ENHANCING NSGA-II FOR MULTI-OBJECTIVE COMBINATORIAL OPTIMIZATION PROBLEM: BALANCED SQUAD FOR INDIAN PREMIER LEAGUE

Project-I (MA57301) report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

5 Yrs Integrated MSc

in

Mathamtics and Computing

by

Chittireddy Akhil Reddy

(20MA20019)

Under the supervision of

Prof. Adrijit Goswami

Department of Mathamtics

Indian Institute of Technology Kharagpur

Autumn Semester, 2024-25

November 13, 2024

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: November 13, 2024 (Chittireddy Akhil Reddy)
Place: Kharagpur (20MA20019)

DEPARTMENT OF MATHAMTICS

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**ENHANCING NSGA-II FOR MULTI-OBJECTIVE COMBINATORIAL OPTIMIZATION PROBLEM: BALANCED SQUAD FOR INDIAN PREMIER LEAGUE** " submitted by **Chittireddy Akhil Reddy** (Roll No. 20MA20019) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of 5 Yrs Integrated MSc in Mathamtics and Computing is a record of bona fide work carried out by him under my supervision and guidance during Autumn Semester, 2024-25.

Date: November 13, 2024
Place: Kharagpur

Prof. Adrijit Goswami
Department of Mathamtics
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Chittireddy Akhil Reddy** Roll No: **20MA20019**

Degree for which submitted: **5 Yrs Integrated MSc**

Department: **Department of Mathamtics**

Thesis title: **ENHANCING NSGA-II FOR MULTI-OBJECTIVE COMBINATORIAL OPTIMIZATION PROBLEM: BALANCED SQUAD FOR INDIAN PREMIER LEAGUE**

Thesis supervisor: **Prof. Adrijit Goswami**

Month and year of thesis submission: **November 13, 2024**

The Indian Premier League (IPL) is one of the world's most competitive and financially significant cricket tournaments, where assembling a balanced team is crucial for optimizing performance within strict regulatory and budgetary constraints. This project addresses the challenge of selecting an optimal IPL squad by utilizing an enhanced version of the Non-dominated Sorting Genetic Algorithm(NSGA-II). Traditional squad selection often fails to consider the nuanced trade-offs between batting and bowling performance while adhering to IPL-specific constraints, such as salary cap, overseas player limits, and role requirements. To tackle this, we formulated a multi-objective optimization problem(MOOP) that seeks to maximize both batting and bowling performance while minimizing squad cost, thereby achieving a balanced and competitive team composition.

The proposed Enhanced NSGA-II incorporates additional modifications to the standard NSGA-II algorithm, specifically tailored to handle the unique constraints and requirements of IPL squad selection. By conducting a comprehensive sensitivity analysis on performance parameters, we identified optimal trade-off squads for varying team priorities, providing insights into the value of different player roles. Our results demonstrate that the Enhanced NSGA-II (BNSGA-II) consistently generates well-balanced squads with a high frequency of key players, indicating robustness in selection choices. Additionally, a comparative analysis with other NSGA-II variants, including MNSGA-II, RNSGA-II, and INSGA-II, confirms the superior performance of our approach in terms of hypervolume, Inverted Generational Distance (IGD), and Non-dominated Population Size (NPS) metrics. This study provides a data-driven solution for IPL squad optimization, enabling franchises to make informed, performance-based squad selection decisions.

Acknowledgements

The success of every project is primarily a result of the dedicated efforts of numerous individuals who consistently provided valuable advice and extended their helping hands. I genuinely want to convey my gratitude for the inspiration, support, and guidance offered by all those who played a crucial role in ensuring the project's success.

I want to express my heartfelt appreciation to my project mentor, Dr.Adrijit Goswami, from the Department of Mathematics at IIT Kharagpur, for his invaluable guidance and unwavering motivation. Without his support, achieving the current state of this project would have been extremely challenging. Collaborating with Dr. Adrijit Goswami has been both enriching and enjoyable, and I am also indebted to him for his insightful discussions on the subject matter and substantial assistance.

I would also like to express my appreciation to Prof. Buddhananda Banerjee and Dr. Bodhayan Roy, our faculty advisor, for their immense support and valuable guidance. My heartfelt thanks to my family members and my friends for their love and support.

Chittireddy Akhil Reddy
5 Yrs Integrated M.Sc
Department of Mathematics
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal

Contents

Chapter 1

Introduction

1.1 Introduction

The Indian Premier League (IPL) has evolved into one of the world's premier Twenty20 (T20) cricket leagues, garnering global viewership and commercial success since its inception in 2008. Unlike traditional cricket formats, IPL requires teams to balance a unique blend of skills to succeed in the shorter, more dynamic T20 format. The performance of an IPL squad hinges on achieving an effective balance of batting and bowling capabilities, which allows teams to adapt to various match situations. However, constructing an optimal IPL squad is a complex, multi-objective problem influenced by several constraints such as budget limits, overseas player restrictions, and positional requirements.

Each IPL franchise operates under a salary cap that limits the total spending on player contracts, requiring team managers to make strategic trade-offs between acquiring star players and building a balanced squad. Furthermore, IPL regulations impose restrictions on the number of overseas players in each squad and in the playing eleven, emphasizing the need to maximize the potential of both domestic and international players. Additionally, achieving a balanced team structure requires the inclusion of specialist batters, bowlers, all-rounders, and wicket-keepers in appropriate proportions.

Traditional methods for IPL squad selection, often based on simple performance metrics and expert intuition, do not account for the complex interactions between players' individual and collective contributions across multiple dimensions. To address this gap, optimization algorithms offer a systematic approach to selecting a balanced IPL squad that maximizes team performance under defined constraints.

This project focuses on applying an enhanced version of the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) to the problem of IPL squad selection. NSGA-II is a widely recognized multi-objective optimization algorithm known for its effectiveness in handling trade-offs between conflicting objectives. However, the standard NSGA-II algorithm alone does not fully accommodate the specific constraints and challenges inherent to IPL squad composition. Hence, we propose an Enhanced NSGA-II variant that incorporates custom constraints and adjustments tailored for IPL's regulatory framework and performance requirements.

