A New Team Selection Methodology using Machine Learning and Memetic Genetic Algorithm for IPL-9

Chellapilla Deep Prakash
Indian Institute of Management

Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 5, Issue 4
April 2016

A New Team Selection Methodology using Machine Learning and Memetic Genetic Algorithm for IPL-9

C Deep Prakash
Deptt. Of Electrical Engg.
Dayalbagh Educational Institute,
Agra-282005, INDIA

Abstract

IPL franchises spend massive amounts of money is ensuring that they have the best batting and bowling options according to their game plan. An important problem is to select the playing XI from the available options. The present work focuses on Machine learning based data analytics to provide a good approach to solve this problem. A detailed performance ranking scheme is developed based on Random Forests based Recursive Feature estimation algorithm to rank the players with respect to the other players in fray in IPL 9. The ranking scheme provides percentile scores to the players for their batting and bowling performance and enables them to be compared against each other. The sum of the percentile scores for batting and bowling for the players in a given selection of Playing XI provide the fitness function for measuring the suitability of the team. An SGA and MGA are designed and implemented for search and optimization of the Playing XI and the results are compared. The MGA provides much more balanced XIs than the SGA and can be gainfully utilized for this purpose.

Keywords

Cricket, IPL IX, Playing XI, team selection

1. Introduction

Test cricket, ODI and Twenty20 (T20) are the three formats in which the game of cricket is being played. T20 is a 20 overs a side match which is usually over in 4 hours. T20 cricket was first introduced in 2003 in England and it became very popular among the cricket fans, because the short format allowed one to enjoy the complete match in one evening. 20 overs (120 = 20 x 6 legal ball deliveries) are allowed in T20 matches for each batting side to score from if they have wickets. The side that scores more runs within the stipulated overs wins.

Indian Premier League (IPL), a T20 tournament, was started in 2008 by Board of Control of Cricket in India (BCCI) [1]. The IPL created eight franchises assigned to eight of the largest cities in India. The teams were franchisee driven. The players were selected through competitive bidding from a pool of available players. The BCCI has been organizing the IPL T20 cricket tournament in each year. 8 IPL tournaments have been held till date and IPL IX is scheduled to be held beginning in April, 2016.

The use of analytical methods is very useful in cricket. Batting, bowling and fielding are the three main departments of the game. There is a huge demand for cricket related statistical studies because of the popularity of the game and the staggering amounts of money involved. These statistics give clear picture of the performance of various players. Followers of the game, especially in India, are keen followers of its statistics also.

In IPL each franchisee has 22-28 players with them. The playing eleven selection problem is to select the playing eleven for a particular match. A number of considerations have to be taken into account for deciding the playing eleven. Some of these are statutory, some are essential and some are desirable.

Statutory condition has been imposed by BCCI to make IPL interesting as well as more competitive. There should not be more than 4 foreign players in the playing eleven in a match. This rule is to be followed in every match by every team.

Essential requirements are those which a team manager ensures while selecting his playing eleven for any match. The captain has to be included in the playing eleven. One wicket keeper has to be selected in the playing eleven. There must be two openers who will open the batting for the team. There must be at least two fast bowlers or fast bowling all-rounders to start the bowling attack with the new ball.

Desirable requirements are those which strengthen a team's batting and bowling lineup. Batting and bowling strengths should be maximized within the constraints applicable. There need to be at least 7 batsmen / batting all-rounders / all-rounders including the wicketkeeper and at least 4 bowlers / bowling all-rounders / all-rounders. The team should have at least six bowling options because if one of the bowlers has an off day there has to be an alternative available.

The problem of team selection or playing eleven selection is difficult because each player (batsmen or bowler) has his own skills and capabilities. It is very difficult to compare two players to decide who is the better one and should be given preference in the Playing XI. For this purpose, some mathematical ranking scheme should be formed which quantifies the capabilities of different players with respect to their batting and bowling skills and enables a comparison.

