Decision Making in Sports using Evolutionary Algorithms.

Sourav Prasad

Indian Institute of Technology (Indian School of Mines) Dhanbad

Acknowledgements

I want to express my deepest gratitude to my supervisor Professor Millie Pant for giving me this opportunity to prepare this project and for her valuable guidance, and kind supervision which has helped shape this project to its present state.

I also acknowledge with a deep sense of gratitude the contribution of my guides Vivekananda Pandey and Shanu Verma for always being there to clarify all my doubts and for their encouraging guidance.

Contents

Introduction

About Cricket

Cricket is a bat-and-ball game played between two teams of eleven players. The batting side scores runs by striking the ball bowled at the wicket with the bat, while the bowling and fielding side tries to prevent this and dismiss each batter. Means of dismissal include being bowled, when the ball hits the stumps and dislodges the bails, and by the fielding side catching the ball after it is hit by the bat, but before it hits the ground. When ten batters have been dismissed, the innings ends and the teams swap roles.

About Indian Premier League

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India.Twenty20 cricket is a shortened format of cricket. In a Twenty20 game, the two teams have a single innings each, which is restricted to a maximum of 20 overs. Currently, with eight teams, each team plays each other twice in a home-and-away round-robin format in the league phase. At the conclusion of the league stage, the top four teams will qualify for the playoffs. The top two teams from the league phase will play against each other in the first Qualifying match, with the winner going straight to the IPL final and the loser getting another chance to qualify for the IPL final by playing the second Qualifying match. Meanwhile, the third and fourth place teams from league phase play against each other in an eliminator match and the winner from that match will play the loser from the first Qualifying match. The winner of the second Qualifying match will move onto the final to play the winner of the first Qualifying match in the IPL Final match, where the winner will be crowned the Indian Premier League champions.

Team Composition Rules of the Indian Premier League

Some of the team composition rules (as of 2020 season) are as follows:

1. The squad strength must be between 18 and 25 players, with a maximum of 8 overseas players.
2. The salary cap of the entire squad must not exceed ₹850 million (US$12 million)
3. Under-19 players cannot be picked unless they have previously played first-class or List A cricket.
4. A team can play a maximum of 4 overseas players in their playing eleven.

Motivation

In any sport, the selection of players for a team is fundamental for its subsequent performance. Factors influencing the selection include the rules of the particular game, the restrictions associated with the game as well the environment in which the game is played. All of this makes the process of selection of a team very complex as it is affected by multiple variables and in many cases marked by a great deal of subjectivity.

Summary

The objective of this study was to select the playing 11 players optimally for an Indian Premier League (IPL) cricket team from the squad of an IPL franchise. It was considered as a Multi-Objective Optimization Problem (MOOP) with objectives as maximising the batting, bowling and fielding performance of the team. The teams were formed subject to the restrictions of the game. The sample was composed of twenty players belonging to the IPL franchise Mumbai Indians. Three different multi-objective evolutionary algorithms viz. NSGA-II, GDE3 and SMPSO were used to construct the teams. To assess the quality of the teams formed by all the three algorithms, these were compared with the official selected teams for the year 2019. Further to account for the variations in the environmental conditions of the game, the formed teams were evaluated based on their performance in different stadiums. According to the results, it was found that the NSGA-II algorithm outperformed all the other algorithms.

This report starts with a review of similar research conducted in the Kanpur Genetic Algorithms Laboratory. Following which the difference between the mathematical formulation of the

previous study and the present study has been discussed. A small introduction to the different algorithms used has been included. Finally, we conclude with the results and the conclusions.

Review of Literature

A similar study was conducted at the Kanpur Genetic Algorithms Laboratory (KanGAL), Indian Institute of Technology, Kanpur in 2012. The goal of this study was the formation of a good and successful cricket team, using batting strength and bowling strength of the team as the major factors affecting its performance and reaching an optimum trade-off.

They proposed a novel gene representation scheme and a multi-objective approach using the NSGA-II algorithm to optimise the overall bowling and batting strength of a team with 11 players as variables. After the initial selection, a multi-criteria decision-making approach consisting of fielding performance and a number of other cricketing criteria was used to reach the final selection of the team.

Mathematical Formulation
In the previous study

In the game of cricket, player statistics have multiple parameters like number of matches played, number of runs, batting strike rate for batsmen, the number of wickets taken for ballers, number of catches taken for fielders etc. Hence it becomes very important to identify the parameters that can reliably indicate a player's performance.

The objective functions chosen were:

In the previous study

1. **Batting Performance:** A player's batting average is the total number of runs he has scored divided by the number of the times he has been out. To avoid preference to new players who have played less number of matches, only those batsmen were considered who have scored at least 300 runs in an international T20 match.

2. **Bowling Performance:** A player's bowling average was calculated as the total number of runs conceded by the baller divided by the number of wickets taken by the baller. So lower the bowling average better is the baller's performance. Again only those players who have taken at least 20 wickets were included as ballers.

3. **Fielding Performance:** A player's fielding performance was calculated as the total number of catches divided by the total number of innings played. It was mainly used as a further criterion for the final selection of the team.