The primary objectives of this project are to maximize the batting and bowling performance of the squad while minimizing the squad's total cost and ensuring adherence to IPL-specific regulations. Through this approach, we aim to generate optimized trade-off squads that provide IPL franchises with a practical, data-driven foundation for squad selection. By leveraging historical player data, performance metrics, and optimization techniques, this project seeks to assist decision-makers in assembling competitive and balanced IPL squads.

1.2 Motivation

Building a winning IPL squad requires navigating complex constraints and making strategic choices to balance player roles, performance metrics, and budgetary limits. Given the fierce competition and high stakes in the IPL, franchises are under pressure to make optimal squad selections that enhance team performance across various scenarios. However, selecting a squad that aligns with objectives such as maximizing batting and bowling strengths, while adhering to salary caps and overseas player limits, represents a challenging multi-objective optimization problem.

Traditional selection methods, which rely heavily on expert judgment and historical data, often miss potential synergies or fail to balance the needs for both specialists and all-rounders. By incorporating data-driven solutions, IPL franchises can optimize squad composition to increase their competitive edge. Advanced optimization techniques, like Enhanced NSGA-II, allow for a structured exploration of trade-offs, identifying squad configurations that best meet a team's objectives.

This project addresses the need for a systematic, data-backed approach to IPL squad selection, offering a model that generates balanced, high-performing squads. Through the use of Enhanced NSGA-II, this solution aims to assist IPL franchises in making informed, optimized decisions that effectively align their resources with competitive performance goals.

1.3 Related Work

The problem of selecting optimal cricket squads has been explored through various computational methods, particularly focusing on the Indian Premier League (IPL) due to its economic significance and competitive format. Early approaches to team selection leveraged Genetic Algorithms (GA), as demonstrated by Sathya and Jamal (2009), who utilized GA for selecting national cricket teams based on player statistics and expert input. Later, in IPL-specific contexts, Ahmed et al. (2013) introduced a bi-objective NSGA-II model, optimizing for both batting and bowling performances under constraints like budget, wicket-keepers, captains, and overseas player limits. Their results showed improvement over actual IPL teams, though they prioritized batting and bowling averages primarily without incorporating strike rate or economy rate measures critical in T20 formats.

Further advancements in team optimization included binary integer programming (BIP) models used for T20 World Cup teams and IPL squads, considering player roles such as openers, middle-order batsmen, and all-rounders. Bhattacharjee Saikia (2014) applied BIP to maximize team strengths across different roles, while Chand et al. (2018) used integer linear programming to optimize batting, bowling, fielding, and cost, also introducing a concept of "star power".

Despite these advances, previous studies often overlooked key T20-specific performance metrics, such as strike rate for batters and economy rate for bowlers, which are critical in the limited-overs format. Additionally, many models failed to ensure a balanced proportion of batters, bowlers, and all-rounders, leading to theoretically optimal but practically unbalanced squads.

To address these gaps, the present study introduces two NSGA-II variants (BNSGA-II and INSGA-II), leveraging enhanced performance factors like Batting Performance Factor (BPF) and Combined Bowling Rate (CBR) to better assess player capabilities in T20 settings. These variants also incorporate constraints on the minimum number of star players, ensuring balanced squad selection aligned with IPL regulations and practical team composition needs

1.4 Problem Statement

The goal of this project is to develop a model that aids in constructing balanced, high-performing squads for the Indian Premier League (IPL) using enhanced multi-objective optimization techniques. IPL team selection is constrained by strict regulations, including budget limits, overseas player restrictions, and role-based quotas for batters, bowlers, and all-rounders. Moreover, franchise owners often have preferences for key players, which adds another layer of complexity to the squad selection process.

This project, seeks to address these gaps by introducing a novel approach using enhanced NSGA-II variants. These variants aim to:

- **Maximize Net Batting and Bowling Performance:** Through performance metrics such as Batting Performance Factor (BPF) and Combined Bowling Rate (CBR), the model evaluates and selects players to maximize the overall batting and bowling strength of the team.
- **Incorporate Franchise Constraints and Preferences:** Adhering to the IPL's squad composition rules and budget limits, the model ensures a balanced mix of batters, bowlers, and all-rounders, while allowing for specific player preferences from franchise owners.
- **Provide Flexible Trade-Off Solutions:** The model generates a Pareto front of trade-off solutions, each representing an optimized squad with varying levels of batting and bowling strength. This allows franchise owners to select a squad that aligns with their strategic priorities.

The outcome of this project will be a multi-objective squad selection tool that not only addresses the constraints of the IPL format but also allows for adaptive, data-driven decision-making in constructing competitive teams.

Chapter 2

Background

2.1 IPL Squad Selection Requirements

The IPL squad selection process is structured around specific regulations to ensure balanced and competitive teams. For the 2024 IPL season, the selection guidelines include budgetary limits, player quotas, and role-based constraints that influence team-building strategies:

- **Salary Cap**: Each franchise has a salary cap of Rs.95 crore, enforcing financial discipline and encouraging efficient player selection.
- **Total Players**: Teams can have a minimum of 18 and a maximum of 25 players, ensuring adequate depth and flexibility while adhering to roster limits.
- **Overseas Players**: Franchises may include up to 8 overseas players in their squad, with only 4 eligible to play in any match, promoting a strong domestic talent base alongside international experience.

- **Captain and Wicket-Keeper**: Typically, each team designates one captain and one primary wicket-keeper, though multiple players capable of fulfilling these roles may be included as backup options.