Even with such a ranking mechanism in place the problem still remains that comparison is difficult in making a choice. For example a difficult choice could be as follows assuming only one slot is left and so only one of them can be selected:

a) An excellent batsman
b) A good batsman who is also a good bowler

Alternatively, we could have the following choice:

a) An excellent bowler
b) A good batsman who is also a good bowler

T20 cricket is predominantly a batsman's game and so strategically it might be preferable to choose an excellent batsman in the first choice and an excellent all-rounder in the second choice. But this choice could also be reversed if the team already has, say six bowling options and does not need any more bowling options. Another factor that could affect the decision is that the team already has a top order that is firing well and so it is not expected that an extra batsmen would actually get to bat.

All these decisions can be made more convenient and result in better team selection if a good ranking scheme can be designed for mathematically quantifying the skills and then using a selection algorithm for automatically examining the huge variety of choices and maximizing some performance index for the team as a whole. Of course, the selection algorithm has to ensure that the final XI that is selected would satisfy all the constraints while maximizing the chosen performance index.

Some studies related to cricket reported in the literature are as follows. Optimal batting strategies using dynamic programming model were developedby Clarke [2]. Alternative batting averages when the batsman remains not-out in one-day cricket was proposed by Kimber and Hansford [3] and Damodaran [4]. Barr and Kantor [5] proposed a method based on batting averages and strike rates. Borooah and Mangan [6] explored batting performance for test matches. Norman and Clark [7] and Ovens and Bukeit [8] applied mathematical modeling approach to optimize the batting order of a team. Lewis [9] analyzed player performance using Duckworth/Lewis percentage values. Van Staden [10] used a graphical method to analyze batting and bowling performance in cricket. Lakkaraju and Sethi [11] described a Sabermetrics style principle to analyze batting performance in cricket. Lemmer [12-14] considered performance analysis using averages and strike rates for bowling and batting. Saikia et al. [15] evaluated the performance of all-rounders in IPL.

The rest of the paper is organized as follows. In section 2, the ranking methodology for ranking the batsmen and bowlers in fray in the IPL IX is described that imposes an ordering on the players in terms of their batting and bowling capabilities. This ranking scheme is based on a machine approach using Recursive Feature elimination to create a comprehensive performance index for batting and another for bowling. In section 3, a memetic Genetic Algorithm is designed that optimizes an appropriate fitness function based on the rankings obtained in section 2. The memetic GA incorporates suitable local search mechanisms within the outer GA loop to provide the optimal solution. In section 4, the results of computational experiments for obtaining the playing XIs out of the available teams of each of the 8 franchises in fray in IPL IX are presented to demonstrate the effectiveness of the proposed approach. The relative strengths and weaknesses are analyzed based on the indices computed in the methodology. Some conclusions and directions are presented in section 5. The references are given in section 6.

2. Proposed Performance Ranking Scheme for IPL IX

Any automated team selection scheme has to be based on a strong mathematical model that evaluates the relative performance of the players in terms of various capabilities including batting, bowling, wicket-keeping and captaincy and enables a decision making process that decides the playing XI. One of the most popular indices for this comparative performance evaluation is the Most Valuable Player Index (MVPI). However, this index has its own shortcomings and favors players who have been playing for a longer time [16].

It is, therefore, necessary to create better indices that enable this comparison to be made in a more comprehensive manner. Deep Prakash et al. [16] present a Deep Performance Index that ranks the performance of bowlers and batsmen based on their performance in the IPL upto season VII and their performance in T20 matches overall upto that point in time. However, T20 is an evolving game and the performances of both batsmen and bowlers are getting better and better with more exposure as shown in [17]. This means that the ranking scheme also has to take into account the latest performance statistics upto IPL season VIII and correspondingly the latest overall T20 performance statistics.

In this paper, a scheme that does this in a precise mathematical manner is presented to devise a ranking scheme that is novel and based on up-to-date statistical performance analysis. This ranking is then utilized to present an automated procedure for selecting the Playing XIs for various franchises for IPL IX.

In order to evaluate the batting capability in T20, five indices are considered as follows.

