Cricket Player Selection using Machine Learning

I. Sri Ram Teja, T. Pavan Kalyan, V. Akhil Kumar Reddy

Abstract: *Cricket has always been a popular game since its invention in the world. Moreover, it became a religion in India. The selection committees like BCCI,PCB,ACB etc. pick the players based on their previous performances in domestic cricket tournaments like IPL,Ranji Trophy, Syed Mushtaq Ali Trophy etc. by committee decisions but there is no application for selection process till now. To develop an application we need player performance analysis and assessment. This paper suggests an important approach for Selecting Cricket players by Evaluating his Statistics and Provides a comparative look at machine learning techniques in cricket player selection. In this paper a model for Bowlers and Batsmen Separately was proposed which was implemented using Random Forest, AdaBoost, Support Vector Machines(SVM), LightGBM,CatBoost, Logistic Regression Linear Discriminant Analysis(LDA), Voting Classifier, Naïve Bayes. The findings obtained by the suggested methodology in this paper are the same as in the Cricket board selected team players.*

Keywords : Support Vector Machines, Naïve Bayes, AdaBoost, LightGBM, random forest, Linear Discriminant Analysis, Voting Classifier,CatBoost,Logistic Regression

I. INTRODUCTION

Cricket is the game with the most popularity in the country India. Test, ODI, T20 are recognized formats. In every format certain statistics of a player are very important for batsmen as well as bowler. Batsmen Statistics such as Batting Average, Strike Rate, Not-outs etc., For Bowler Bowling Average, Economy, Strike Rate. Although, there are many, the mentioned statistics are much important for batsmen as well as bowler in every format These make batsmen selected ,Moreover Experience plays a crucial role in Big Matches like World Cup, ICC events etc. But if the senior player is Selected and he gets injured in those Conditions ,Then our System plays a major role in selecting the players who are good at first class matches. It gives a better opportunity for players who are willing to showcase their talent in international matches .

I,Sri Ram Teja* ,Department of Computer Science and Engineering, K L University,Email:iylamsriramteja@gmail.com

T.Pavan Kalyan, Department of Computer Science and Engineering, K L University, Email: pavan.tummala55@gmail.com

V.Akhil Kumar Reddy, Department of Computer Science and Engineering, K L University,Email: vakhilreddy555@gmail.com

The main objectives of this paper:

1.To build a machine learning model for ipl franchise and BCCI selection Committee to predict the player Selection based on simple parameters – Not outs,, player Name, , Average, Strike rate.

2.To study advanced Classification techniques and understand the classification models for getting better results of prediction.

3. To deploy the machine learning model as a web application for usage.

II. LITERATURE SURVEY

Riju Chaudhary & Sahil Bharadwaj, Sakshi Lakra (2019), discusses the Active strategy for selecting Cricket team members by using Data Envelopment Analysis for the performance of cricket players and the outcomes of these papers correspond with team players selected by the cricket board .[1] Md. Muhaimenur Rahman , Md. Omar Faruque Sham, Sabir Ismail, they proposed a system that is divided into three sections: before the game begins, after one innings played and continuous wicket fall estimating the chances of winning and losing a cricket game. During the game ,they used algorithms like decision tree algorithm C5.0 ,SVM and got accuracy 80% .[2] K.Ananthapadmanabha, Dr. K. Uday Kumar (2017) proposed a system that Focus our efforts on using internal participants to match network analysis, they use techniques like graph mining, greedy technique , link mining, community mining. [3] Deepak Saraswat,Vijai Devi(2018) discusses Analyzing the Indian Cricket Team's success using weighted association Theory Mining This study is used by the team to frame game winning schemes. Factors like tossing outcome, home / away state, bat first / second, game outcome. They have used algorithms like Linear Regression, Naïve Bayes.[4] D. Thenmozhi, P. Mirunalini (2019) makes an Evaluation of whether a team can win the ongoing IPL match or not ,by using Machine Learning Algorithms such as SVM, Gaussian Naïve Bayes, Random Forest, KNN. Finally, they got the Accuracy of 75%, 80%, 55%, 75%, 80%, 80%, 75% and 84% for respective teams for Indian premier league teams respectively. [5]Md. Jakir Hossain, Md.Abul Kashem(2018) based on Statistical data they predict the performance of each player and finds the top 30 players .They used genetic algorithm and predict the final Bangladesh 14 player squad, From this t20 and Test match team can be predicted .[6] A.G,M.Imam Hossian, Saki, Yeaser,

Cricket Player Selection using Machine Learning

They were trying to predict the success of ODI cricket players using Machine Learning in the upcoming match ,they get the statistical data from trusted websites like ESPNcricinfo, Cricbuzz etc. They used ML algorithms like SVM, Linear Regression and got accuracy 91.5% accuracy for batsmen,75.3% for Bowler.