Role-Based Constraints:

- **Batters and Batting All-Rounders (a1)**: Teams typically include 5-6 pure batters and 1-2 batting all-rounders, setting a lower bound of 6-8 players to strengthen the batting lineup.
- **Bowlers and Bowling All-Rounders (a2)**: Teams usually include 4-5 pure bowlers and 1-2 bowling all-rounders, with a minimum of 5-7 players in this category for a balanced bowling attack.
- **Pure Batters (m1) and Pure Bowlers (m2)**: Teams generally include a minimum of 5 pure batters and at least 4 pure bowlers to cover essential batting and bowling positions.
- **All-Rounders (m3)**: Teams aim for around 3-5 all-rounders who contribute to both batting and bowling, adding strategic flexibility.
- **Star Players (n1, n2, n3)**: Teams strive to include 1-2 star batters, 1-2 star bowlers, and at least 1 star all-rounder, who serve as key players within the squad. A player with high-performance records in IPL and T20I is considered a 'star player'

These requirements shape the squad composition, balancing financial and positional constraints to help each franchise build a competitive team. This project leverages optimization algorithms to construct IPL squads within these guidelines, prioritizing balanced teams that maximize overall batting and bowling strength while meeting regulatory limits.

2.2 Optimization Problems

Optimization problems are computational challenges that involve finding the best solution among the set of possible solutions. The primary goal is to either maximize or minimize a particular value, often referred to as an *objective function*, while adhering to a set of constraints. The key concepts related to optimization problems include:

- **Objective Function**: This function defines the things which need to be optimized, such as maximizing profits, minimizing cost, or minimizing travel time. It enables the possibility to compare the performance of all possible solutions for a certain optimization problem.
- **Decision Variables**: These are independent numerical variables that can be adjusted to optimize the objective function. In optimization problems, a decision vector $(x_1, x_2, \dots)$ represents two or more decision variables $\{x_1, x_2, \dots\}$. The collective set of all decision vectors is known as the decision space.
- **Constraints**: These are the conditions or limitations that restrict the feasible solutions of an optimization problem. The constraints can be equality constraints (e.g., resources must be fully utilized) or inequality constraints (e.g., budgets cannot be exceeded).

2.3 Multi-objective Optimization

MOOPs represent a complex category of optimization challenges where the goal is to optimize multiple conflicting objectives simultaneously.

In a MOOP with k objectives ($k \geq 2$) and a decision space D, an objective function f maps a set of decision variables to a k-dimensional vector representing the quality of a solution. Mathematically, the function f can be defined as per Eq. 1-1 and Eq. 1-2:

$$f : D \to \mathbb{R}^k$$

$$f(x) = (f_1(x), f_2(x), \ldots, f_k(x)), \quad \forall x \in D$$

Unlike traditional single-objective optimization problems, there is no single optimal solution to MOOP, as the objectives often compete with each other. MOOPs aim to find a set of trade-off solutions known as the Pareto optimal solutions.

Pareto optimal solutions, named after Vilfredo Pareto, are a fundamental concept in multi-objective optimization. They represent the set of non-dominated solutions in a MOOP, where no solution outperforms another across all objectives simultaneously. The objective vectors corresponding to these Pareto optimal solutions collectively form the Pareto front. Each point on the Pareto front corresponds to a solution, where improving one objective comes at the expense of another. Thus, Pareto optimal solutions capture a range of feasible solutions representing optimal trade-offs between competing objectives.

The difficulty of MOOPs lies in the challenge of balancing competing objectives, which amplifies the complexity of decision-making. Navigating this complex landscape of trade-offs poses a significant challenge for decision-makers, especially in fields such as engineering, finance, and logistics, where multiple and conflicting goals must be carefully balanced.

2.4 NSGA-II Algorithm

The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is a well-known multi-objective optimization algorithm widely used for solving problems with conflicting objectives. NSGA-II has been applied successfully in diverse domains due to its ability to produce a set of trade-off solutions, also known as Pareto-optimal solutions, in a single run. These solutions represent the best possible compromises among conflicting objectives without favoring any particular solution.

NSGA-II operates with a population-based approach, iteratively evolving a set of potential solutions to converge on the Pareto front. It employs three main features that enhance its performance and efficiency:

1. **Non-dominated Sorting:** NSGA-II organizes solutions into distinct fronts based on dominance. The first front consists of non-dominated solutions, where each solution is not worse than any other in all objectives. Subsequent fronts are formed by excluding previously selected solutions, ensuring that higher-ranked fronts contain the most optimal solutions available.

2. **Crowding Distance:** NSGA-II uses crowding distance as a diversity preservation mechanism to maintain a spread of solutions across the Pareto front. It calculates crowding distance for each solution based on its proximity to other solutions in the same front, helping avoid convergence to a single area of the Pareto front.

3. **Selection and Variation Operators:** The algorithm uses genetic operators, such as crossover and mutation, to evolve the population in each generation. The selection process considers both non-dominated sorting rank and crowding distance, favoring solutions that are in better-ranked fronts and have higher crowding distances, thus preserving diversity.

In the context of IPL squad selection, NSGA-II is used to balance two conflicting objectives: maximizing batting and bowling strengths. By defining these objectives and applying NSGA-II's evolutionary process, a set of balanced trade-off squads can be generated. These squads adhere to IPL regulations and constraints while maximizing team performance. The resulting Pareto front from NSGA-II represents diverse combinations of squads, each optimized differently in terms of batting and bowling strengths, providing a spectrum of solutions that allow for informed selection of squads based on specific strategic goals.

Chapter 3

Methodology

3.1 Data Collection

For this project, data on 244 IPL players' performance, roles, and auction prices were collected from various reliable sources. The data sources and the process for categorizing players based on performance are detailed below.

3.1.1 Data Sources

Auction Prices: The player auction prices for IPL 2024 were collected from NDTV Sports IPL 2024 Auction. This data included the player's base price, team, and designation as an overseas player if applicable.

Player Statistics: To gather players' T20I and IPL performance statistics, I used the ESPNcricinfo website, which contains comprehensive player data. This data includes:

- **Batting Metrics:** Matches, innings, runs, average (Bat_Ave), strike rate (Bat_SR), highest score (HS), and boundary counts (4s and 6s).
- **Bowling Metrics:** Wickets, bowling average (Bow_Ave), economy rate (Econ), strike rate (Bow_SR), best bowling figures (BBI and BBM), and count of 4-wicket and 5-wicket hauls.
- **Fielding Metrics:** Catches and stumpings.

