Predicting Best XI

Manasvi Dobariya, Aisha Nagrecha, Dishita Madani, Dhruv Shah

[#]*School of Engineering and Applied Sciences, Ahmedabad University*

[1]manasvi.d@ahduni.edu.in

[2]aisha.n@ahduni.edu.in

[3]dishita.m@ahduni.edu.in

[4]dhruv.s3@ahduni.edu.in

***Abstract* — Cricket is a well known sport, especially the T20 format. It incurs maximum uncertainty, which rises due to the player's performance, venue, opposition team, player's form, etc. The craze of cricket has given rise to several fantasy league platforms such as Dream 11, where we are expected to form a team of 11 players out of playing 22, which are predicted to perform very well for that match. While selecting the players, all these factors are to be considered: for batsmen, they must be able to score maximum runs while the bowlers must be able to take maximum wickets. In this project, we have taken these two main features into account i.e. 'number of runs' and 'number of wickets' and used them to classify using different classification algorithms. We have rated the players from the data obtained from the previous years tournament and classified into different clusters. Thereafter, the players are classified into different classes using Random Forest and Naive Bayes classification models.**

***Keywords*— Clustering, Regression, Fantasy Points, Random forest, Classification, Cricket, Optimisation. Linear Programming, Indian Premier League.**

I. Introduction

Cricket is a very popular game it includes following types of players:

1) Batsman
2) Bowler
3) All Rounder
4) Wicket keeper

Performance of a team depends highly upon performance of individual players and a player's performance in each match depends on several parameters like location, past records, current form, average rate, strike rate (batsman), economy rate (bowler), number of innings, opposition team etc. So we are developing a model which could predict the best 11 players for that particular match.In this project we attempt to predict individual player's performance in IPL, by analyzing their previous records through supervised learning and we rate every player's performance through clustering.

II. Literature Survey

Nilesh M. Patil, Bevan H. Sequeira, Neil N. Gonsalves and Abhishek A. Singh [1] predicted a Cricket Team based on the past performances using Random Forest Algorithm and Decision Tree Classifier. The proposed model tries to eradicate the biased selections and gives the best decision.

Sonu Kumar and Sneha Roy [2] predicted score and classified players using Multiple Linear Regression (MLR), Multilayer Perceptron (MLP) and Classification Algorithms. MLR is used for determining the relationship between the random variables. MLP, a feed-forward artificial neural network, is used for mapping input to designated output.

Nihal Patel and Mrudang Pandya [3] predicted the player's performance using Supervised Machine Learning techniques like Decision Tree, Random Forest, XG Boost and Stacking. Decision Tree was used for creating class label training tuples. Random Forest is basically an ensemble of Classification and Regression where each tree is dependent on a random vector sampled independently and with the same distribution of all the trees in the forest. XG Boost is basically used to get higher predictive power and faster results compared to it's contemporary gradient boosting techniques. Stacking is used to learn from the multiple models (XGBoost and Random Forest) to build a new model.

Kalpdrum Passi and Niravkumar Pandey [4] predicted the success of players in One-Day International Cricket Matches using the Regression Model. The Regression Model aims at predicting how many runs a batsman will score and how many wickets a bowler will be able to get.

Madhav Goswami and Abhishek Anand [5] predicted the team for fantasy application Dream11 using Random Forest, XG boost, CatBoost and Shapley Additive Explanations.

Lakshmi Ajay [6] has predicted the fantasy cricket team using Linear Programming. The main aim has been to maximise the fantasy points following various constraints in team selection.

Implementation

A. Working with the Dataset

The dataset chosen for working on the problem, contained ball to ball information of all the matches that have been contested in the history of Indian Premier League right from 2008 to 2020. The chosen dataset was separated into three categories - one for batsmen, one for bowlers and one for all-rounders. After dividing, the ball by ball data of a particular match was combined to make it a match-wise data of the player.

B. Working with Clustering

Clustering came into the scenario for separating the player performances into different categories of points for the players. Here we have used K-means clustering algorithm (K=5). Batsmen in particular were clustered using the information of Strike Rate, Runs Scored and No. of balls faced. Bowlers were clustered using the Economy, Wickets taken and No. of overs delivered. While the all-rounders were

clustered using the combined effects of Runs scored, Strike-Rate, Wickets taken and Economy, also considering a decent amount of participation in the match.
For eg. an uncapped player, would be in the category of 8 points while a player of calibre like Virat Kohli would be in the 10 point category. Here we know the difference but for each and every player it would be difficult hence clustering was used.