1) HardHitter = (4*Fours + 6*Sixes) / Balls faced by player
2) Finisher = Number of times Not Out/ Total number of played Innings
3) FastScorer =Total runs scored/ Total balls faced
4) Consistent = Total runs scored/ Total number of innings in which he got out
5) RunningBetweenWickets = (Total runs scored - (4*Fours + 6*Sixes))/ Number of balls faced without boundary

The above five measures are typical T20 measures and provide a more detailed analysis of the performance of the batsmen. Similarly, in order to define Bowling Capability five indices are considered as follows.

1) Economy= Total number of runs conceded / Total number of overs bowled
2) WicketTaker= Total number of balls bowled / Total number of wickets taken
3) Consistent = Total number of runs conceded / Total number of wickets taken
4) BigWicketTaker = Number of times four wickets or five wickets taken / Number of innings played
5) ShortPerformance = (Number of wickets taken – 4* Number of times four wickets – 5* Number of times five wickets taken) / (Number of innings played - Number of times four wickets or five wickets taken)

Ten indices were computed for each batsman and each bowler considering their latest T20 performance statistics available and their IPL performance upto season VIII. In order to select the features which matter most and their relative importance, Recursive Feature Elimination algorithm with Random Forests, a popular machine learning approach is used.Recursive Features Elimination using the Random Forests Algorithm works as follows. Random Forests algorithm is a state of the art procedure for various classification and regression tasks.

In addition to constructing each tree using a different bootstrap sample of data, random forests change how the classification or regression trees are constructed. In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node. This strategy has been shown to perform better than many other classifiers, including discriminant analysis, support vector machines and neural networks, and is robust against overfitting. It is very user-friendly because it has only two parameters (the number of variables in the random subset at each node and the number of trees in the forest) and is usually not very sensitive to their values.

The algorithm proposed in this paper for determining the relevant features for player ranking performs Recursive Feature Elimination (RFE) [18]. In this approach, the algorithm fits the model to all predictors which are the indices in the current work. Each predictor is ranked according to its importance to the model. At each iteration of feature selection, the N top ranked predictors are retained, the model is refit and performance is assessed. The value of N with the best performance is determined and the top N predictors are used to fit the final model. The predictor rankings are recomputed on the model on the reduced feature set. Resampling methods (e.g. cross-validation, the bootstrap) are used to reduce variability caused by feature selection when calculating performance. The steps in the algorithm are encapsulated inside an outer layer of 10-fold cross-validation to ensure better robustness of results and provide better estimates of performance. However, this makes the algorithm compute intensive. A consensus ranking is used to finally determine the best predictors to retain.

The weightages and features for batting and bowling obtained through the above procedure are shown in table 1.

Table 1: Batting and Bowling features and their relative importance

Batting Features	Weightage	Bowling Features	Weightage
T20_Consistency	0.4851	T20_Consistency	0.3011
IPL_Consistency	0.3574	IPL_ShortPerformance	0.3001
T20_FastScorer	0.1251	IPL_Consistency	0.2327
T20_FScore	0.0275	T20_WicketTaker	0.1429
IPL_FastScorer	0.0047	IPL_BigWicketTaking	0.0480

After selecting the features, they have been normalized and composite ranking index is determined for evaluating batting performance and bowling performance. The formulae obtained are as follows.

Batting Index: T20_Consistency*0.4851+ IPL_Consistency*0.3574+ T20_FastScorer*0.1251+ T20_FScore*0.0275+ IPL_FastScorer*0.0047

Bowling Index: T20_Consistency*0.3011+ IPL_ShortPerformance*0.3001+IPL_Consistency*0.2327+ T20_WicketTaker*0.1429+ IPL_BigWicketTaking*0.0480

The indices have then been calculated using the above formulae. The top 10 batsmen and bowlers in fray in IPL IX according to these indices are shown in table 2.