[7] TejinderSingh,Vishal Singla, Parteek Bhatia (2015) Proposed a scheme of two methods in which the score for a cricket match is estimated first by using the factors like run rate, wickets fallen, venue, batting team and next The outcomes of the match are predicted by the same variables. These two methods are implemented using the classifier Linear regression, naïve Bayes. It also found that Linear regression classifier has less error than naive Bayes classifier in the results .[8]

III. RESEARCH METHODOLOGY

All the required data scraped from espncricinfo.com [9].we collected batsmen and bowler data separately includes statistics such as name, matches, average, strike rate, not outs, high score,innings,runs,50's,100's etc. . Based on the player previous performances in international cricket, we check whether the player is selected at that time ,If the player is Selected then we add the value as yes(Y) otherwise no(N).

Batsmen Dataset. The batsmen dataset is taken form espncricinfo.com, consists of 700 observations with 15 attributes.

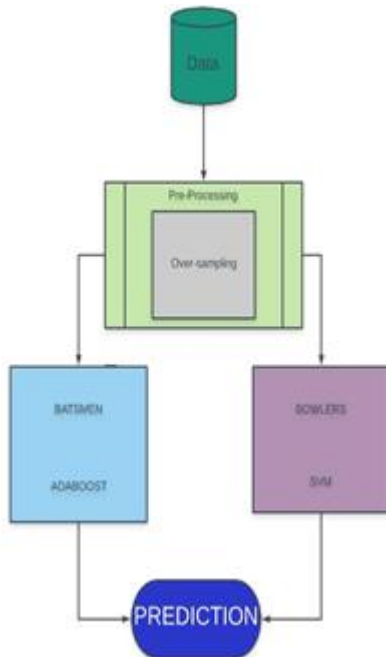Bowler Dataset. The bowler dataset is also taken espncricinfo.com consists of 650 observations with 13 features

Fig.1. Block Diagram of our model.

A. Pre-processing

In pre-processing data cleaning and analysis of relevance are performed because not all the features in the data set are needed for the report.

- **Oversampling**:

By using Libraries Rose, Smote family we increased the observations around 150, for this we used "ovun.sample" in R language that creates possibly balanced samples by random over-sampling minority examples or combination of over-and under-sampling.

B. Modelling and Experimental Results:

Following pre-processing of data, the following models were implemented: Random Forest, AdaBoost, Support Vector Machines (SVM), Light-GBM, Linear Discriminant Analysis(LDA), Voting Classifier, Naïve Bayes. The datasets are divided into Training and testing results, for evaluation respectively. The precision of the parameters is used for model performance assessment. When we're concerned with classification, we measure the accuracy after test data is expected. And Accuracy is the percentage of our model which shows how good enough our model is to predict with data testing. Accuracy is a measure of our success.

$$Accuracy = (TN + TP) / (TN + FP + FN + TP). \qquad (1)$$

C. RESULTS AND ANALYSIS:

The output parameter is tabulated in this section.

TABLE I.   ACCURACY OF MODELS OVER BATSMEN DATASET

Algorithm	Accuracy
Random Forest	98.68%
AdaBoost	98.68%
SVM	88.15%
Light-GBM	98.54%
Linear Discriminant Analysis	82.89%
Voting Classifier	97.64%
Naïve Bayes	84.21%

From the table 1, we can clearly observe that Random Forest and AdaBoost It reached the highest can

TABLE II.   ACCURACY OF MODELS OVER BOWLER DATASET

Algorithm	Accuracy
Catboost	59.14%
Logistic Regression	57.34%
SVM	82.07%
Naïve Bayes	80.76%

From the table 2, we can state that SVM has the highest accuracy 82.07% and Logistic Regression has the least with 57.34%.
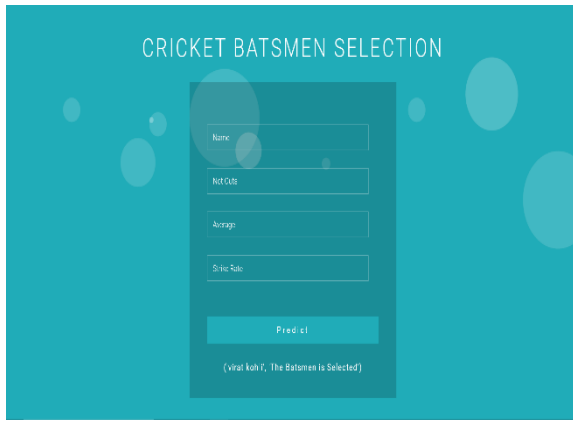
Fig.2. Output of Web Application for a batsman

Fig.3. Output of Web Application for a bowler

We can clearly predict that whether the batsmen or bowler is selected based on given input Statistics.

