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

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.

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, balling 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.

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 balling 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 balling 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

$$arg \max_{\mathbf{t}=\{c,w,p_1,p_2,\ldots,p_9\}} \begin{cases} f_1(t) = \sum_{i=c,w,p_1,\ldots,p_9} \text{batting performance}(\mathbf{i}), \\ f_2(t) = \sum_{i=c,p_1,\ldots,p_9} \text{bowling performance}(\mathbf{i}), \\ f_3(t) = \sum_{i=c,w,p_1,\ldots,p_9} \text{fielding performance}(\mathbf{i}). \end{cases}$$

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. **In this study the parameters chosen were:**

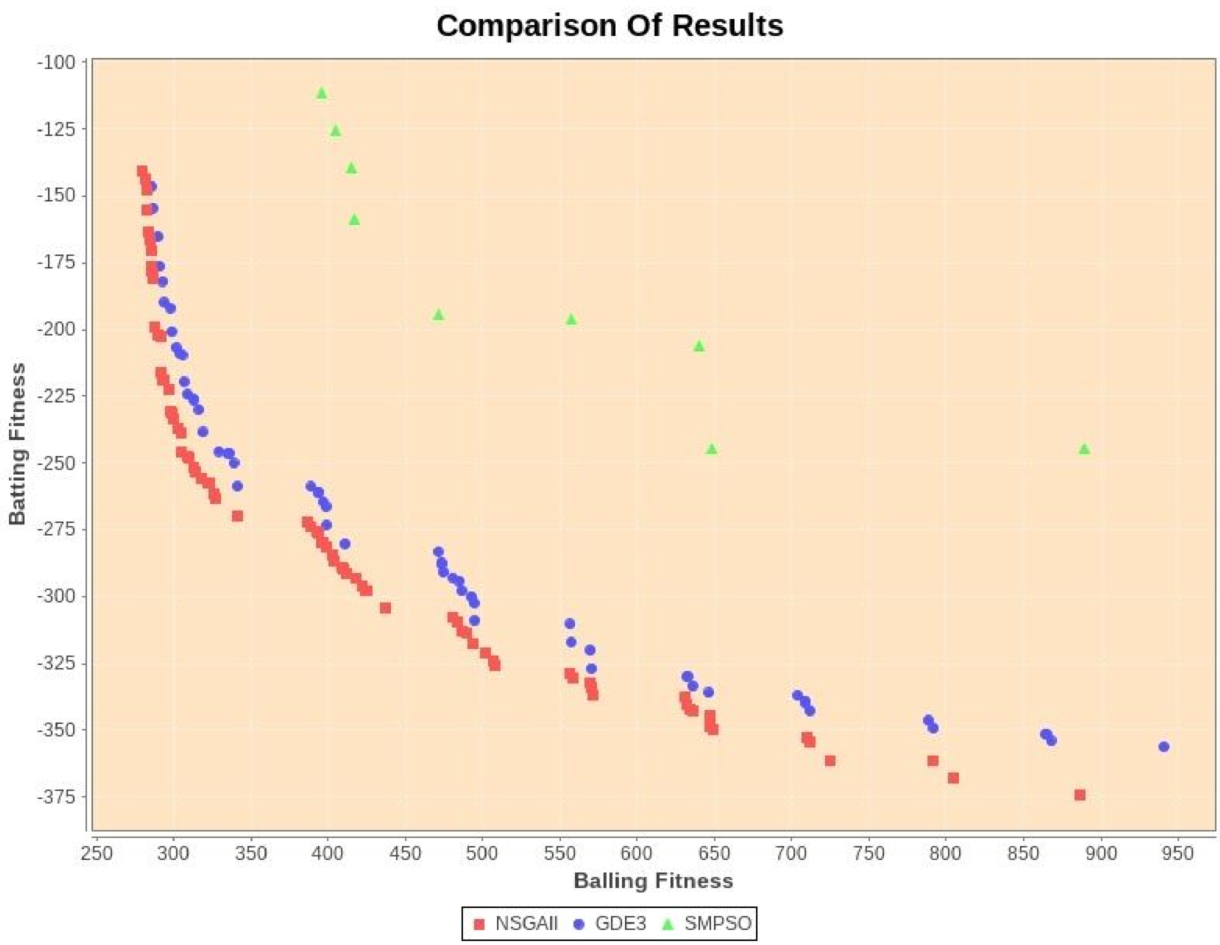
1. **Batting Performance:** A player's batting average is the total number of runs he has scored divided by the number of the times he has been out. To avoid preference to new players who have played less number of matches, only those batsmen were considered who have scored at least 300 runs in an international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

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.