Table 2: Top 10 batsmen and bowlers in fray in IPL IX according to the proposed tanking scheme

Category	Top Ten Players [Rank]
Batting	C.Gayle [0.968], D.Miller [0.934], MS Dhoni [0.912], S.Marsh [0.912], K.Peterson [0.884], S.Raina [0.877], JP Duminy [0.863], V.Kohli [0.843], AB De Villiers [0.838], D.Warner [0.838]
Bowling	L.Malinga [0.972], M.Starc [0.957], M.McCleneghan [0.921], S.Gopal [0.901], I.Tahir [0.901],SA.Hasan [0.895],S.Narine [0.894], S.Sharma [0.892], N.Coulter Nile [0.859], D.Wiese [0.859]

3. Memetic Genetic Algorithm for selection of Playing XI

Simple Genetic Algorithm (SGA) is a search and optimization technique that is inspired from the natural selection occurring in the evolution of species. It is a population based technique where a population of solutions is generated and goes through a process that simulates evolution in order to generate better solutions with better values of objective function. The approach is very generic and can be used in almost all kinds of search and optimization problems as long as one can define a function that mathematically characterizes the "goodness" of solutions represented in the form of a solution string. It does not need any additional information like the gradient information or the Hessians. Following the theory of natural evolution it provides higher probability of surviving and producing offsprings to the solutions that have better values of the objective function or the fitness function. This kind of guided randomness lies at the heart of the genetic search algorithm.

Pseudo Code for SGA

Initialize: Initialize population randomly using binary representation
While stipulated Number of iterations are not completed **do**
Evaluate the fitness value of each member of the population
Assign number of children to be produced by each member
Evolve a new population using crossover and mutation
Select best fitness solutions for next generation
end while

SGAs are very general and can be and have been used with advantage in a variety of applications. However, the main problem with this approach is the slow convergence rate. Attempts to speed up the convergence reduce the generality of the approach as the strategies that work with one optimization problem do not work with the others. This is also in line with the No Free Lunch Theorem in this context. However, methods that improve definitely of use in solving some particular problem. One of the approaches to speed up convergence that has been suggested in literature is to incorporate local search based on detailed domain information along with the outer GA loop. The resulting algorithms are broadly classified as Memetic Genetic Algorithms (MGAs).

Thus a Memetic GA is an improved version of SGA with local improvements incorporated within the SGA for population search. MGAs are referred to as Baldwinian algorithms because they work on the principle of Baldwinian theory of evolution which suggests a mechanism whereby evolutionary progress can be guided towards favourable adaptation without the changes in individual's fitness arising from learning or development being reflected in changed genetic characteristics.

Pseudo Code for MGA

Initialize: Initialize population randomly using binary representation

While: Number of iterations are not completed **do**

Evaluate the fitness value of each string in the population

Evolve a new population using crossover and mutation
Select solutions with better fitness for next generation
foreach new string in the population **do**
ifcaptain is missing in the playing XI **then** introduce captain by replacing a lower index player
ifwicket keeper is missing in the team **then** introduce wicket keeper by replacing a lower index player
if team has too many bowlers thus lowering batting index
t**hen**introduce specialist batsman to replace a bowler
else if team does not have enough bowlers thus affecting bowling index
thenintroduce specialist bowler by replacing a lower indexplayer
else

for each player with higher batting or bowling index than the one in playing XI who is left out

replace thelowest index player in the Playing XI with this player.

Accept the change only if it results in improvement in the overall fitness of the selection.

endfor

endfor

endwhile

A binary representation is used to represent a chromosome where a 1 implies that the player is selected in the playing eleven and a 0 implies that the player is not selected in the team. Random Initialization with 4 overseas players and 7 Indian players is used to start the Memetic GA. A fitness function is defined and the best parents are selected for crossover. Mutation with 0.01 % probability is used. After each generation 20% of the old chromosomes from the previous generation are allowed to pass to the next generation. The algorithm terminates after 1000 runs and gives the optimal solution found so far. The objective function that is to be maximized is the sum of batting and bowling indices of the players in the playing XI. Since both batting and bowling indices are part of the fitness function the GA evolves strings that have good values for both together. This may result in some solutions being skewed in favor of batting capabilities or in favor of bowling capabilities. Some key players like the captain or the wicketkeeper may also be missed out in the string formed by random crossover or mutation. Local heuristics are employed to rectify the solution in such cases.