The constraints considered were:

1. A captain should be included in a team.
2. There should be at least one wicketkeeper in a team.
3. No two players are identical in a team.
4. Not more than four international players can be included in a team.
5. The total cost of the players selected must be less than the total budget.
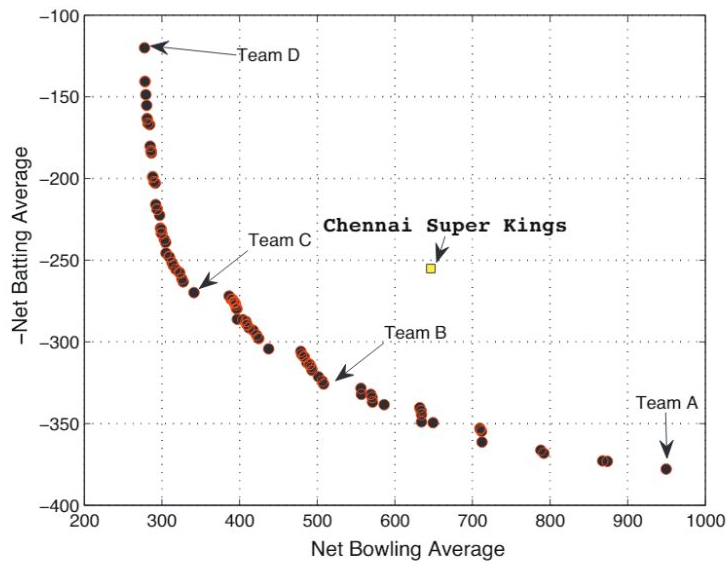
Results Obtained

Fig. 1. Multi-objective trade-off front obtained by NSGA-II is shown. CSK team is outperformed by Teams B and C (obtained in this study) on batting and bowling performances.

The simulation was run with constraints such as TotalBudget=6 million dollars. At least one wicketkeeper, one captain, and a maximum of four international players may be included in the team.

The standard parameter settings used were population size=400, crossover probability=0.9, maximum generations=400, mutation probability=0.05, SBX distribution rate=10 and mutation distribution index=20

To compare their results they used the team of Chennai Super Kings (CSK), the team that also won the finals. Using the same fitness functions the performance of the team was calculated and plotted in fig. 1 as a yellow square. The CSK team was outperformed by the teams found by the NSGA algorithm despite having a bigger budget of 7.5 million dollars as compared to the constraint of 6 million dollars.

Comparison of Results

As a part of the internship, the selection of players for IPL teams was initially performed using the same mathematical formulations as mentioned above. The sample was a dataset having the statistics of 129 IPL players up till the year of 2010. The results produced using the three algorithms viz. NSGA-II, GDE3 and SMPSO have been compared graphically.

Comparison Of Results
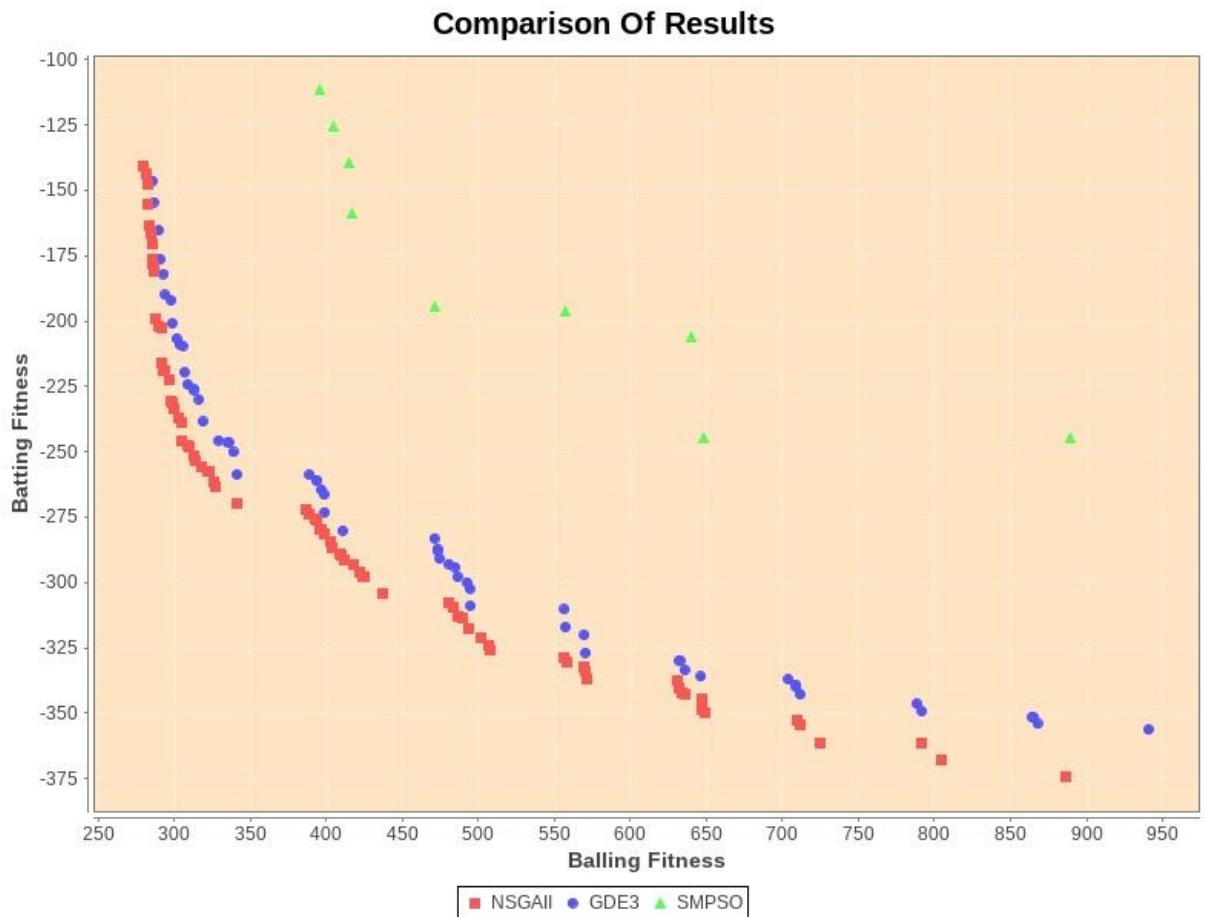
As determined by the graph:

1. **NSGA-II and GDE3 gave comparable results, and they were able to find solutions near the Pareto front.**

2. **SMPSO was not able to find solutions near the Pareto front.**

Report on the Present Investigation

Data Collection

The data has been aggregated for the years 2008 - 2018 for the players of the Mumbai Indians franchise. The data was aggregated using a python script attached in the appendix.

Sources

1. https://www.cricbuzz.com/cricket-scorecard-archives
2. https://cricsheet.org/

Players

Player Name	Role
Rohit Sharma	BAT
Surya Kumar Yadav	BAT
Evin Lewis	BAT
Quinton de Kock	WK
Ishan Kishan	WK
Aditya Tare	WK
Kieron Pollard	BAT/BALL
Ben Cutting	BAT/BALL
Kurnal Pandya	BAT/BALL
Hardik Pandya	BAT/BALL
Yuvraj Singh	BAT/BALL
Jayant Yadav	BALL
Lasith Malinga	BALL
Mayank Markande	BALL

Mitchell McClenaghan	**BALL**
Bauran Hendricks	**BALL**
Jasprit Bumrah	**BALL**
Rahul Chahar	**BALL**
Barinder Sran	**BALL**

Highlighted Players are the International Players

Stadiums

1. Bengaluru
2. Delhi
3. Kolkata
4. Mumbai
5. Chennai
6. Hyderabad
7. Kochi
8. Mohali
9. Indore
10. Visakhapatnam
11. Rajkot
12. Raipur
13. Ranchi
14. Jaipur
15. Ahmedabad
16. Cuttack
17. Nagpur
18. Dharamshala
19. Pune
20. Kanpur

Mathematical Formulation

In the present study, the problem was reformulated as selecting the playing eleven members from the complete squad of the franchise depending upon the stadium.

The parameters chosen in the present study were:

1. **Batting Performance:** A player's batting average was the total number of runs he has scored divided by the number of innings subtracted by the number of times he has been not-out.

2. **Bowling Performance:** A player's bowling average was calculated as the total number of runs conceded by the baller divided by the number of wickets taken by the baller. So lower the bowling average better is the baller's performance.

3. **Fielding Performance:** A player's fielding performance was calculated as the total number of catches summed with the total number of run-outs divided by the total number of innings played.

The constraints considered in the present study were:

1. There should be at least one wicketkeeper in a team.
2. There should be at least five ballers in the team.
3. No two players are identical in a team.
4. Not more than four international players can be included in a team.

Method

Representation of the Players

Initially, a database of all the players was created as mentioned in the Appendix a snippet of the final table has been included below. Each player was identified by a unique identification number.

id	Player Name	Matches	Runs	...
25	Rohit Sharma	188	4898	...
26	Yuvraj Singh	132	2750	...

Genotype   Phenotype

Representation of the Solution

A solution was generated by generating ten different random numbers, each of which mapped to a player in the database. The captain for all the teams was selected as Rohit Sharma, and hence he was always included in each solution. Thus each solution represented a selection of playing eleven members. There may be a problem of generating similar random numbers, This problem was solved by using a special data structure viz. HashSet. A HashSet data structure does not allow duplicates.

| 2 | 7 | 1 | 54 | 23 | 25 | 9 | 34 | 16 | 27 | 8 |

A Solution represented as a collection of Player IDs

Evaluation of the Solution

Each solution was evaluated based on the parameters mentioned in the mathematical formulation. **The validity of a solution was determined using constraint penalty values.** For each of the constraints, if the solution followed that particular constraint, then the penalty was 0 other-wise, a non-negative penalty value was assigned depending upon the extent of the violation.