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

Mathematical Formulation

In the present study, the problem was reformulated as selecting the playing eleven members from the complete squad of the franchise. This was done using the data collected for all the major Indian Stadiums for all the players of the Mumbai Indians 2019 Squad.

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. **Balling Performance:** A player's balling 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 balling 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..

Data Collection

The stadium wise performance data (2008-2018) was collected for each of the players of the Mumbai Indians Squad 2019. This data was compiled from various sources and has been attached as an excel sheet in the Appendix.

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.

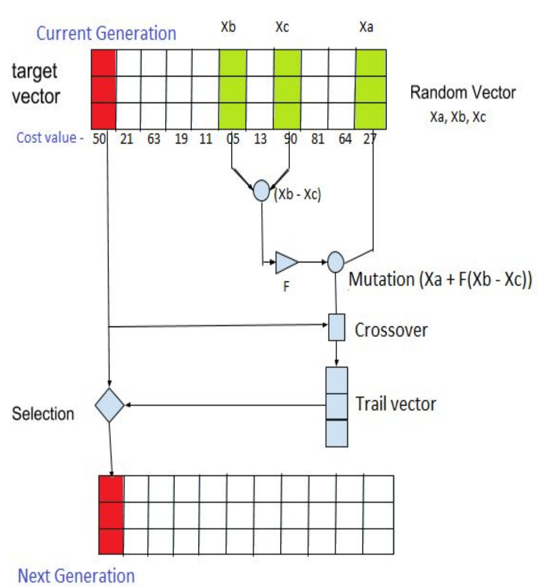
The standard parameters used for carrying out the simulations of NSGA-II were:

- **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.

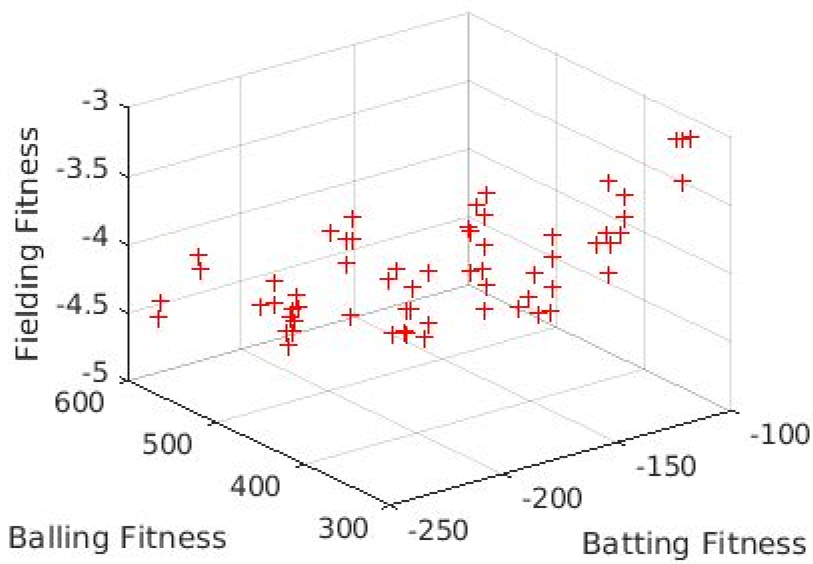
The standard parameters used for carrying out the simulations of SMPSO were:

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

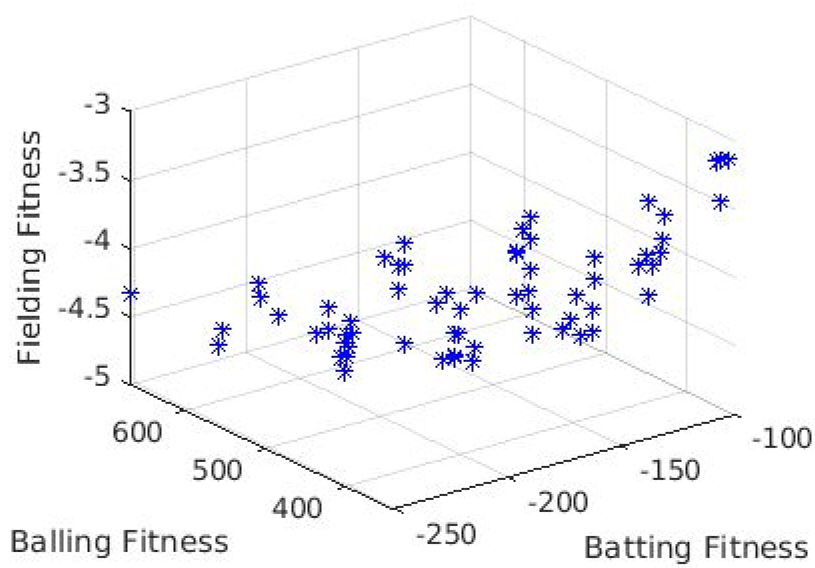
Results and Discussion

Graphical Comparison

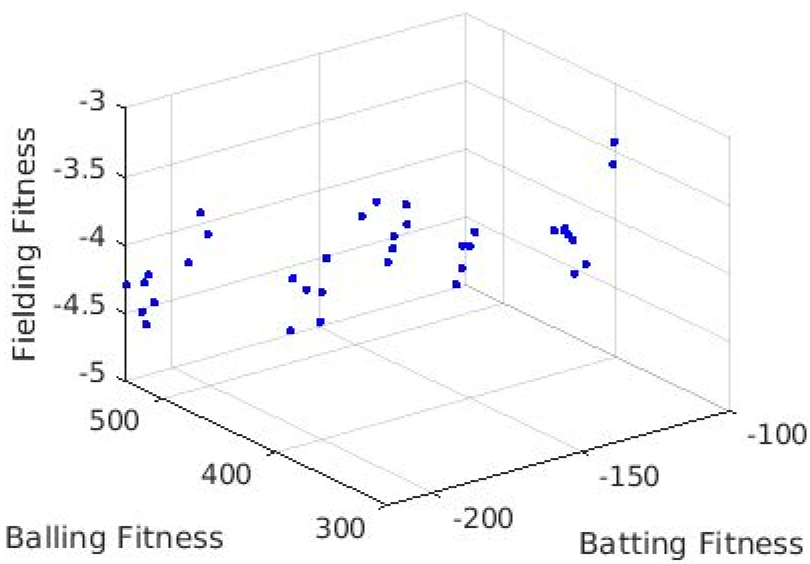
The Pareto front generated by the NSGA-II Algorithm is attached below:



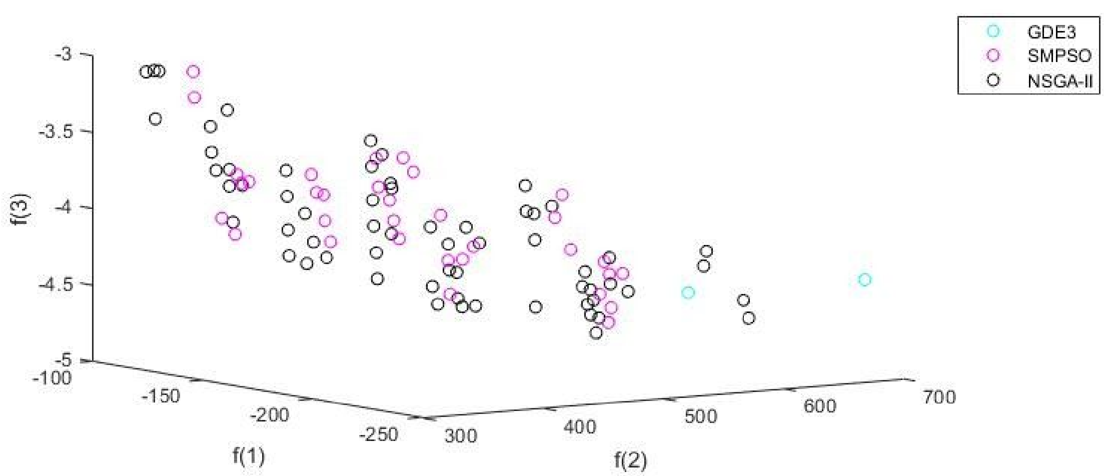
The Pareto front generated by the GDE3 Algorithm is attached below:



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.