4. Results and Discussion

The Playing XIs which have been obtained using SGA are shown in table 3. In SGA the teams obtained are sometimes unbalanced i.e. they are slightly biased towards bowlers or batsmen. There are cases when wicket-keepers are not included by the SGA in the playing eleven as their batting index does not warrant their inclusion. However, no Playing XI can be created without a wicketkeeper and so corrective measure has to be taken.

Table3: Teams obtained with SGA

Team	Player [Batting, Bowling]
DD	S.Iyer [0.727,0], S.Samson [0.452,0], JP Duminy [0.863,0.363], C.Morris [0.424,0.799], A.Mishra [0.109,0.854], N.Coulter Nile [0,0.859], I.Tahir [0,0.901], Z.Khan [0,0.624], M.Shami [0,0.613], J.Yadav [0,0.459], C.Miland [0,0.568]
GLR	A.Finch [0.747,0.144], D.Smith [0.609,0.460], S.Raina [0.877,0.223], E.Dwivedi [0.610,0], D.Bravo [0.558,804], J.Faulkner [0.412,0.735], R.Jadeja [0.370,0.419], P.Sangwan [0,0.543], D.Kulkarni [0,0.718], P.Tambe [0,0.655], S.Jakati [0,0.535]
KKR	G.Gambhir [0.622,0], R.Uthappa [0.705,0], Y.Pathan [0.654,0.326], S.Hasan [0.285,0.895], A.Russel [0.586,0.425], J.Hastings [0.187,0.760], P.Chawla [0.170,0.654], S.Narine [0,0.894], A.Rajpoot [0,0.638], J.Unadkat [0,0.740], K.Yadav [0,0.542]
KXIP	S.Marsh [0.912,0], M.Vohra [0.511,0], D.Miller [0.934,0] ,G.Maxwell [0.456,0.197], M.Singh [0.390,0.735], A.Patel [0,0.629], R.Dhawan [0.433,0.308], M.Johnson [0,0.751], M.Sharma [0,0.833], S.Sharma [0,0.892], S.Thakur [0,0.737]
MI	R.Sharma [0.812,0.376], L.Simmons [0.776,0.512], K.Pollard [0.789,0.630], H.Pandya [0.670,0.246], H.Singh [0.241,0.646], R.Vinay Kumar [0.126,0.621], K.Pandya [0,0.542], L.Malinga [0,0.972], M.McCleneghan [0,0.921], J.Bumrah [0,0.516], S.Gopal [0,0.901]
RCB	C.Gayle [0.968,0.241], S.Watson [0.819,0.546], V.Kohli [0.843,0.128], S.Binny [0.278,0.476], D.Wiese [0.490,0.859], H.Patel [0,0.609], A.Karnwar [0,0.643], M.Starc [0,0.957], V.Aron [0,0.525], V.Malik [0,0.523], Y.Chahal [0,0.687]
RPS	A.Rahane [0.685,0], F.DuPlesis [0.650,0.575], S.Smith [0.713,0.509], K.Peterson [0.884,0.296], MS Dhoni [0.912,0], S.Tiwary [0.678,0], A.Morkel [0.569,0.587], I.Pathan [0.431,0.547], R.Bhatia [0.225,0.504], A.Dinda [0,0.568], RP.Singh [0,0.722]
SRH	D.Warner [0.838,0], S.Dhawan [0.668,0], K.Williamson [0.616,0.086], T.Suman [0.383,0.241], Y.Singh [0.601,0.613], M.Henriques [0.618,0.616], Ben Cutting [0.331,0.410], A.Reddy [0.193,0.787], K.Sharma [0.215,0.655], B.Kumar [0,0.668], A.Nehra [0,0.812]

Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 5, Issue 4
April 2016

Team DD is biased towards bowlers and their wicket-keeper is not included in the playing eleven. In team GLR their wicket-keeper is not included in the playing eleven. In KKR again bowlers are dominating the team, the batting strength is very poor. In KXIP the wicket-keeper is missing in the playing eleven. In MI the wicket-keeper is missing and the team is dominated by bowlers. RCB is also wicket-keeper less and a bowlers dominated team. In RPS, spinner is missing in the playing eleven. In SRH, wicket-keeper is missing.