Algorithms

The MOEA framework was used to implement the algorithms. In the following subsection, a general introduction to each of the algorithms has been provided.

Non-dominated Sorting Genetic Algorithm II

The steps for the NSGA-II algorithm are as described below:

1. Using an initial parent population, generate a child population using Genetic Operators.
2. Perform a non-dominated sorting in the combination of parent and offspring populations and classify them by fronts, i.e. they are sorted according to an ascending level of non-domination.
3. Fill new population according to the front-ranking
4. If one front is being included partially, perform Crowding-sort that uses crowding distance that is related to the density of solutions around each solution. The less-dense solutions are preferred for the diversity of solutions.
5. Create an offspring population from this new population using crowded tournament selection (It compares by front-ranking if equal then by crowding distance), crossover and mutation operators.

The standard parameters used for carrying out the simulations of NSGA-II were:

❖ **Simulated Binary Crossover**
   ➢ Rate = 1.0
   ➢ Distribution Index = 15.0
❖ **Polynomial Mutation**
   ➢ Rate = 1/N
   ➢ Distribution Index = 20.0

Generalised Differential Evolution 3

The steps for the GDE3 algorithm are as described below:

1. Initialisation of the population, mutation, recombination, replacement and evaluation
2. As is explained in the figure below, three other random vectors are chosen, and a new mutant vector is produced by using a mutation formula.

3. A new Trial vector is generated by the recombination of information from the mutant vector and the source vector.

4. The selection depends upon the comparison of the non-dominated ranks and the crowded distance of the trial and the source vector.
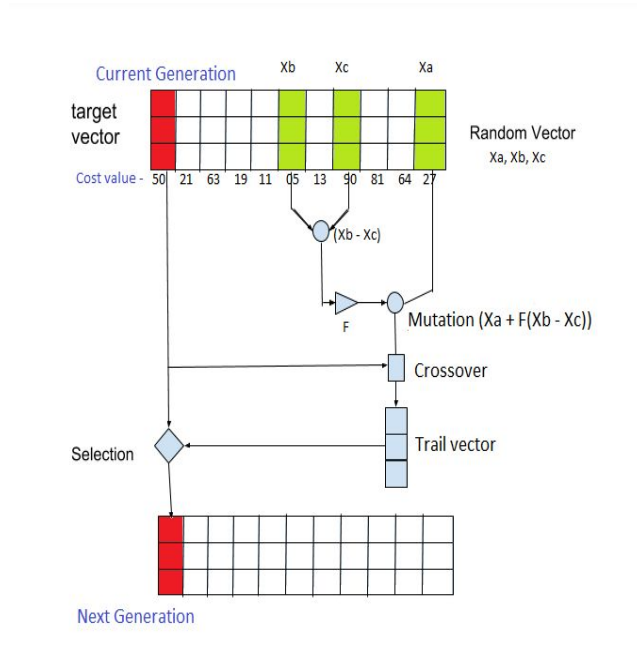
The standard parameters used for carrying out the simulations of GDE3 were:

❖ **Crossover Rate = 0.1**
❖ **Step Size = 15.0**

Speed Constrained Multi-Objective Particle Swarm Optimization

This Algorithm is based on Swarm Intelligence, Rather than using the mutation/crossover, it uses real-number randomness and global communication among the swarm particles for optimisation

The steps for the SMPSO algorithm are as described below:

1. It solves a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity.

2. The movement of each particle is influenced by its local best-known position but is also guided toward the best-known positions in the search-space, which are updated as better positions are found by other particles.

3. The determination of these particles with best positions is according to the Pareto dominance of the solutions. The non-dominated solutions are known as leaders, and these are stored in an archive to approximate the Pareto Front.

4. Finally, the swarm moves towards the best solutions.

The standard parameters used for carrying out the simulations of SMPSO were:

❖ **Archive Size = 100**
❖ **Polynomial Mutation**
  ➢ Rate = 1/N
  ➢ Distribution Index = 20.0

Results using Two Objectives

NSGA-II

The formation of teams was first carried out using only Two objectives i.e. Batting and the Bowling fitness of the teams. The following graphs show the results produced by the NSGA-II Algorithm using different numbers of maximum iterations as the stopping criteria.
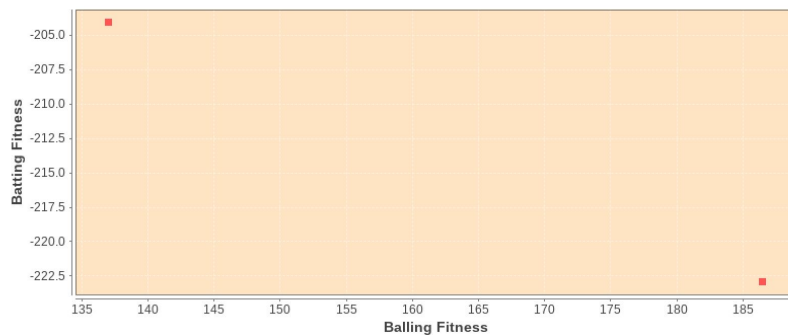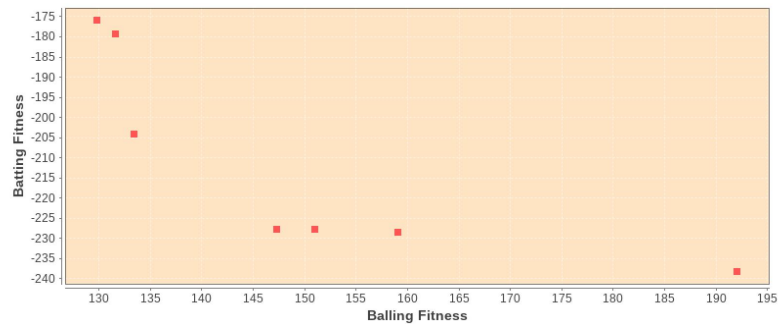

