

IPL Player's Performance Prediction

Nihal Patel^{1*}, Mrudang Pandya²

^{1,2} Dept. of Information Technology, CSPIT, Charotar University of Science and Technology, Changa, Gujarat, India.

Corresponding Author: nihalhari86@gmail.com, Tel.: +91 8469992314

DOI: <https://doi.org/10.26438/ijcse/v7i5.478481> | Available online at: www.ijcseonline.org

Accepted: 21/May/2019, Published: 31/May/2019

Abstract— Fantasy cricket league is a rapidly growing industry in India. It has around 70 million users. Lots of people are really making money from it. Player selection is the most important task in the Fantasy League. The performance of a player depends on various factors such as opposition team, venue, his current form and many more. Fantasy league user has to make own team of 11 players from both the team players. In this paper, we are going to predict the performance of a player in IPL matches by analysing previous year's ball by ball data (2008-2018) using supervised machine learning techniques. Here we classified the batsman's runs and bowler's wickets in a different range to pick or not to pick. We used Decision Tree, Random Forest, Xgboost, Stacking for prediction of the players[6]. Stacking technique found the most accurate classifier for the problem.

Keywords— Fantasy League, Machine Learning, Decision Tree, Random Forest, Xgboost, Stacking.

I. INTRODUCTION

Cricket is one of the most popular sport around the world. But in India, it is the most played sport. It has three formats test, one day and t20[7]. Here we are dealing with the t20 format. I 2003, T20 Cricket was first introduced in England. Due to its high voltage action in every over and on every ball T20 cricket came to India also. Indian premium language(IPL) was founded by the Board of Control for Cricket in India (BCCI) in 2008. IPL completed 11 successful seasons.

Fantasy sports is a \$7.22 billion industry. Nowadays, earning money from Fantasy Cricket is a major topic to discuss with friends.

Fantasy cricket is classified as a "game of skill", similar to fantasy sports in the United States. In this game, the user registers their teams and then compete with other users to see who gets the maximum points. In IPL, these leagues are very popular.

So, how to make the best team for Fantasy League? Is this player going to perform well in next inning? How many runs will VIRAT KOHLI make in next match? This all are the primary questions for those people who want to make money from Fantasy League.

"You can make lots of money in IPL match with enough preparation"

In this paper, we predict the players' performance in One Day International (ODI) matches by

analysing their characteristics and stats using supervised machine learning techniques. For this,

we predict batsmen's and bowlers' performance separately as how many runs will a batsman

score and how many wickets will a bowler take in a particular match.

In this paper, we are going to predict the performance of a player in IPL matches by analyzing previous year's ball by ball data (2008-2018) using supervised machine learning techniques. In section III we mention the point system of fantasy cricket league and construct the dataset according to that in section IV. After that we apply machine learning models on dataset in section V.

II. RELATED WORK

A very few articles related to players' performance prediction in the game of cricket out there. Kalpdram Passi and Niravkumar Pandey[3] 'predicting player's performance in one-day international cricket match using machine learning'. C. Deep Prakash, C. Patvardhan, C. Vasantha Lak-

shmi[8] 'Data Analytics based Deep Mayo Predictor for IPL-9'. Debarghya Das[9] 'An Integer Optimization Framework for Fantasy Cricket League Selection and Substitution'. Tim B. Swartz[10] 'Research Directions in Cricket'.

Our work is probably the first generalized approach to predict how many runs will a batsman score and how many wickets will a player take on a particular match day and also to predict the performance of player for fantasy league cricket. We used some supervised machine learning algorithms to build prediction models that can be used to predict the performance of any player in a IPL match.

III. FANTASY POINT SYSTEM

In the fantasy league, we have to make a strong team of 11 players which is going to perform well in the match. Each format has its own point system here we are talking about t20 game.so according to the dream11[4] point system

Batting: run= +0.5

Boundary bonus = +0.5

Six bonus= +1

Half-century bonus= +4

Century bonus= +8

Dismissal for duck= -2

Bowling: wicket= +10

4 wicket haul bonus= +4

8 wicket haul bonus= +8

Maiden over= +4

The Captain you choose in your team will get 2 times the points and for Vice-Captain it will be 1.5 times.

So according to the above point system, we understood that we have to choose that batsmen where they are going to make the highest run and pick those bowlers where they are going to take highest wickets to make our team point total highest.