3.1.2 Data Extraction Process

For each player, the following process was used to extract relevant statistics:

1. First, I located the player's profile on ESPNcricinfo by performing a search query on Google with the player's name and "ESPNcricinfo." This was automated using SerpAPI, where I selected the first link from each search result.
2. Once the profile URLs were collected, I automated data extraction for all relevant statistics for each player.

3.1.3 Player Categorization

To enhance squad selection, players were classified as "star players" based on thresholds in T20I and IPL performance:

- **Star Batter:** Players who scored at least 1,000 runs in T20I or 2,000 runs in IPL were categorized as star batters.
- **Star Bowler:** Players with a minimum of 30 wickets in T20I or 60 wickets in IPL qualified as star bowlers.

- **Star All-Rounder:** Players who scored at least 300 runs and took 15 wickets in T20I, or scored 600 runs and took 30 wickets in IPL, were identified as star all-rounders.

This classification was essential for ensuring that the selected squads included well-rounded players and met expertise-based constraints. By focusing on auction prices and comprehensive T20 performance data, this dataset supports the creation of balanced, competitive IPL squads. Below is the details of ipl 2024 auctioned players

Role	**Number of Players**
Pure Batter	49
Pure Bowler	93
Wicket-keeper	31
All-rounder	71
Batting all-rounder	35
Bowling all-rounder	36
Captain	10

3.2 Problem Formulation

The IPL squad selection problem can be formulated as a multi-objective optimization problem, where the objective is to maximize the overall team performance by selecting a set of players subject to budget and selection constraints. Each squad represents a "Knapsack" with Maximum salary cap (Knapsack Capacity) where each player represents an "item" with a "weight" (auction price) and "profit" (performance value). The two main objectives are:

Objectives

1. Batting Performance Factor (BPF)

The Batting Performance Factor (BPF) is a metric that evaluates a player's batting capability by combining their batting average and strike rate. The parameter α allows flexibility in the weighting of the two factors. The objective is to maximize the net batting performance by minimizing the negative sum of BPF values for the selected players:

$$\text{BPF} = (\text{Batting Strike Rate})^{\alpha} \times (\text{Batting Average})^{1-\alpha}$$

The objective function for batting performance is:

$$\min f_1 = -\sum_{i=1}^{n} x(i) \times \text{BPF}(i)$$

where $x(i)$ is a binary decision variable, defined as:

$$x(i) = \begin{cases} 1 & \text{if player } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

2. Combined Bowling Rate (CBR)

The Combined Bowling Rate (CBR) evaluates a player's bowling effectiveness, integrating their bowling average, strike rate, and economy rate. The objective is to maximize net bowling performance by minimizing the total CBR for the selected players:

$$\text{CBR} = \frac{1}{\frac{1}{\text{Bowling Average}} + \frac{1}{\text{Economy Rate}} + \frac{1}{\text{Strike Rate}}}$$

The objective function for bowling performance is:

$$\min f_2 = \sum_{i=1}^{n} x(i) \times \text{CBR}(i)$$

where $x(i)$ again indicates whether player i is selected.

Constraints

The basic IPL regulations impose the following constraints on the team selection:

- **Salary Cap:** Total cost of squad $\leq T$
- **Squad Size:** Number of players $= m$
- **Overseas Players:** Maximum 8 overseas players.
- **Captain and Wicketkeeper:** At least one captain and one wicketkeeper.
- **Expertise Bounds:**
  - Sum of batters and batting all-rounders $\geq a_1$
  - Sum of bowlers and bowling all-rounders $\geq a_2$
  - Number of pure batters $\geq m_1$
  - Number of pure bowlers $\geq m_2$
  - Number of all-rounders $\leq m_3$
  - Number of star pure batters $\geq n_1$

– Number of star pure bowlers $\geq n_2$

– Number of star all-rounders $\geq n_3$

Problem Complexity

The problem involves selecting an optimal combination of players from a large pool, with the potential combinations represented by C_n^m, where n is the total number of players, and m is the maximum squad size. This setup ensures that a balance of player roles and skills is achieved within the constraints, optimizing the squad for both batting and bowling capabilities.

3.3 Implementation of Enhanced NSGA-II

To address the complex requirements of IPL squad selection, we implemented two enhanced variants of the NSGA-II algorithm: **BNSGA-II** (Binary NSGA-II) and **INSGA-II** (Integer NSGA-II). These variants were tailored to handle multiple objectives and constraints while generating trade-off squads optimized for both batting and bowling strengths. The development of these algorithms aims to improve performance over traditional NSGA-II methods by focusing on efficient solution representation, population initialization, constraint handling, and repair mechanisms.

3.3.1 Chromosome Representation

The choice of chromosome representation is crucial as it impacts both the algorithm's efficiency and the quality of solutions. Both of these representations denote a squad with m number of players. We used two chromosome representations to suit different optimization requirements:

- **Binary Chromosome Representation:** In BNSGA-II, each chromosome is represented as a binary vector, where a gene value of '1' indicates the inclusion of a player in the squad. This representation ensures unique combinations and prevents the re-selection of players.

Player Tag	1	2	3	4	...	n
Chromosome	1	0	0	1	...	0

- **Integer Chromosome Representation:** In INSGA-II, chromosomes are represented as integer vectors, where each gene directly refers to a specific player in the pool. This representation allows for more flexible operations while ensuring that the squad is composed of unique players.

Player Tag	1	2	3	4	...	m
Chromosome	15	14	6	21	...	3

3.3.2 Population Initialization

For effective optimization, both algorithms require an initial population that adheres to squad constraints:

- **BNSGA-II:** The population is initialized as binary chromosomes, each containing exactly m '1's to represent the squad size.
- **INSGA-II:** The population comprises integer chromosomes, where each chromosome has unique player indices, ensuring no duplicates within a squad.