Two solutions were produced for 500 iterations of the NSGA-II algorithm
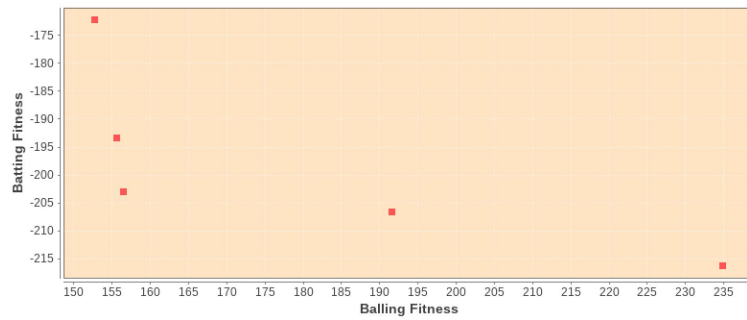
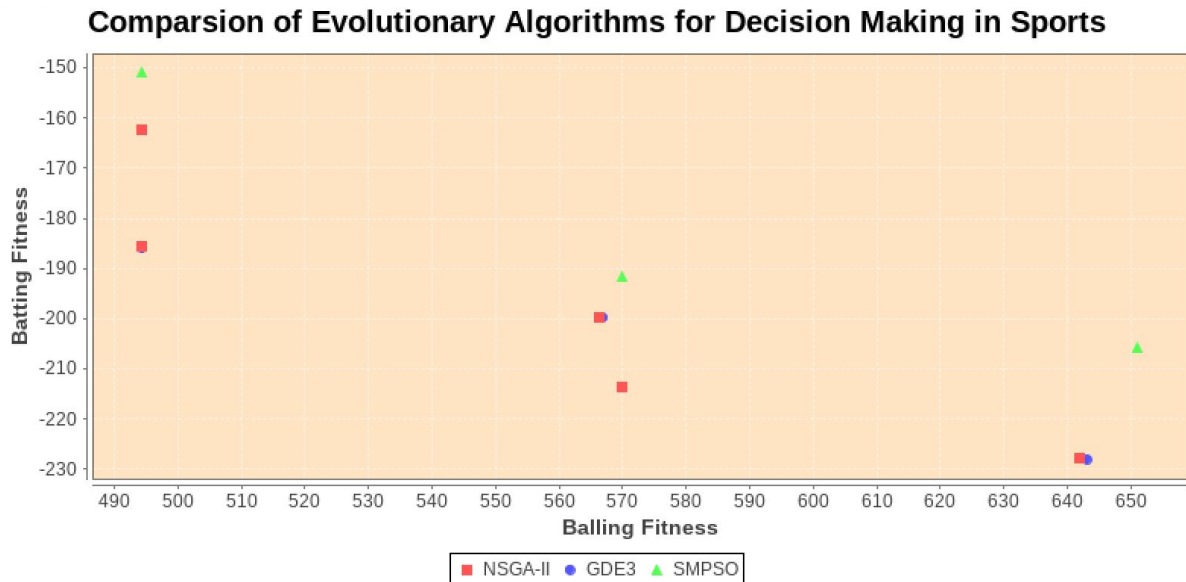*The solutions produced for 1000 iterations of the NSGA-II algorithm*
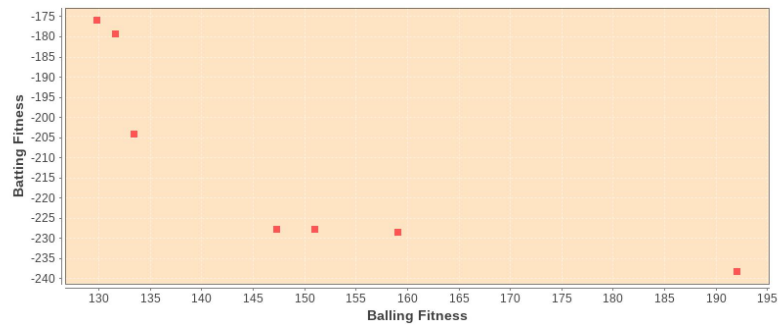
The solutions produced for 5000 iterations of the NSGA-II algorithm

