. It is an outside game on a cricket field at 22-yard rectangular long pitch, between 2 groups consisting of eleven players. It is played in three formats namely Test, One Day International (ODI) and Twenty Over International (T20). In ODI each team takes its chance to bat, trying to score as many runs which can be scored in 50 overs while the other team fields for that much amount of overs. Each chance is termed as an innings. The batsman looks to make runs by hitting the ball being bowled to him. The bowler on the opposite hand tries to urge the batter out. There are unit sure rules outlined to urge the batter out by the

bowlers or the fielders. every batsmen keeps on batting till he gets out. So, the innings of the batting team is over when either the 10 batsmen out or the 50 overs have been bowled by the fielding team; in either of the situations the batting team now gets the chance of bowling and vice-versa. The team which scores more runs will win the match.

Unlike different sports, cricket stadium's size and form isn't mounted except the scale of the pitch and coterie that area unit twenty two yards and thirty yards severally. The cricket rules don't mention the scale and also the form of the sphere of the construction. Pitch and outfield variations can have a substantial effect on batting and bowling. The bounce, seam movement and spin of the ball depends on the character of the pitch. The sport is additionally full of atmospheric conditions like altitude and weather. a singular set of taking part in conditions area unit created because of these physical variations at every venue. Depending on these set of variations a particular venue may be batsman friendly or a bowler friendly. Currently, in an ODI match the projected scores can be seen displayed at the score card during the first innings, which is basically the final score of the batting team at the end of that innings if it scores according to the current run rate or a particular rate. Run rate is defined as the amount of runs scored per the number of overs bowled. However, run rate is taken into account because the sole criteria for calculating the ultimate score. But there are other factors too which may affect the final score like number of wickets fallen, the venue and the batting team itself.

2. BRIEF DESCRIPTION

In this project we develop a model in order to predict outcomes of the Indian Premier League and 1st inning score over the years 2008-2019. We used a multi-step approach to research the info that created over five hundred records. There are different attributes used in the project which are id number, Season in which the match had been played which ranges from 2008 to 2019, city where the match had been played, the date on which the match was played, names of two teams participating in the match, the toss winner, toss decision, result of match which can be normal, tie or no result. No result is found thanks to some interruption within the game, the main reason for these may well be thanks to rain. The other attributes are whether D/L method is applied or not

which stands for Duckworth–Lewis which comes into play when rain has occurred. The other attributes are the Winner of the game, win by how much runs if the winning batted 1st and win by how much wickets if the winning team batted second. Several different attributes are player of the match or man of the match, venue or the name of stadium where the match was played and name of the two standing umpires present on the field.

The prediction of the match is made by eliminating some of the features which is Data Cleaning method. For the prediction of the match we have used some algorithms which are LASSO Regression and Logistic Regression. The prediction of the match is made mainly on teams participating, toss winner, toss.

3. METHODOLOGY

In this section, the methods related to the work are presented. The diagram below shows the workflow of the model considering the following aspects: credit datasets, data cleaning and feature selection, machine learning algorithms and evaluation measures of model performance.

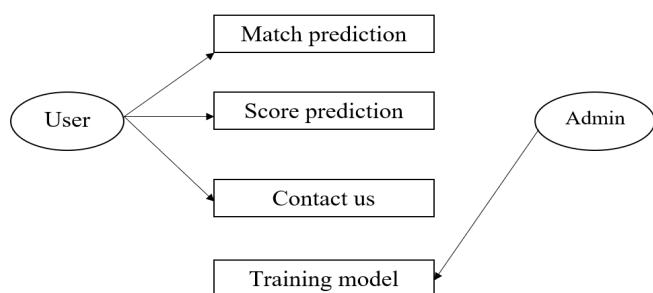


Figure 01. Overview of proposed model

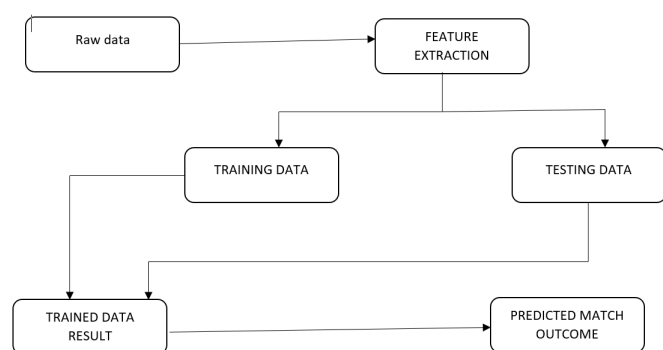


Figure 02. Flow diagram for proposed system

3.1 Cricket dataset:

We took our dataset from the kaggle website which

contains attributes like team1, team2, toss decision, toss winner, city, venue, season, win by runs, win by wickets, umpire1, winner which data from year 2008 - 2019.