3.3.3 Crossover and Mutation Operators

To explore a diverse set of solutions, the following crossover and mutation operators were implemented for each variant:

- **BNSGA-II:** The algorithm employs two-point crossover and swap mutation. The two-point crossover exchanges genes between two points randomly selected on the parent chromosomes, while swap mutation exchanges two random genes within a chromosome to introduce variability.
- **INSGA-II:** Partially mapped crossover (PMX) and random mutation were applied to maintain unique player selection in each squad. PMX ensures that offspring inherit unique players from both parents, while random mutation replaces a player with one not currently in the squad.

3.3.4 Constraint Handling and Repair Mechanism

Constraint satisfaction in BNSGA-II and INSGA-II is critical, given the limitations on the number of players, budget, and roles. We implemented a repair mechanism, to ensure that each solution in the population remains feasible:

Repair Mechanism: After applying crossover and mutation, the repair mechanism adjusts chromosomes to meet squad size and role constraints. For BNSGA-II, the repair mechanism maintains a fixed number of players m in each squad across generations by adjusting genes if the number of selected players deviates from m. This repair process minimizes infeasibility and enhances population diversity by preserving feasible solutions. The pseudocode for the repair mechanism is given in Algorithm 3-1. The other procedures, including the non-dominating sorting, crowding distance calculation, and selection procedures, are the same as in conventional NSGA-II. The algorithmic steps of the BNSGA-II are shown in Algorithm 3-2

Algorithm 3-1 Repair Mechanism

Parameters: Mutated chromosome (x), Number of variables (n), Squad length (m)

1: Find the number of non-zero elements in x and their indices set: N and Index

2: **IF** $N > m$, then

  Randomly select m indices from the set Index: Index1

  $y(\text{Index1}) = 1$

ELSE IF $N < m$, then

  Take the complement of the set Index: Index2

  Randomly select $m - N$ indices from Index2: Index3

  $x(\text{Index3}) = 1$

END IF

Algorithm 3-2 BNSGA-II

Algorithm parameters: Population size (pop_size), Number of variables (n), Maximum number of generations (max_gen), Mutation probability (p_m), Crossover probability (p_c)
Problem parameters: chromosome (x), Squad length (m)

% Population initialization

1: **FOR** $i = 1 : pop\_size$

  1.1: Generate a matrix $[A]_{1\times m}$ with unique elements $a_i \in \{1, 2, \ldots, n\}$

  1.2: **FOR** $j = 1 : n$

    IF $j \in A$, then

      $x_i(j) = 1$

    ELSE

      $x_i(j) = 0$

    END IF

  END FOR

END FOR

% **Generating next generations** P_t**,** $t = 2, 3, \dots, max\_gen$

2: **FOR** $t = 1 : max\_gen$

2.1: Perform two-point crossover on P_t using p_c

2.2: Perform swap mutation and generate the offspring population

2.3: Perform the Repair mechanism on the offspring population using Algorithm 3-1

2.4: Compare parent population and repaired offspring population

2.5: Put the combined population into different non-dominated fronts $F = \{F_1, F_2, \dots\}$ using non-dominated sorting procedure

2.6: Set the parent population P_{t+1}, φ, $k = 1$

2.7: **WHILE** $(P_{t+1} \leq pop\_size)$ **DO**

Include the k^{th} non-dominated front F_k in the P_{t+1} $(P_{t+1} = P_{t+1} \cup F_k)$

ELSEIF $(P_{t+1} \cup F_k > pop\_size)$ then

Calculate the crowding distance of members of the front F_k

Sort the members of F_k in decreasing order of crowded distance

Add the top $pop\_size - |P_{t+1}|$ members of F_k to P_{t+1}

END IF

END WHILE

END FOR

3: **Output:** A set of Pareto optimal solutions F_k

3.3.5 Summary of Enhanced NSGA-II Variants

The table below summarizes the features of BNSGA-II and INSGA-II, highlighting differences in chromosome representation, crossover, mutation, and constraint handling:

Feature	**BNSGA-II**	**INSGA-II**
Chromosome Type	Binary	Integer
Crossover	Two-point	Partially Mapped (PMX)
Mutation	Swap Mutation	Random Mutation
Constraint Handling	Constraint dominance + Repair Mechanism	Constraint Dominance

Table 3.3: Summary of BNSGA-II and INSGA-II Features

Both BNSGA-II and INSGA-II enhance NSGA-II by ensuring feasible and optimized squads. By balancing objectives and maintaining IPL-specific constraints, these variants offer franchise owners a flexible and effective method for building well-rounded squads.

3.4 Algorithm Comparison

The experimental design involves two distinct cases for the IPL squad selection problem. Each case introduces unique constraints aimed at achieving a balanced squad based on various player expertise requirements.

3.4.1 Case 1: Expertise Bounds

In this case, the following expertise bounds are applied:

- Sum of batters and batting all-rounders $\geq a_1$
- Sum of bowlers and bowling all-rounders $\geq a_2$

These constraints allow for flexibility in player roles but ensure a sufficient number of batters, bowlers, and all-rounders within the squad.

3.4.2 Case 2: Specific Expertise Bounds

In this case, the expertise bounds are defined more specifically:

- Number of pure batters $\geq m_1$
- Number of pure bowlers $\geq m_2$
- Number of all-rounders $\leq m_3$
- Number of star pure batters $\geq n_1$
- Number of star pure bowlers $\geq n_2$
- Number of star all-rounders $\geq n_3$

These constraints aim to maintain a well-rounded balance of players by defining precise lower bounds on pure batters and bowlers, limiting the number of all-rounders, and including star players in each expertise category.

3.4.3 Experiments

The comparison includes two main experiments:

- The first experiment evaluates the efficiency and effectiveness of the proposed algorithms, BNSGA-II and INSGA-II, in optimizing the squad selection problem for IPL.
- The second experiment analyzes the impact of the constraints in Cases 1 and 2 on the squad composition, specifically examining how these constraints affect the overall balance of player roles.