The solutions produced for 10000 iterations of the NSGA-II algorithm. It was almost similar to the results of 5000 iterations

NSGA-II, GDE3 and SMPSO

**Comparsion of Evolutionary Algorithms for Decision Making in Sports**

However any team's strength cannot be reliably and adequately evaluated using only its batting and balling performance because many times the cricket games are on by teams which show excellent fielding performance. Hence further the performance of the teams was evaluated using all three objective functions.
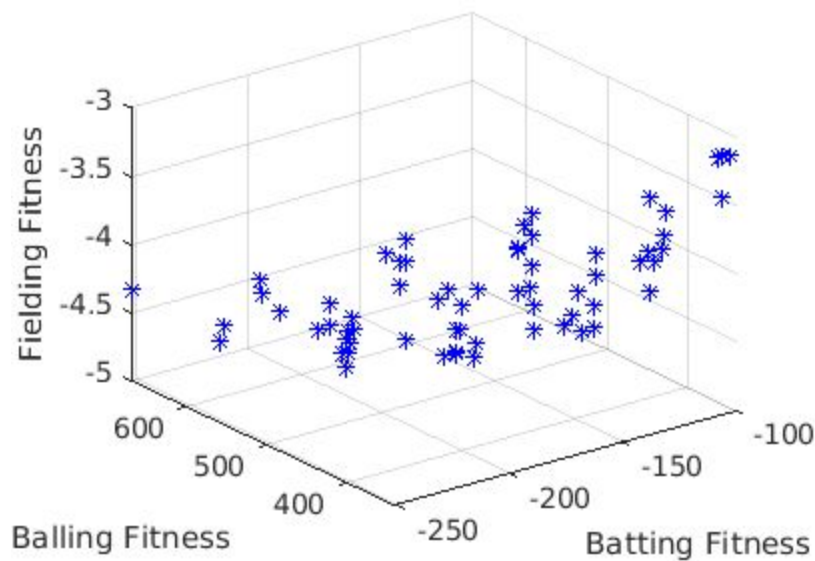
Results and Discussion

Graphical Comparison

The Pareto front generated by the NSGA-II Algorithm is attached below:
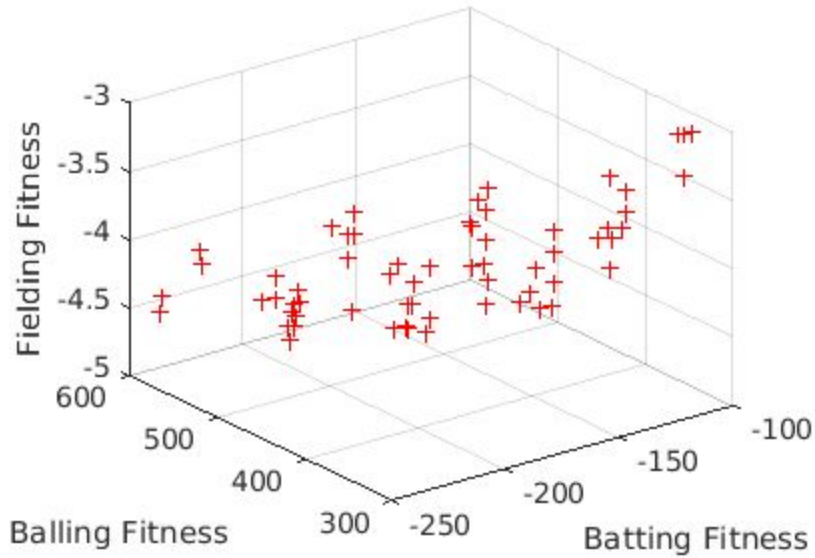
The Pareto front generated by the GDE3 Algorithm is attached below:

The Pareto front generated by the SMPSO Algorithm is attached below:
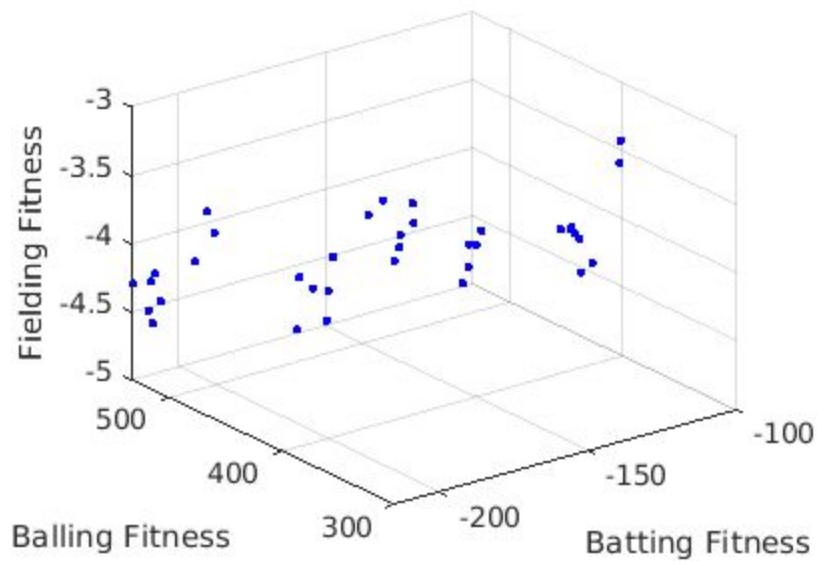
Graphical Comparison Graph of all the three algorithms
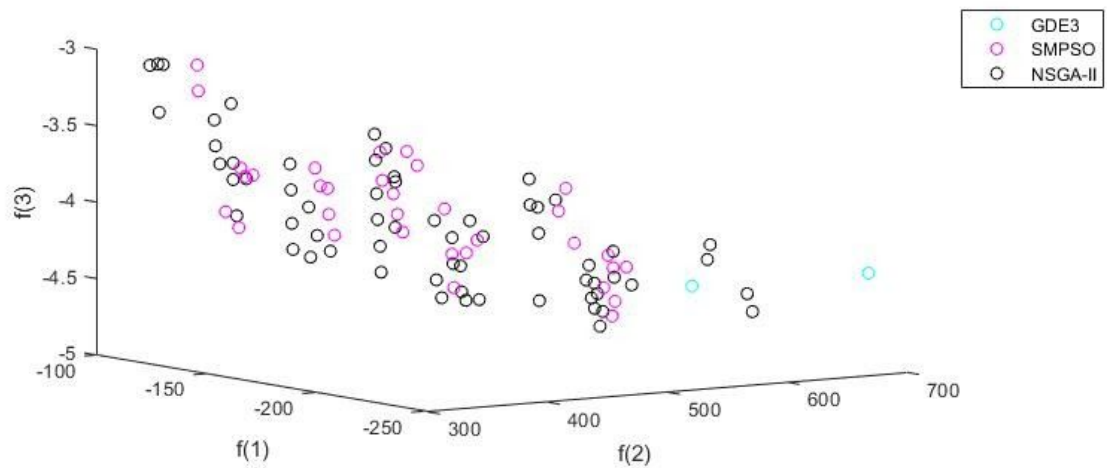
Analytical Comparison

In the following attached table, the frequency percentage of the players selected for the final playing 11 by each of the algorithms has been compared to the original percentage. **The percentages were calculated as follows:**

1. **Original:** The number of times that particular player has played in the playing 11 divided by the total number of matches played by the team

2. **Algorithm:** The number of times that particular player was selected by the particular algorithm divided by the total number of solutions produced by the algorithm

3. **Delta_algorithm:** The absolute difference between the original percentage and the algorithm percentage

Player Name	Original	NSGAII	GDE3	SMPSO	delta_NSGA	delta_GDE3	delta_SMPSO
Hardik Pandya	100	96.77	96.88	97.22	3.22580645	3.125	2.77777778
Rohit Sharma	93.75	100.00	100.00	100.00	6.25	6.25	6.25
Kieron Pollard	100	96.77	96.88	88.89	3.22580645	3.125	11.11111111
Surya Kumar Yadav	100	61.29	62.50	41.67	38.70967742	37.5	58.33333333
Kurnal Pandya	100	100.00	100.00	94.44	0	0	5.55555556
Ishan Kishan	43.75	53.23	54.69	63.89	9.47580645	10.9375	20.13888889
Yuvraj Singh	25	91.94	90.63	100.00	66.93548387	65.625	75
Lasith Malinga	75	66.13	64.06	72.22	8.87096774	10.9375	2.77777778
Ben Cutting	18.75	64.52	65.63	80.56	45.76612903	46.875	61.80555556
Jasprit Bumrah	100	40.32	39.06	30.56	59.67741935	60.9375	69.44444444
Quinton de Kock	100	59.68	60.94	38.89	40.32258065	39.0625	61.11111111
Evin Lewis	18.75	29.03	31.25	22.22	10.28225806	12.5	3.47222222
Rahul Chahar	81.25	66.13	65.63	47.22	15.12096774	15.625	34.02777778
Mitchell McClenaghan	31.25	20.97	20.31	27.78	10.28225806	10.9375	3.47222222
Bauran Hendricks	0	56.45	54.69	55.56	56.4516129	54.6875	55.55555556
Aditya Tare	0	4.84	6.25	19.44	4.838709677	6.25	19.44444444
Mayank Markande	18.75	77.42	76.56	52.78	58.66935484	57.8125	34.02777778
Barinder Sran	12.5	14.52	14.06	52.78	2.01612903	1.5625	40.27777778
Jayant Yadav	12.5	0.00	0.00	13.89	12.5	12.5	1.38888889
Alzarri Joseph	18.75	0.00	0.00	0.00	18.75	18.75	18.75
Behrendorff	31.25	0.00	0.00	0.00	31.25	31.25	31.25
Siddesh Lad	6.25	0.00	0.00	0.00	6.25	6.25	6.25
Rasikh Salam	6.25	0.00	0.00	0.00	6.25	6.25	6.25
Anukul Roy	6.25	0.00	0.00	0.00	6.25	6.25	6.25
Anmolpreet Singh	0	0.00	0.00	0.00	0	0	0
Pankaj Jaswal	0	0.00	0.00	0.00	0	0	0
Total Deviation					521.3709677	525	634.7222222

As is evident from the last row of the table, NSGA-II has the least absolute deviation.

