

Cricket Prediction using Machine Learning Algorithms

Sudhanshu Akarshe¹, Rohit Khade¹, Nikhil Bankar¹, Prashant Khedkar¹, Prof. Prashant Ahire²

Student, Department of Computer, D Y Patil Institute of Technology, Pimpri, Savitribai Phule Pune University, Pune, Maharashtra, India¹

Professor, Department of Computer, D Y Patil Institute of Technology, Pimpri, Savitribai Phule Pune University, Pune, Maharashtra, India²

ABSTRACT

Cricket is most popular sport played in India. It has huge spectator support and the masses show great interest in predicting the outcome of games in their Test, One-day international as well as in T-20 matches. The game is having number of rules and scoring system. Numerous parameters are present such as, cricketing skills and performances, match venues which has significant effect on the outcome of a game. Such parameters, along with their interdependence create a challenge to create an accurate prediction of a game. In this project, we are going to build a rigid prediction system that takes in historical match data, player performance and predicts future match events such as final results in a victory or loss. Our system will perform this prediction using various machine learning algorithms. We describe our system and algorithms and finally present quantitative results displayed by best suited algorithm having highest accuracy. Also, representing the winning team even before the match starts and provide best suited squad of both teams.

Keywords : Prediction System, Historical Match Data.

I. INTRODUCTION

Cricket first started in the 16th century in England. Cricket is played in multiple formats, different playing standards and varying duration. There are three current forms of Cricket which is recognized by the International Cricket Council (ICC) which are Twenty 20, One Day International and Test match.

Indies was first country to hold cricket world cup. Further, in 1987 overs per innings were reduced to 50 overs each side. Australia is the first ever country to hold three consecutive world cups. Pervious world cup held in England and Wales in 2019 was won by England.

II. METHODS AND MATERIAL

CRICKET WORLD CUP

First ever international cricket world cup was hosted by England in 1975 as one-day match where each team had 60 overs to bowl out opponent team. West

CRICKET PREDICTION SYSTEM

Yearly, various tournaments conducted at both national and international level. Prediction of player performance has huge demand. With such demand, it has motivated many people for analyzing team performance and even prediction of future games, across all formats of the game. Currently, many strategists are present where combination of player

experience and team building for making instantaneous strategic decisions.

III. PROPOSED SYSTEM

In this system, primary aim is to predict the match outcome and to provide a best team against opposite team. In order to achieve a reliable accuracy, we need to analyze a large amount of data. Therefore, the first step of the system's implementation was to collect data for all possible matches. Dataset is collected from various websites such as ESPN website. Dataset for team winning prediction consist of all ODI matches played since 2000. Dataset used for squad selection consist of all international players which are not retired till date.

Cleaning of data is done in order to obtain meaningful data. This data is used to train machine learning models.

Analytical rules are applied to filter collected data with respect to the selected feature. Features are match venue, playing 11, weather condition, performance of individual player.

Further, cleaned data is split into training and testing data. Training data is fed to all machine learning models and accuracy of each model is noted. Algorithms having highest accuracy is selected in the model. Three algorithms were used which are as follows-

DECISION TREE:

A decision tree uses tree-like chart to classify data based on all features to include all possible outcomes. Max_depth in function is selected based on number of features present in dataset.

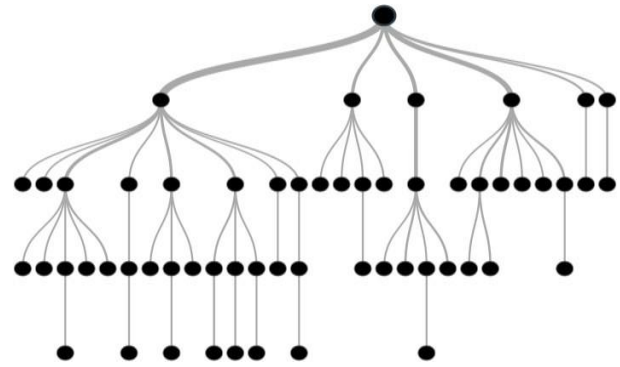


Fig 1. Decision Tree

GRADIENT BOOSTING

Gradient boosting is greedy algorithm and can overfit a training dataset quickly. N-estimators and random_state differs as per features selected in dataset.

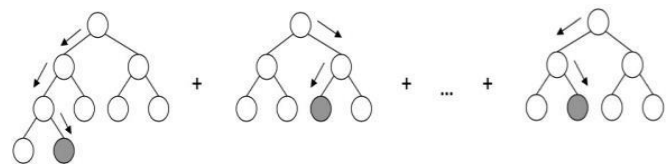


Fig 2. Gradient boosting.

RANDOM FOREST

Random forest is a classification algorithm consisting of multiple decision trees. Each tree is built using bagging and feature randomness.

Entropy is used as criterion for quality of split as it focuses on information gain. Also, number of trees in forest differ with respect to data.

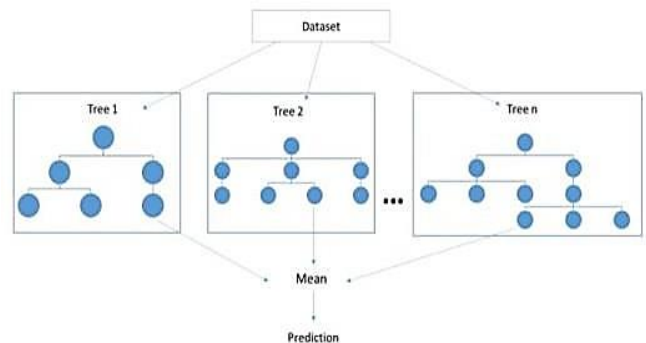


Fig 3. Random Forest

IV. IMPLEMENTATION

I. TEAM WINNING PREDICTION A] Dataset fetching:

We are using BeautifulSoup for parsing the webpage and searching for specific elements from that webpage. The data is scraped from ESPN website and stored in CSV formats for further processing.