3.4.4 Benchmark Algorithms for Comparison

To validate the efficiency of BNSGA-II and INSGA-II, the study compares them with two prominent algorithms that have been adapted for the IPL team selection problem:

- **MNSGA-II**: This modified NSGA-II algorithm has been adapted for the bi-objective IPL team selection problem and uses a chromosome representation with fixed positions for the captain and wicketkeeper. Originally developed for an 11-player team, MNSGA-II has been extended here to a 23-player squad.
- **RNSGA-II**: This variant of real-coded NSGA-II, known as RNSGA-II, is also applied to the IPL squad selection. Along with MNSGA-II, it uses a continuous chromosome representation that is converted to binary using a rounding function. The rounding function is applied as follows:

$$b_i = \begin{cases} 1 & \text{if } x_i \leq \frac{n}{m} \\ 0 & \text{otherwise} \end{cases}$$

  where x_i represents the value of the continuous chromosome and b_i is the corresponding binary value, helping select m players from n in alignment with squad constraints. Algorithm 3-1 is used to repair any infeasible chromosomes generated by these algorithms.

The experiments compare these four algorithms—BNSGA-II, INSGA-II, MNSGA-II, and RNSGA-II—based on several performance metrics, including:

- **Hypervolume (HV)**: This metric evaluates the extent of the Pareto front produced by the algorithms.

- **Inverse Generational Distance (IGD)**: Measures the average distance between the Pareto front obtained and the true Pareto front.
- **Number of Pareto Solutions (NPS)**: The total number of non-dominated solutions generated, indicating the diversity of the solution set.
- **Computation Time**: The time taken by each algorithm to reach its final solution, indicating computational efficiency.

3.4.5 Parameter Settings for Experiments

The parameters for the proposed model and algorithms are summarized in tables below. The optimal values for these parameters were determined through fine-tuning, with several configurations tested to ensure comprehensive coverage of the Pareto front while achieving good convergence. For the experimental runs, a population size of 200 and a generation limit of 500 were selected, with fine-tuning also applied to crossover and mutation probabilities.

The upcoming sections detail the experimental results, providing insights into the comparative performance of BNSGA-II and INSGA-II alongside the benchmark algorithms.

Algorithm	Crossover Probability	Mutation Probability	Other Parameters
RNSGA-II	1	0.1	-
BNSGA-II	1	0.1	-
INSGA-II	1	0.1	-
MNSGA-II	0.9	0.05	Crossover rate = 0.1, SBX distribution index = 10, Mutation distribution index = 20

TABLE 3.4: Algorithm Parameters

Chapter 4

Results and Discussion

4.1 Algorithm Performance Comparison

This section presents a detailed comparison of the four algorithms—BNSGA-II, INSGA-II, MNSGA-II, and RNSGA-II—based on hypervolume, IGD (Inverse Generational Distance), NPS (Number of Pareto Solutions), and computational time for both Case I and Case II. Each performance metric provides insights into different aspects of the algorithm's effectiveness in optimizing the IPL squad selection problem under given constraints.

4.1.1 Hypervolume Analysis

Hypervolume is a critical measure to evaluate the performance of the algorithms along the Pareto front. In Case I, BNSGA-II achieves the highest average hypervolume (0.9427), indicating a superior spread and proximity to the Pareto front, followed by INSGA-II (0.9252) and RNSGA-II (0.9120). MNSGA-II shows a lower hypervolume (0.8039), suggesting a less effective coverage of the front.

In Case II, BNSGA-II maintains a higher hypervolume average (0.7101), showing an effective balance in solution diversity under stricter conditions. INSGA-II follows with a close hypervolume value of 0.7023, while MNSGA-II and RNSGA-II fall behind, reflecting their struggle under these constraints.

4.1.2 Inverse Generational Distance (IGD)

The IGD metric measures the convergence of solutions to the true Pareto front. In Case I, BNSGA-II leads with the lowest IGD average (0.0043), indicating strong convergence, closely followed by RNSGA-II (0.0171). MNSGA-II exhibits a higher IGD of 0.0515, which suggests more distance from the Pareto front.

In Case II, BNSGA-II also shows a low IGD (0.0192), indicating efficient convergence, followed by INSGA-II (0.0299). The increase in IGD values across algorithms in Case II reflects the complexity introduced by the restrictive constraints.

4.1.3 Number of Pareto Solutions (NPS)

The NPS metric measures the diversity of solutions. In Case I, INSGA-II achieves the highest NPS average (100.0), highlighting its effectiveness in generating diverse solutions, followed by BNSGA-II (95.2). MNSGA-II has the lowest NPS (67.4), indicating limited diversity.

In Case II, BNSGA-II shows a notable reduction in NPS (56.4), consistent with its focus on meeting strict constraints. MNSGA-II and RNSGA-II exhibit lower NPS values (28.0 and 34.8, respectively), indicating a more limited solution diversity under the tighter conditions.

4.1.4 Computation Time

Computation time is essential to evaluate each algorithm's efficiency. MNSGA-II achieves the lowest average computation time in both cases (126.33s in Case I and 127.97s in Case II), making it the most computationally efficient. BNSGA-II and INSGA-II have moderate computation times, with RNSGA-II taking the longest in both cases due to its resource-intensive operations.

In summary, BNSGA-II demonstrates strong performance across hypervolume and IGD metrics in Case I and performs robustly in Case II despite stricter constraints. INSGA-II and MNSGA-II exhibit diversity and efficiency benefits, with MNSGA-II achieving the fastest computation times.