Teams which have been obtained by using MGA are shown in table 4. Many of the shortcomings of the teams in Table 3 have been effectively taken care of by the local search mechanism built into the algorithm.

Table 4: Teams obtained with Memetic GA

Team	Player [Batting, Bowling]
DD	S.Iyer [0.727,0], Q.DeKock [0.664,0], M.Aggarwal [0.338,0],K.Nair [0.308,0], S.Samson [0.452,0], JP Duminy [0.863,0.363], C.Morris [0.424,0.799], A.Mishra [0.109,0.854], N.Coulter Nile [0,0.859], Z.Khan [0,0.624], M.Shami [0,0.613]
GLR	A.Finch [0.747,0.144], D.Smith [0.609,0.460], S.Raina [0.877,0.223], E.Dwivedi [0.610,0],D.Karthik [0.458,0],D.Bravo [0.558,804], J.Faulkner [0.412,0.735], R.Jadeja [0.370,0.419], P.Sangwan [0,0.543], D.Kulkarni [0,0.718], P.Tambe [0,0.655]
KKR	G.Gambhir [0.622,0], R.Uthappa [0.705,0], M.Pandey [0.488,0],C.Lynn [0.588,0],Y.Pathan [0.654,0.326], S.Hasan [0.285,0.895], A.Russel [0.586,0.425], P.Chawla [0.170,0.654], S.Narine [0,0.894], A.Rajpoot [0,0.638], J.Unadkat [0,0.740]
KXIP	S.Marsh [0.912,0], M.Vohra [0.511,0], D.Miller [0.934,0] ,G.Maxwell [0.456,0.197],W.Saha [0.547,0], M.Singh [0.390,0.735], A.Patel [0,0.629], R.Dhawan [0.433,0.308], M.Johnson [0,0.751], M.Sharma [0,0.833], S.Sharma [0,0.892]
MI	R.Sharma [0.812,0.376], J.Butler [0.414,0], A.Rayudu [0.488,0], C.Anderson [0.573,0.253], K.Pollard [0.789,0.630], H.Pandya [0.670,0.246], H.Singh [0.241,0.646], R.Vinay Kumar [0.126,0.621],L.Malinga [0,0.972], J.Bumrah [0,0.516], S.Gopal [0,0.901]
RCB	C.Gayle [0.968,0.241], S.Watson [0.819,0.546], V.Kohli [0.843,0.128],AB De Villiers [0.838,0], K.Jadhav [0.45,0], S.Binny [0.278,0.476], D.Wiese [0.490,0.859], H.Patel[0,0.609], A.Karnwar [0,0.643], V.Aron [0,0.525],Y.Chahal [0,0.687]
RPS	A.Rahane [0.685,0], F.DuPlesis [0.650,0.575], S.Smith [0.713,0.509], K.Peterson [0.884,0.296], MS Dhoni [0.912,0], S.Tiwary [0.678,0], A.Morkel [0.569,0.587], I.Pathan [0.431,0.547], R.Ashwin [0,0.665], A.Dinda [0,0.568], RP.Singh [0,0.722]
SRH	D.Warner [0.838,0], S.Dhawan [0.668,0], K.Williamson [0.616,0.086], Y.Singh [0.601,0.613], A.Tare [0.486,0] , M.Henriques [0.618,0.616], Ben Cutting [0.331,0.410], A.Reddy [0.193,0.787], K.Sharma [0.215,0.655], B.Kumar [0,0.668], A.Nehra [0,0.812]

The teams which have been obtained by MGA more or less balanced in their batting and bowling strengths. They have one wicket-keeper in their playing eleven and they have batting and bowling options.A detailed comparison of the teams obtained using SGA and MGA is shown in table 5.