3.2 Data cleaning:

It is the process of making data fit for use, prediction and removing the unwanted data from the dataset.

Data cleaning can be done using

- Preprocessing of data
- Feature of Engineering
- Feature Selection

Preprocessing of data:

This includes handling the missing data. This dataset has lots of missing values. To handle this we had given the average of that column.

Removing the duplicate data while going through our dataset I found Team Mumbai Indians were duplicated in columns team_1, team_2, winner, and toss_winner.

Feature of engineering:

In the cricket dataset we have to deal with a lot of categorical values. And that is not understandable by computers, so we need to convert all categorical data into numeric data. Columns like toss_winner, toss_decision, and winner are understandable to us, but to do this in better form we can add the three columns which contain all the information above and delete columns like toss_winner, toss_decision, and winner.

We can add toss_winner to team1_toss_win: if team1 won the toss then we replace it with 1 else 0. Where 0 means team 2 won the toss.

Toss_decision to team1_bat_decision: if team1 won the toss and made a decision to bat then we replace with 1 else 0.

Where 0 means team1's own toss and took fielding or team2 won the toss.

Winner to team1_win: if team1 won the match then we replace it with 1 else 0. Where 0 means team 2 won the toss.

Feature selection:

It is a process of making data to a certain range so that our machine can predict with high accuracy.

3.3 Algorithm

We have used algorithms like Lasso regression, Logistic Regression.

LASSO Regression:

The LASSO algorithm stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a L1 regularization technique. It is used over regression

methods for a more correct prediction. It is efficiently used when we have more features. It has the property of Shrinkage where data values are shrunk towards a central point as the mean.

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Logistic Regression:

A provision regression model predicts a dependent information variable by analyzing the connection between one or a lot of existing freelance variables.

$$g(E(y)) = \alpha + \beta \times 1 + \gamma \times 2$$

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The algorithm has been tested with the Ipl matches dataset taken from the kaggle website. The performance metrics used to appraise this model are Accuracy, Recall, Precision, F1 Score and Confusion matrix.

	Precision	Recall	F1-Score	Support
Team1_won	0.68	0.75	0.71	51
Team2_won	0.59	0.51	0.55	37
Accuracy			0.65	88
Macro avg	0.64	0.64	0.64	88
Weighted avg	0.64	0.65	0.64	88

	Precision	Recall	F1-Score	Support
Team1_score	0.77	0.86	0.81	51
Team1_score	0.84	0.75	0.79	37
Accuracy			0.80	88
Macro avg	0.81	0.80	0.80	88
Weighted avg	0.81	0.80	0.80	88

Using lasso regression we try to predict the first inning score, as shown in the confusion matrix. Where we have accuracy upto 80. which is better than currently used models. The accuracy of this can be incremented further by doing more research.

5. CONCLUSIONS AND FUTURE WORK

Our Cricket Match Winning Prediction Project will be very useful in the coming time. The Prediction will be made on statistical records and using some proposed model specifically LASSO regression and Logistic regression so it will help the people to have a look over it. The results also show that our prediction will be almost accurate.

The main focus will be on increasing the accuracy of the model. We also to consider some of the main factors which we have not considered yet in this project. Like in this project we have not considered the role umpires play in the match but in reality their role is very crucial. Also we can include the role played by Third Umpire.

REFERENCES

- Bandulasiri, A. (2008). Predicting the team winner in one day international (ODI) cricket. Journal of Mathematical Sciences Mathematics Education , 3 (1), 6-17.
- Brooks, R. D., Fa, R. W., Sokulsky, D. (2002). An ordered responses model of test cricket performance. Applied Economics, 34 (18), 2353-2365. ICC. (2012-2013). ICC Annual Report.
- Aminul Islam Anik, Sakif Yeaser, A.G.M Imam Hossian and Amitabha Chakrabarty (2018). A Players performance prediction in ODI Cricket.
- Harshit Barot ,Arya Kothari ,Pramod Bide, Bhavya Ahir, Romit Kankaria (2020). an analysis and prediction for IPL (Indian Premier League).
- Hatharasinghe, M.M., and Poravi, G. (2019). Cricket Match outcome Prediction.
- Wickramasinghe, A.N., and Yapa, R.D. (2018). Cricket Match Outcome Prediction Using Tweets and Prediction of Man of the Match using Social Networks Analysis.
- Abdul Basit, Muhammad Bux Alvi, Fawad Hassan Jaskani (2020). ICC World cup 2020 winner prediction using Machine Learning Techniques.
- Amal Kaluarachchi, Aparna S. Varde (2010). A classification based tool to predict the outcome of ODI cricket.
- Rameshwari Lokhande, Parmila M. Chawan (2018). Prediction of Live score and winning.