This information presented in the form of a bar chart is attached below:

Original, NSGAII, GDE3 and SMPSO
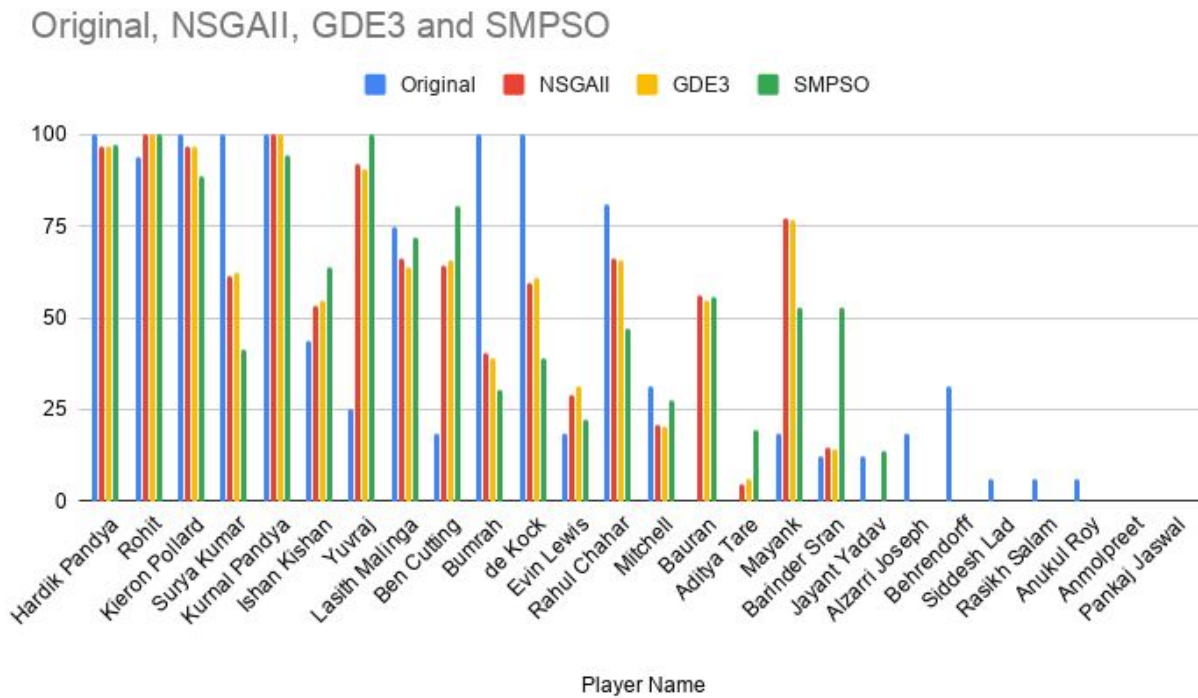
Future Work

The Selection of Teams for a particular stadium could not be completed in the course of the internship and hence it would be completed in the future.

Appendix

Stadium-Data   File:
https://docs.google.com/spreadsheets/d/1HasUeo4eDB2OrVQMcMB3qMApIngziyry3OR7FTr3L_M/edit?usp=sharing

Algorithm-Data-File:
https://docs.google.com/spreadsheets/d/1WPoV5pj0M28s0y0y5tQ8E0lpbOKyAe7Yc8_vuMljrz8/edit?usp=sharing

Python-Script

https://www.kaggle.com/souravpd/ipl-data-set-spark/

References

1. An explanation of cricket, http://www.cs.purdue.edu/homes/hosking/cricket/explanation.htm

2. https://en.wikipedia.org/wiki/Indian_Premier_League#Player_acquisition,_squad_composition_and_salaries

3. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-ii. IEEE Transactions on Evolutionary Computation 6(2), 182-197 (2002)

4. A. J. Nebro, J. J. Durillo, J. Garcia-Nieto, C. A. Coello Coello, F. Luna and E. Alba, "SMPSO: A new PSO-based metaheuristic for multi-objective optimization," 2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making(MCDM), Nashville, TN, 2009, pp. 66-73, doi: 10.1109/MCDM.2009.4938830.

5. S. Kukkonen and J. Lampinen, "GDE3: the third evolution step of generalized differential evolution," 2005 IEEE Congress on Evolutionary Computation, Edinburgh, Scotland, 2005, pp. 443-450 Vol.1, doi: 10.1109/CEC.2005.1554717.

6. Ahmed, Faez & Jindal, Abhilash & Deb, Kalyanmoy. (2011). Cricket Team Selection Using Evolutionary Multi-objective Optimization. 71-78. 10.1007/978-3-642-27242-4_9.

7. MOEA Framework, http://moeaframework.org/