Table 5: Comparison of Batting and Bowling strengths with SGA and MGA

Team	SGA_ Batting	SGA_ Bowling	SGA_ Score	MGA_ Batting	MGA_ Bowling	MGA_ Score
DD	2.575	6.04	8.615	3.885	4.112	7.997
GLR	4.183	5.236	9.419	4.641	4.701	9.342
KKR	3.209	5.874	9.083	4.098	4.572	8.67
KXIP	3.636	5.082	8.718	4.183	4.345	8.528
MI	3.414	6.883	10.297	4.113	5.161	9.274
RCB	3.398	6.194	9.592	4.686	4.714	9.4
RPS	5.747	4.308	10.055	5.222	4.469	9.991
SRH	4.463	4.888	9.351	4.566	4.647	9.213

The MGA has improved the batting strength of 7 teams while reducing their bowling strength and in one case it has reduced the batting strength for increasing the bowling strength. Thus more realistic and better teams are obtained using the MGA.

The changes in the team while using MGA in comparison with the SGA are shown in table 6.

Table 6: Changes in the team using MGA

Team	In	Out
DD	Q.DeKock, M.Aggarwal, K.Nair	I.Tahir, J.Yadav, C.Miland
GLR	D.Karthik	S.Jakati
KKR	M.Pandey, C.Lynn	K.Yadav, J.Hastings
KXIP	W.Saha	S.Thakur
MI	J.Butler, A.Rayudu, C.Anderson	L.Simmons, K.Pandya, M.McCleneghan
RCB	K.Jadhav, AB De Villiers	V.Malik, M.Starc
RPS	R.Ashwin	R.Bhatia
SRH	A.Tare	T.Suman

Team details, key players, strengths and weaknesses are shown in table 7.

Table 7: How Teams Stand in IPL 9

Team	**Captain**	**Wicket Keeper**	**Key Players**	**Strengths**	**Weaknesses**
DD	Z.Khan	Q.DeKock	S.Samson, C.Morris, Coulter Nile	Finishing and bowling capability of allrounderC.Morris, Consistency of Q.DeKock and JP Duminy	Unpredictable batting line up with emerging Indian players S.Iyer, M.Aggarwal, S.Samson and K.Nair
GLR	S.Raina	D.Karthik	D.Bravo, J.Faulkner, R.Jadeja	Three best T20 allrounders in the form of D.Bravo, J.Faulkner and R.Jadeja	Main bowling attack lacks the fire and international experience
KKR	G.Gambhir	R.Uthappa	SA.Hasan, A.Russel, Y.Pathan	Big bash star Chris Lynn's form, All-rounders Shakib and Russel can test the opponent with both bat and bowl	Form of openers Gambhir and Uthappa can cause trouble since they haven't played any international matches recently
KXIP	D.Miller	W.Saha	S.Marsh, G.Maxwell, M.Johnson	Explosive batting line-up where anybody can change gears especially Miller, Maxwell	Bowling line up is very strong but bowling options are limited.

MI	R.Sharma	J.Butler	K.Pollard, H.Pandya, C.Anderson	Many match winners in the form of J.Butler, R.Sharma , C.Anderson and K.Pollard	Overdependent on foreign players
RCB	V.Kohli	AB De Villiers	C.Gayle, S.Watson, D.Wiese	Best top order with C.Gayle, S.Watson, V.Kohli, AB De Villiers	No big names in the bowling attack
RPS	MS.Dhoni	MS Dhoni	F.DuPlesis, S.Smith, R.Ashwin	Consistent players in the batting line up with Rahane, DuPlesis, Smith, K.Peterson	Bowling is overdependent on Ashwin. The pace attack lacks fire.
SRH	S.Dhawan	A.Tare	D.Warner, K.Williamson, Yuvraj Singh	Perfect combination of consistent, aggressive and attacking batting line up with a proper T20 approach.	No frontline bowlers apart from Nehra and Bhuvneshwar. All-rounders have to perform well with the ball.