C. Working with Classification

Classification was applied to consider the performance of the players on the basis of venue, pitch type and other factors regarding the ground. Applied algorithms were Naive Bayes Classifier and Random Forest Classifiers. The model was fitted well on the training dataset as well as test dataset.

D. Working with Regression

Regression analysis basically allows us to predict an outcome variable (y) based on the value/s of one or more predictor variables (x). The proposed work uses Regression to predict the fantasy points of a player on the basis of data feeded to it. Random Forest Regressor is used for building the model. Data feeded as predictor variables includes teams contested, the average fantasy points of the player, venue, and attributes related to the player. Overfitting was observed as the accuracy on training data was higher comparatively. Hyper Parameter tuning was applied to overcome it. The importance of the features gave a better understanding for tuning the model. Finding the best parameters, data was fitted on the model. The problem of overfitting was somewhat solved doing so. The models were built separately for different roles (wicket-keeper batsman, bowlers and all-rounders).

E. Working with Linear Programming

Linear Programming has been used to obtain the fantasy cricket team from the playing teams. In this model, the predicted fantasy points obtained after Regression were taken into consideration. There were various constraints taken in account. The main constraints are:

i. Team constraints:

- Min players from each team = 4
- Max players from each team = 7

ii. Player Role:

- No of wicket keepers = 1
- No of batsmen = (3-5)
- No of all-rounders = (1-3)
- No of bowlers = (3-5)

Keeping these constraints in mind and to maximise the fantasy points, this model was used.

III. Results

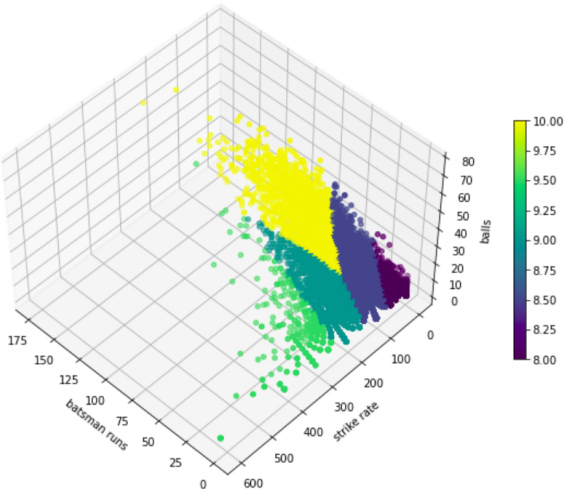

Fig. 1 Clusters of batsmen with points ranging from 8 to 10 (8,8.5,9,9.5,10) for each match performance.

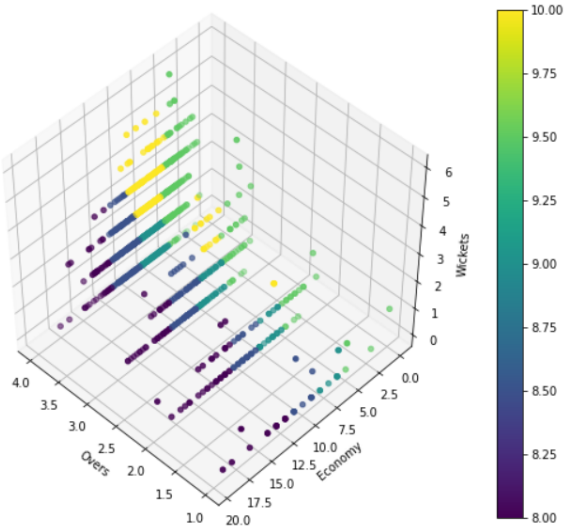

Fig. 2 Bowler clusters separated based on match wise performance with point range 8 to 10 (8,8.5,9,9.5,10)

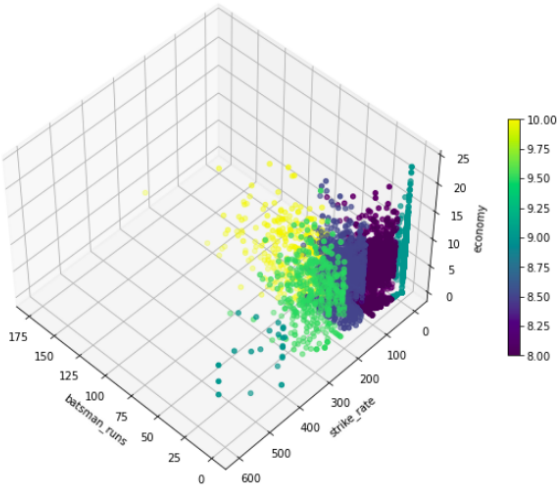

Fig. 2 All rounder clusters separated based on match wise performance with point range 8 to 10 (8,8.5,9,9.5,10)