B] Preprocessing:

The dataset is preprocessed to eliminate the missing data values or missing information which is of no use for prediction. In data preprocessing, the steps like data cleaning, feature reduction, selection are done which are independent methods. Elimination of redundant means columns which containing multiple values because it will hamper the prediction. The dataset is then split into training dataset and testing dataset.

C] Model selection:

Three different model are used having unique features which are as follows:

- Ground based prediction: This model is trained over features such as venue location, host country, type of pitch.
- Weather based prediction: This model is mainly trained over the weather at that venue where match was held.
- Toss based prediction: This model is trained over features such as toss winner, toss choice, winner based on innings played.

D] Model Training:

All three models are trained using following algorithms

- Random Forest.
- Gradient Boosting Algorithm.
- Decision Tree.

All three algorithms are uniquely featured with respect to each model to obtain higher accuracy and models are trained.

E] Results:

Initial average highest accuracy of all three models is about 70.02%. When predicted value of all models is boiled down to single prediction the accuracy obtained is 79.20%.

V. SQUAD SELECTION

A] Dataset Fetching and Preprocessing:

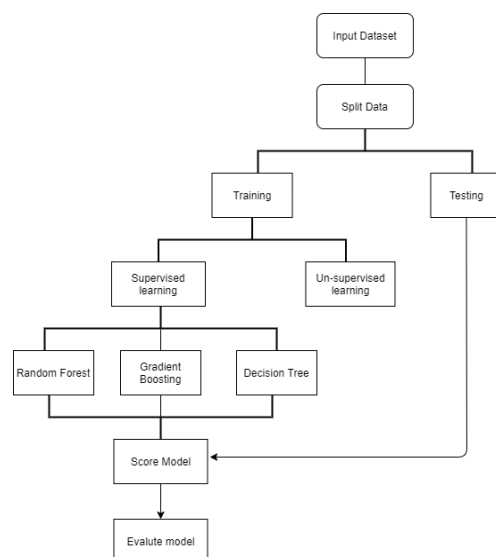
Data fetched and preprocessed in Team Winning Prediction is used.

B] Player Selection:

We are using the data of grounds from scraped data. This data is in CSV format. There are files of each ground which consists of top performed players on that ground. This data is used for selecting the best players on a particular ground based on their performance.

We are also selecting players based on their average score, efficiency, strike rate, etc.

VI. SYSTEM ARCHITECTURE



WORK FLOW DIAGRAM

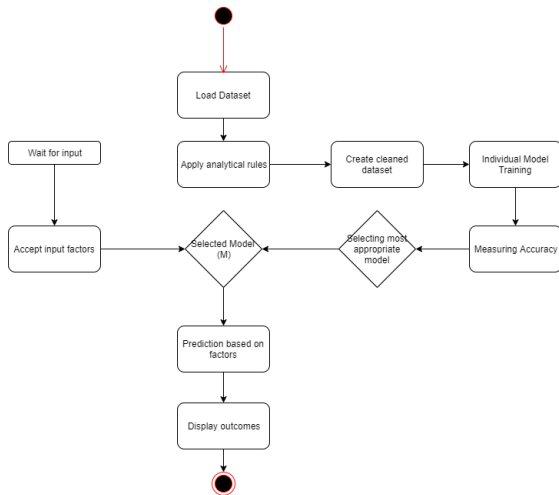


Fig 2. Work flow diagram

TOOLS

Web Scrapping –

Web scraping is used for the extraction of data from a website. Data on the website is in an unstructured format. Scraping helps to collect this data and convert/store it into a structured format. Data is scraped from <https://www.espncricinfo.com/>.

Python -

Python is interpreted, high level and object-oriented scripting language. Python is designed to be highly readable. It is widely used in data science domain.

Android -

Java language is used for developing Android applications also using Android Software Development Kit. User application is built in android.

VII.FUTURE SCOPE

Previously, a single algorithm is used in these prediction systems to measure performance. Instead, we have used multiple algorithms and merged their performance to improve overall accuracy.

VIII. ACKNOWLEDGEMENT

In this way, with the help of prediction system we can predict the outcome of matches on prior basis and provide proper squad selection

IX. REFERENCES

- [1]. 'Cricket Score Prediction system using clustering algorithm', A. Tripathi, J. Vanker, B. Vaje, V. Varekar.
- [2]. 'Identifying optimal set of attributes that impose high impact on end results of cricket match using machine learning', P. Somaskandhan, G. Wijesinghe, L. Bashitha.
- [3]. 'An analysis of Bangladesh One day international cricket: Machine learning approach', M. Rehman, O. Shamim, S. Ismail.
- [4]. 'Predicting the Match Outcome in One Day International Cricket Matches, while the Game is in Progress', M. Bailey and S. Clarke.
- [5]. 'Applications of Modern Classification Techniques to Predict the Outcome of ODI Cricket', N. Pathak and H. Wadhwa.
- [6]. 'Prediction of rising stars in cricket', M. Ahmad, A. Doud, L. Wang, H. Hong.
- [7]. ESPN Cricinfo, <http://www.espncricinfo.com>

Cite this article as :

Sudhanshu Akarshe, Rohit Khade, Nikhil Bankar, Prashant Khedkar, Prof. Prashant Ahire, "Cricket Prediction using Machine Learning Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 3, pp.1128-1131, May-June-2020. Available at doi : <https://doi.org/10.32628/CSEIT2063195>
 Journal URL : <http://ijsrcseit.com/CSEIT2063195>