So, now the question is how are we going to pick 11 players from two teams? here we make a range for batsman and bowler to pick in our 11.

IV. DATA PRE-PROCESSING

Batsman: If the batsman make

<40 runs = not pick

40<runs<80 = pick

>80 = candidate for captain or vice captain

Bowler: If the bowler take

<2 wickets = not pick

2,3 wickets = pick

>3 wickets = candidate for captain or vice captain.

Here we have the ball by ball dataset available of IPL matches from 2008 to 2018 and we make a range for batsman and bowler to pick or not so we convert this problem into a classification problem.

With this, we get the data of batsman and bowler how they performed in each inning vs each team and how they performed at a particular venue.

So, on this data, we apply supervised learning technique like random forest, decision tree, logistic regression, naïve base classification and get different- different accuracy.

In this paper, we choose 'VIRAT KOHLI' as our batsman and 'LASITH MALINGA' as our bowler and analysing their performance and do prediction on that.

V. MACHINE LEARNING MODELS

5.1 DECISION TREE

Decision tree induction is the process of creating decision trees for class label training tuples. A decision tree is basically a tree structure like a flowchart. The first node at the top of the tree is the root node. To classify a given tuple X, the attributes of the tuple are tested against the decision tree starting from the root node to the leaf node which holds the class prediction of the tuple[6].

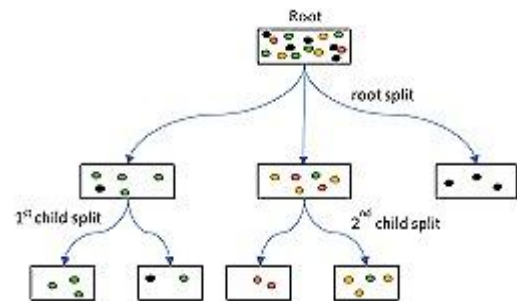


Figure 1: Decision Tree

For batsman Decision tree gives us 59.12% cross validation score and 4.43% standard deviation while for bowler 59.70% cross validation score and 7.93% standard deviation.

5.2 RANDOM FOREST

Random Forests is an ensemble method for classification and regression. Random forest is a set of decision trees. The algorithm generates a number of decision trees to create a forest. Where each tree is dependent on a random vector sampled independently and with the same distribution of all the trees in the forest.

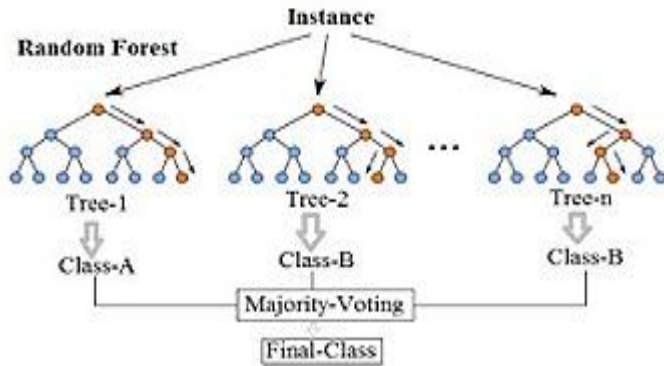


Figure 2: Random Forest

For batsman Random Forest gives us 72.14% cross validation score and 2.95% standard deviation while for bowler 62.05% cross validation score and 4.49% standard deviation.

5.3 XGBOOST (eXtreme Gradient Boosting)

XGBoost (extreme Gradient Boosting) is an advanced implementation of the gradient boosting algorithm. It has high predictive power and is almost 10 times faster than the other gradient boosting techniques[5].

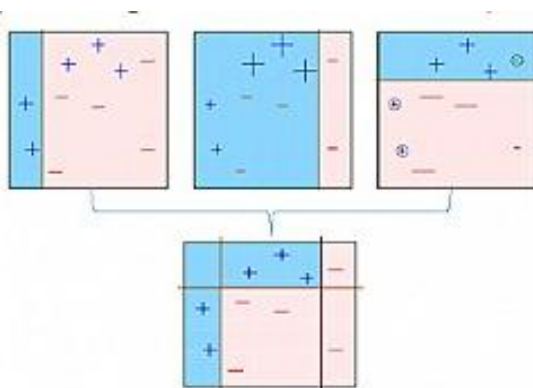


Figure 3 : XGBOOST

For batsman XGBoost gives us 70.12% cross validation score and 3.28% standard deviation while for bowler 70.70% cross validation score and 3.68% standard deviation.