4.2 Distribution of Player Roles in Trade-Off Squads

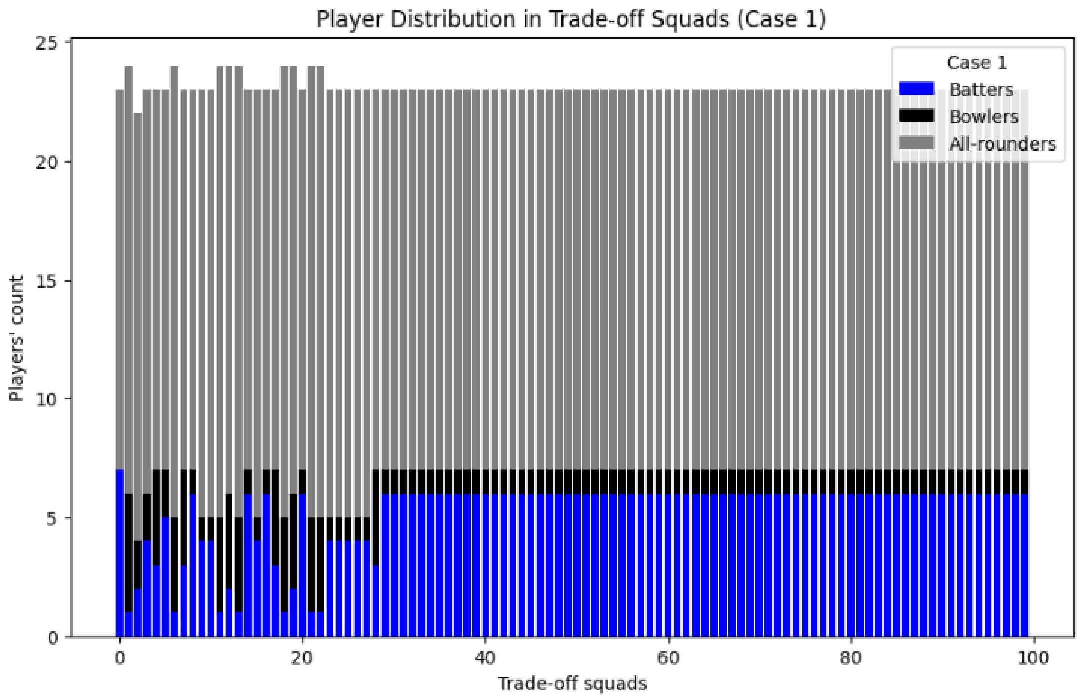
In a balanced IPL squad, it is essential to have an appropriate mix of pure batters, pure bowlers, and all-rounders to ensure both batting depth and bowling variety. However, the player distributions observed in Cases I and II reveal distinct differences in how each approach meets this requirement, highlighting the importance of tailored constraints for creating a functional squad.

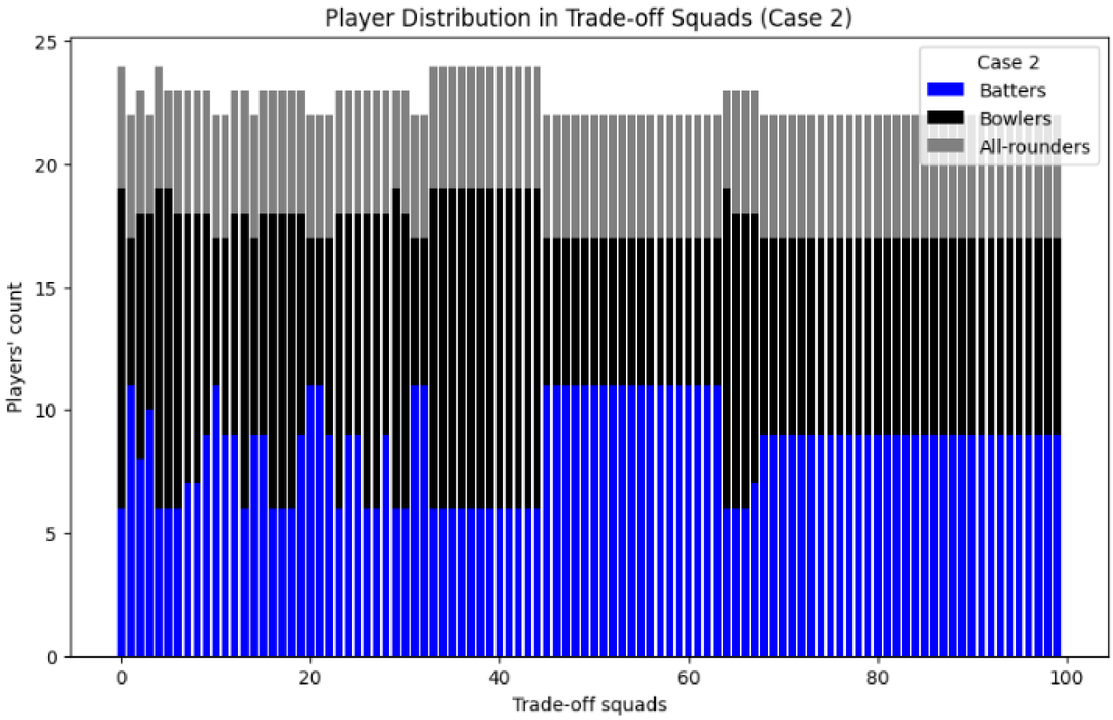
4.2.1 Role Distribution in Case 1

In Case I, the constraints primarily focus on the combined totals of batters and batting all-rounders, as well as bowlers and bowling all-rounders. This approach has led to a strong preference for all-rounders, who can contribute to both batting and bowling. Consequently, over 60% of the players in most Case I squads are all-rounders. While all-rounders provide dual utility, an over-reliance on them results in a lack of specialized pure batters and bowlers, potentially compromising the squad's focus on core skills. Some squads formed in Case I even lack dedicated bowlers, which is unrealistic for a competitive IPL squad where dedicated specialists are crucial for high-stakes matches.

4.2.2 Role Distribution in Case 2

In contrast, Case II imposes direct lower and upper bounds on each category of expertise, ensuring that the final squads maintain a suitable number of pure batters, bowlers, and all-rounders. This balance is visually evident in the distribution of players across roles in the generated squads. With specific minimums for pure batters and pure bowlers, Case II squads exhibit a more balanced structure, aligning better with typical IPL team compositions and avoiding an excessive tilt toward any single role. Each squad in Case II includes a reasonable proportion of pure bowlers, making it more viable for actual gameplay scenarios.





4.3 Trade-Off Squads with Preferred Players

Using BNSGA-II under Case-II conditions, two optimized trade-off squads were generated with preferred players Rohit Sharma and Jasprit Bumrah. These squads balance batting and bowling strengths while varying in cost.