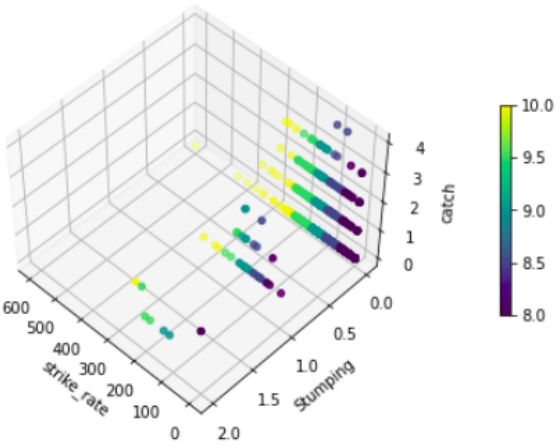

Fig. 4 Wicket keeper clusters separated based on match wise performance with point range 8 to 10 (8,8.5,9,9.5,10)

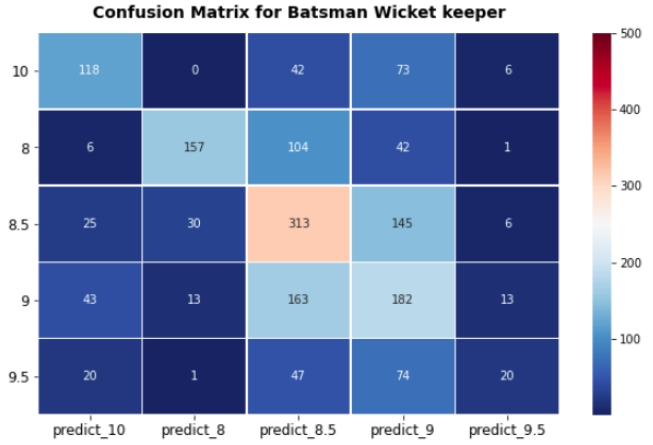

Fig. 5 Confusion matrix of the random forest classifier used to classify match performances of batsmen and wicket keepers

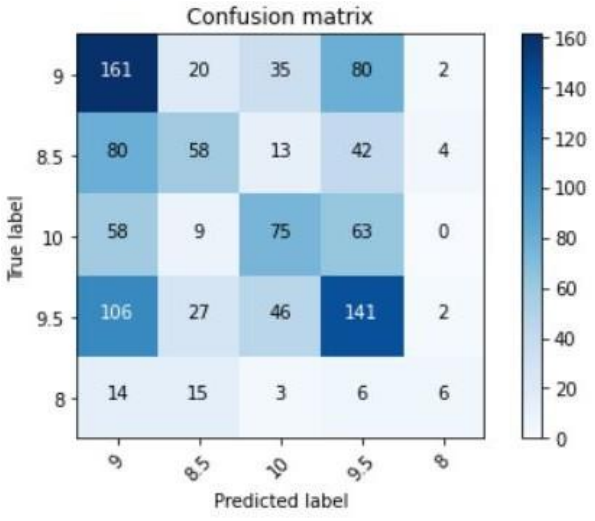

Fig. 6 Confusion matrix of the random forest classifier used to classify match performances of bowlers

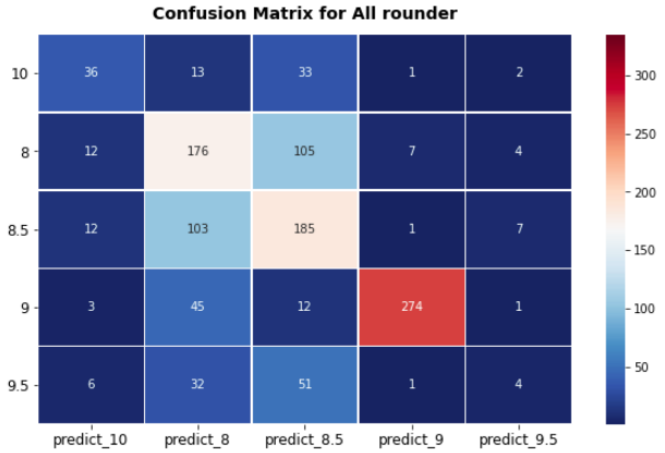

Fig. 7 Confusion matrix of the random forest classifier used to classify match performances of all rounders