5.4 STACKING (Advanced Ensemble Tech-nique)

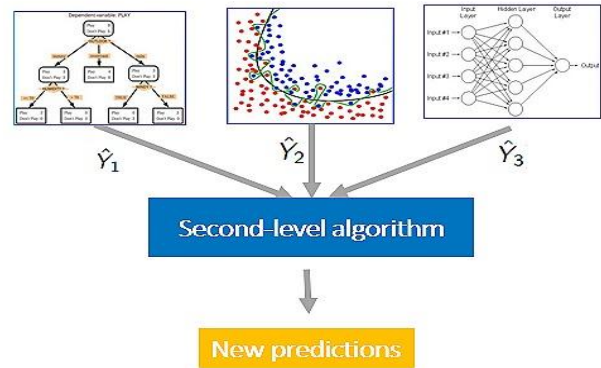


Figure 4: Stacking

Stacking is an ensemble learning technique that uses predictions from multiple models (In this experiment we use XGB Classifier and Random Forest Classifier) to build a new model[5].

For batsman it gives us 74.33% cross validation score and 3.88% standard deviation while for bowler 82.41% cross validation score and 3.94% standard deviation.

VI. RESULTS AND DISCUSSION

Classifier	Batsman	Bowler
Decision tree	59.12% (4.43%std)	59.70% (7.93%std)
Random forests	72.14% (2.95%std)	62.05% (4.49%std)
XGBoost	70.12% (3.28%std)	70.70% (3.68%std)
Stacking	74.33% (3.88%std)	82.41% (3.94%std)

We can see in the above table that the Stacking technique gives us 74.33% validation score for batsman and 82.41% validation score for bowler performance prediction.

Some other features that affect players' performance such as weather or the nature of the wicket could not be included in this study due to the un-availability of data.

So, in this IPL season, we can select our best team in a fantasy league and win a cash prize with this machine learning model.

VII. CONCLUSION AND FUTURE SCOPE

We used four machine learning algorithms. Decision Tree, Random Forest, Xgboost and Stacking in our experiment. Stacking technique gives us best result.

There is also a point system in the fantasy league for fielding perspective which may increase points but it could not be included in this study because we found that it is all about luck. Some other features that affect players' performance such as weather could not be included in this study due to the unavailability of data. The training data can be made larger in the future work for better model learning and classification. Various other data analytics techniques can also be used for improving the accuracy of the model in future work.

REFERENCE

- [1] Fantasy Cricket, https://en.wikipedia.org/wiki/Fantasy_cricket
- [2] <https://www.iplt20.com/teams>
- [3] Kalpdrum Passi and Niravkumar Pandey, 'predicting player's performance in one-day international cricket match using machine learning', February 2018.
- [4] Fantasy cricket league dream11, <https://www.dream11.com/>
- [5] <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>
- [6] https://en.wikipedia.org/wiki/Machine_learning
- [7] <https://en.wikipedia.org/wiki/Cricket>
- [8] C. Deep Prakash, C. Patvardhan, C. Vasantha Lakshmi, 'Data Analytics based Deep Mayo Predictor for IPL-9', Volume 152 – No.6, October 2016
- [9] Debarghya Das, 'An Integer Optimization Framework for Fantasy Cricket League Selection and Substitution'.
- [10] Tim B. Swartz, 'Research Directions in Cricket'

Authors Profile

Nihal Patel Pursuing B.Tech degree in Information Technology from Charotar University of Science and Technology, Changa, Gujarat, India. Area of interest is Machine learning.



Mrudang Pandya received the B.E. degree in Information Technology from sankalchand Patel Collage of Engineering (SPCE), North Gujarat Visnagar, India, in 2010. M.Tech degree in Information Technology from U.V Patel Collage of Engineering (UVPC) Ganpat University Kherwa, Mehsana, India in 2012. Presently working as a Assistant Professor, CHARUSAT, Changa, Gujrat, India. Also pursuing PHD in computer engineering from same university. My research interest includes Soft Computing, Deep Learning, Software Engineering, Enterprise Resource Planning, Image Processing, Data mining.