IV. CONCLUSION AND FUTURE WORK

The paper presents a performance study of classification algorithms for the machine learning using Cricket batsmen and bowler datasets. For batsmen dataset Random Forest and AdaBoost outperformed among all other algorithms. For bowler dataset SVM has high accuracy among other algorithms. By conducting those experiments on a large-scale real-world dataset, this work can be extended. Our experimental findings show for batsmen Random Forest and AdaBoost achieved an accuracy of 98.68% whereas for bowler SVM made a high percentage in accuracy with 82.07%. The results obtained by the model are synchronizing with the data of recent Indian Cricket team against South Africa. Our model is a Web application where it can be used as an application for selecting cricket players.
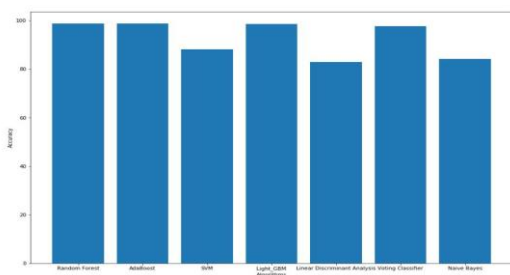
1.IPL Franchise.

2.BCCI Selection Committee.

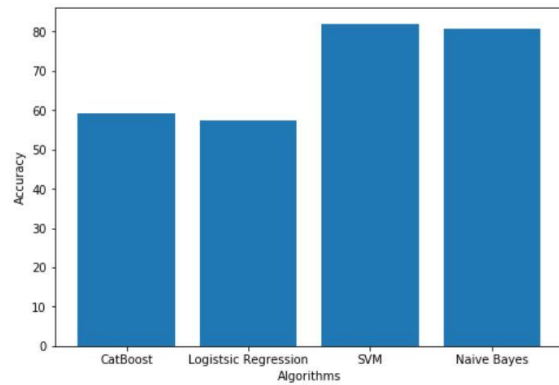**Fig.4. Accuracy of Algorithms for Batsmen**

Fig.5. Accuracy of Algorithms for Bowlers

From Fig.4, we can infer that Voting Classifier was exceptional for the batsmen data. Furthermore, the Random Forest, AdaBoost, Light GBM outperformed very well and there is no such disparity in accuracies for these algorithms. Whereas in Fig.5, Support Vectors was the best algorithm and comparatively Naïve Bayes did well in terms of Accuracy for Bowlers data.

The whole results were carried out in Python programming language. The data considered is legit and can be used in any forms of Cricket as this was done on the basis of One-Day International format of the game.

REFERENCES

1. Chaudhary, R., Bhardwaj, S., & Lakra, S. (2019, February). A DEA Model for Selection of Indian Cricket Team Players. In *2019 Amity International Conference on Artificial Intelligence (AICAI)* (pp. 224-227). IEEE.
2. Rahman, M. M., Shamim, M. O. F., & Ismail, S. (2018, October). An Analysis of Bangladesh One Day International Cricket Data: A Machine Learning Approach. In *2018 International Conference on Innovations in Science Engineering and Technology (ICISET)* (pp. 190-194). IEEE.
3. Ananthapadmanabha, K., & Uday Kumar, K. (2017, May). Match fixing network analysis to verify nearness among internal participants of a cricket match. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1043-1048). IEEE.
4. Saraswat, D., Dev, V., & Singh, P. (2018, September). Analyzing the performance of the Indian cricket team using weighted association rule mining. In *2018 International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 161-164). IEEE.
5. Thenmozhi, D., Mirunalini, P., Jaisakthi, S. M., Vasudevan, S., Kannan, V. V., & Sadiq, S. S. (2019, February). Moneyball-Data Mining on Cricket Dataset. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)* (pp. 1-5). IEEE.
6. Hossain, M. J., Kashem, M. A., Islam, M. S., & Marium, E. (2018, September). Bangladesh Cricket Squad Prediction Using Statistical Data and Genetic Algorithm. In *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)* (pp. 178-181). IEEE.
7. Anik, A. I., Yeaser, S., Hossain, A. I., & Chakrabarty, A. (2018, September). Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms. In *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)* (pp. 500-505). IEEE.
8. Singh, T., Singla, V., & Bhatia, P. (2015, October). Score and winning prediction in cricket through data mining. In *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)* (pp. 60-66). IEEE.
9. ESPN cricinfo, http://www.espncricinfo.com/.

AUTHORS PROFILE

Ialam Sri Ram Teja is a B.Tech student in the K L University ,India. Currently, persuing his final year undergraduation .Recently placed in reputed IT company .His research interests are Machine Learning,Data mining. He is a five star coder in Hacker rank

Tummala Pavan Kalyan is a B.Tech Student in the K L University,India,He is also a final year student ,Recently placed in Top IT Company in India,He is well specialized in computational Intelligece.

V,Akhil Kumar Reddy is a B.Tech student in KL University,India,Currently studying my undergraduation final year,Looking for persuing higher studies in Abroad.His research interest is Data mining.

71