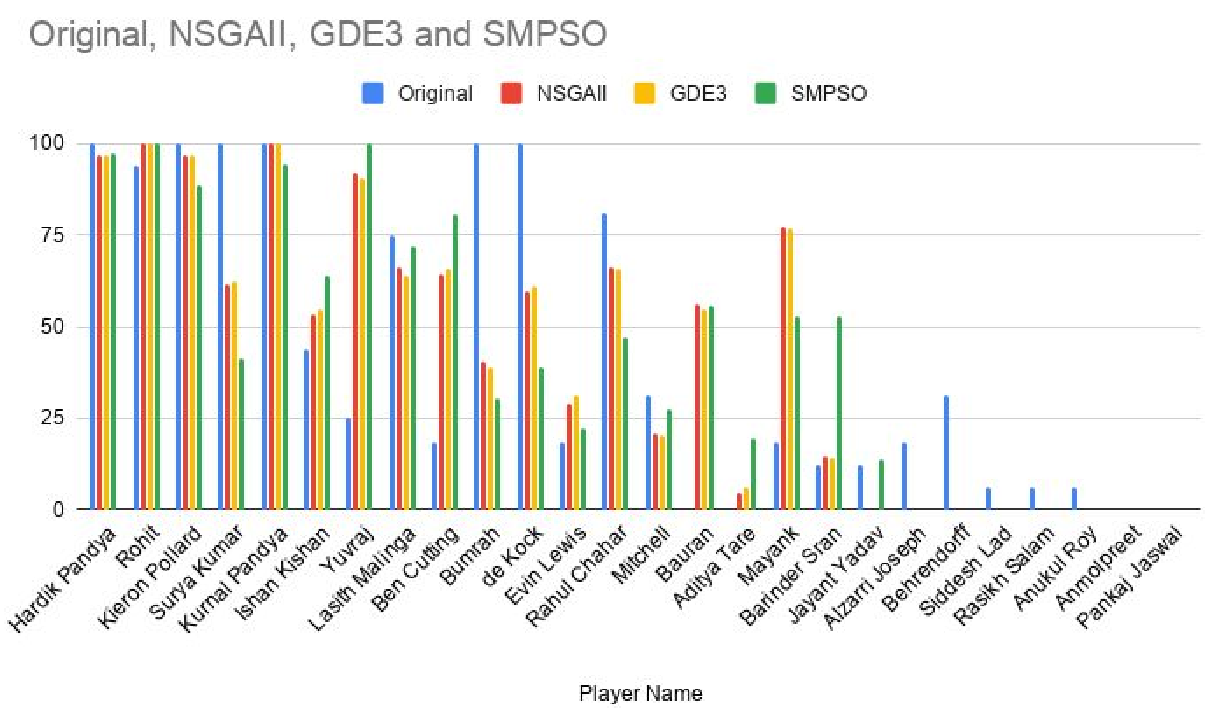
The percentages were calculated as follows:

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:



Selection of the Team for a Particular Stadium

Further, the produced solution teams were evaluated for their performance in three particular stadiums Mumbai, Mohali and Bengaluru. A generic method was developed that could be used to evaluate the fitness of the solutions in any city. **One of the Pareto Teams produced for the city of Mumbai is:**

1. Rohit Sharma	6. Ishan Kishan
2. Kieron Pollard	7. Ben Cutting
3. Surya Kumar Yadav	8. Yuvraj Singh
4. Mitchell McClenaghan	9. Lasith Malinga
5. Kurnal Pandya	10. Jasprit Bumrah
	11. Hardik Pandya

One of the Pareto Teams produced for the city of Mohali is:

1. Rohit Sharma 2. Rahul Chahar 3. Surya Kumar Yadav 4. Evin Lewis 5. Kurnal Pandya	6. Ishan Kishan 7. Quinton de Kock 8. Yuvraj Singh 9. Lasith Malinga 10. Jasprit Bumrah 11. Hardik Pandya

One of the Pareto Teams produced for the city of Bengaluru is:

1. Rohit Sharma 2. Kieron Pollard 3. Surya Kumar Yadav 4. Rahul Chahar 5. Kurnal Pandya	6. Ishan Kishan 7. Ben Cutting 8. Yuvraj Singh 9. Lasith Malinga 10. Mitchell McClenaghan 11. Hardik Pandya

The highlighted team members in the first table are the players that are common to all the three cities. Hence we can observe that depending on the stadium, different combinations of players may perform better.

Future Work

In the following study, lots of important factors such as pitch conditions, weather conditions, as well as the opponent team have been neglected. It may happen that certain players might be better in tackling a particular opponent. The future scope of this problem would be to include all these particular factors as well as factors to account for the uncertainty in the player performances.

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

References

1. An explanation of cricket, http://www.cs.purdue.edu/homes/hosking/cricket/explanation.htm
2. 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)
3. 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.
4. 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.
5. 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.
6. MOEA Framework,http://moeaframework.org/