5. Conclusions

IPL has captured the imagination of the cricket crazy nation that is India like nothing before. Franchises spend massive amounts of money is ensuring that they have the best batting and bowling options according to their game plan. An important problem is to select the playing XI from the available options. The present work focuses on Machine learning based data analytics to provide a good approach to solve this problem. A detailed performance ranking scheme is developed based on Random Forests based Recursive Feature estimation algorithm to rank the players with respect to the other players in fray in IPL 9. The ranking scheme provides percentile scores to the players for their batting and bowling performance and enables them to be compared against each other. An SGA and MGA are designed and implemented for search and optimization of the Playing XI and the results are compared. The MGA provides much more balanced XIs than the SGA and can be gainfully utilized for this purpose. A multiobjective version of the GA can also be attempted to find multiple solutions according to different criteria considering even those that are not yet included in the formulation.

6. References

[1] Indian Premier League, https://en.wikipedia.org/wiki/Indian_Premier_League
[2] Clarke, S R, "Dynamic programming in one day cricket - optimal scoring rates," Journal of the Operational Research Society, 50, 1988, pp 536 – 545.
[3] Kimber, A C and Hansford, A R, "A Statistical Analysis of Batting in Cricket," Journal of Royal Statistical Society, 156, 1993, pp 443 – 455.
[4] Damodaran, U, "Stochastic Dominance and Analysis of ODI Batting Performance: The Indian Cricket Team, 1989-2005," Journal of Sports Science and Medicine, 5, 2006, pp 503 – 508.

[5] Barr, G. D. I., and Kantor, B.S., "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Journal of the Operational Research Society, 55, 2004, pp 1266-1274.
[6] Borooah, V. K., and Mangan, J E, "The 'Bradman Class': An Exploration of Some Issues in the Evaluation of Batsmenfor Test Matches, 1877–2006.", Journal of Quantitative Analysis in Sports, 6 (3), Article 14, 2010.
[7] Norman, J and Clarke, S R, "Dynamic programming in cricket: Batting on sticky wicket," Proceedings of the 7th Australasian Conference on Mathematics and Computers in Sport, 2004, pp 226 – 232.
[8] Ovens, M and Bukeit, B, "A mathematical modeling approach to one day cricket batting orders," Journal of Sports Science and Medicine, 5, 2006, pp 495-502.
[9] Lewis, A., "Extending the Range of Player-Performance Measures in One-Day Cricket," Journal of Operational Research Society, 59, 2008, pp 729-742.
[10] Van Staden, P., "Comparison of Cricketers' Bowling and Batting Performance using Graphical Displays," Current Science, 96, 2009, pp 764-766.
[11] Lakkaraju, P., and Sethi, S., "Correlating the Analysis of Opinionated Texts Using SAS® Text Analytics with Application of Sabermetrics to Cricket Statistics," Proceedings of SAS Global Forum 2012, 136-2012, pp 1-10.
[12] Lemmer, H., "A Measure for the Batting performance of Cricket Players," South African Journal for Research in Sport, Physical Education and Recreation, 26, 2004, pp 55-64.
[13] Lemmer, H., "An Analysis of Players' Performances in the First Cricket Twenty20 World Cup Series," South African Journal for Research in Sport, Physical Education and Recreation 30, 2008, pp 71-77.
[14] Lemmer, H., "The Single Match Approach to Strike Rate Adjustments in Batting Performance Measures in Cricket," Journal of Sports Science and Medicine, 10, 2012, pp 630-634.
[15] Saikia, Hemanta and BhattacharjeeDibojyoti, "A Bayesian Classification Model for Predicting the Performance of All-Rounders in the Indian Premier League, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=16220 60.
[16] C. Deep Prakash, C.Patvardhan and Sushobhit Singh, "A new Machine Learning based Deep Performance Index for Ranking IPL T20 Cricketers", International Journal of Computer Applications (0975 – 8887) Volume 137 – No.10, March 2016
[17] C. Deep Prakash, C.Patvardhan and Sushobhit Singh," A new Category based Deep Performance Index using Machine Learning for ranking IPL Cricketers", International Journal of Electronics, Electrical and Computational System IJEECS ISSN 2348-117X Volume 5, Issue 2 February 2016
[18] Leo Breiman. Random forests. Machine Learning, 45(1): 5–32, 2001