Classification report for Batsman and Wicket Keeper

	precision	recall	f1-score	support
10	0.57	0.50	0.53	239
8	0.79	0.51	0.62	310
8.5	0.47	0.59	0.53	519
9	0.34	0.45	0.39	414
9.5	0.40	0.12	0.18	162
accuracy			0.48	1644
macro avg	0.52	0.43	0.45	1644
weighted avg	0.51	0.48	0.48	1644

Fig. 8 Classification report for batsmen and wicket keepers

	precision	recall	f1-score	support
0	0.42	0.51	0.46	312
1	0.41	0.47	0.44	349
2	0.46	0.34	0.39	163
3	0.38	0.09	0.15	32
4	0.46	0.36	0.40	210
accuracy			0.43	1066
macro avg	0.43	0.35	0.37	1066
weighted avg	0.43	0.43	0.42	1066

Fig. 9 Classification report for bowlers

classification report of All rounder

	precision	recall	f1-score	support
10	0.52	0.44	0.47	85
8	0.49	0.59	0.53	304
8.5	0.48	0.60	0.53	308
9	0.97	0.82	0.89	335
9.5	0.39	0.10	0.15	94
accuracy			0.61	1126
macro avg	0.57	0.51	0.52	1126
weighted avg	0.62	0.61	0.60	1126

Fig. 10 Classification report for all rounders

	Before Hyper parameter tuning		After Hyper parameter tuning	
	Train	Test	Train	Test
Bowler	95.21%	66.58%	82.23%	67.23%
Batsman and Wicket Keeper	94.46%	54.49%	77.11%	55.73%
All rounder	95.44%	71.11%	94.05%	71.34%

Table 1: Hyper Parameter Tuning accuracy comparison - Regression model

	Players	Rating	Fantasy points	Roles
0	AR Patel	8.5	89.066	All Rounder
1	A Nortje	9.0	36.750	Bowler
2	JC Archer	10.0	49.856	Bowler
3	JC Buttler	9.5	9.008	Wicketkeeper batsman
4	K Rabada	9.5	67.806	Bowler
5	MP Stoinis	8.5	40.826	Batsman
6	PP Shaw	8.0	21.268	Batsman
7	R Ashwin	9.0	39.764	All Rounder
8	R Parag	8.0	26.244	Batsman
9	SS Iyer	10.0	33.440	Batsman
10	S Gopal	8.5	25.582	All Rounder

Table 2: Optimised team obtained using Linear Programming

IV. Conclusions

After proper selection of features, we were able to get considerable good accuracy after performing Regression. Since, in this project, we were supposed to use pure Machine Learning techniques, the accuracy obtained can further be improved if some Deep Learning techniques would have been applied.

V. References

[1] Nilesh M. Patil, Bevan H. Sequeira, Neil N. Gonsalves and Abhishek A. Singh, "Cricket Team Prediction Using Machine Learning Techniques", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3572740

[2] Sonu Kumar and Sneha Roy, "Score Prediction and Player Classification Model in the Game of Cricket using Machine Learning", International Journal of Scientific & Engineering Research, Volume 9, Issue 8, August-2018, ISSN 2229-5518, https://www.ijser.org/researchpaper/Score-Prediction-and-Player-Classification-Model-in-the-Game-of-Cricket-Using-Machine-Learning.pdf

[3] Nihal Patel and Mrudang Pandya, "IPL Player's Performance Prediction", International Journal of Computer Sciences and Engineering, Volume 7, Issue 5, May-2019, E-ISSN 2347-2693 https://www.ijcseonline.org/pdf_paper_view.php?paper_id=4268&

[4] Kalpdrum Passi and Niravkumar Pandey, "Predicting Players' Performance in One Day Internation Cricket Matches Using Machine Learning", https://airccj.org/CSCP/vol8/csit88310.pdf

[5] Madhav Goswami and Abhishek Anand, "Dream11 Team Predictor with Python and Machine Learning", https://medium.com/analytics-vidhya/dream11-team-predictor-with-python-and-machine-learning-f0dfce1489eb

[6] L. Ajay, "Creating a Fantasy Cricket Team - Application of Linear Programming," *Medium*, 10-Feb-2021 https://towardsdatascience.com/creating-a-fantasy-cricket-team-application-of-linear-programming-4b60c261702d.

[7] Dream11 Fantasy Points distribution system. https://www.dream11.com/games/point-system