Highest-Priced Squad:

The highest-priced squad, totaling Rs.84.0 crore, features a balanced mix of batters, bowlers, and all-rounders. Key players include Prithvi Shaw, Moeen Ali, and Sunil Narine, along with emerging talents like Dhruv Jurel and Abdul Samad. This squad achieves a batting strength of -8.88 and a bowling strength of 901.14, focusing on a solid batting lineup with reliable bowling support, especially with the inclusion of Bumrah and Ngidi.

Lowest-Priced Squad:

The lowest-priced squad, costing Rs.80.1 crore, emphasizes a strong bowling unit, achieving a higher bowling strength of 1101.01 compared to the highest-priced squad. It includes additional bowlers like Chris Woakes and Mukesh Kumar to reinforce its bowling capabilities, while maintaining a batting strength of -9.09 with dependable players such as Rilee Rossouw and Kyle Mayers. This lineup offers a more economical option while still fulfilling the dual objective of batting and bowling balance.

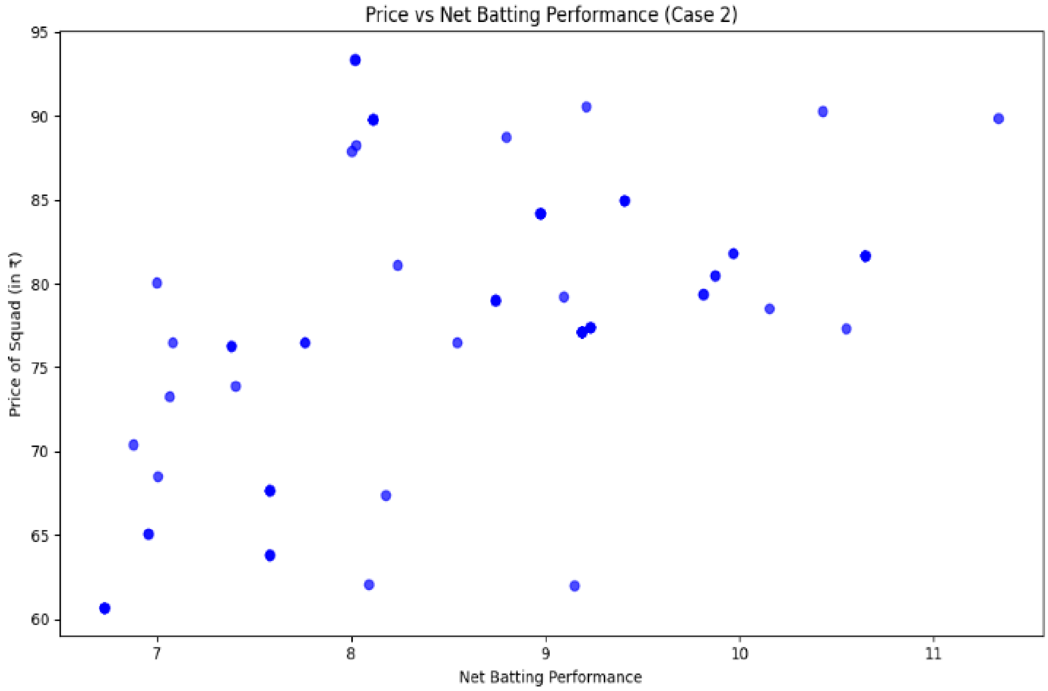
Both squads provide strategic trade-offs between batting and bowling, allowing flexibility in team composition while meeting budget constraints.

4.4 Price vs Performance and Star Players Analysis

The following analysis examines the relationship between squad cost and key performance indicators, as well as the impact of including star players in the lineup. The graphs illustrate trends in BNSGA-II results for Case II, highlighting the trade-offs between squad price, batting and bowling performance, and the number of star players.

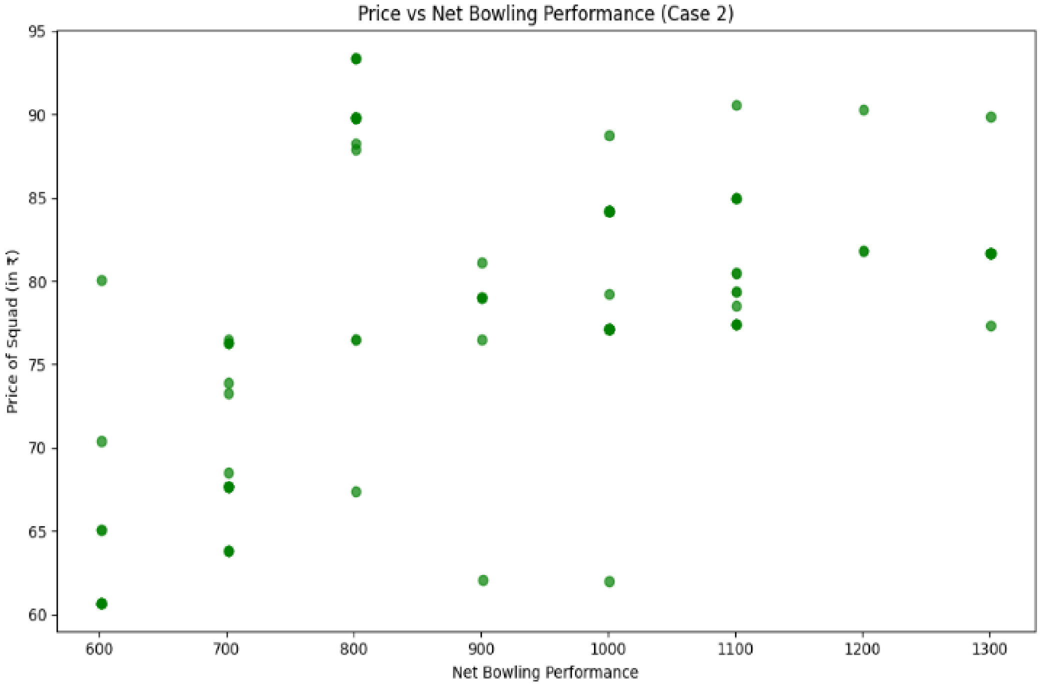
4.4.1 Price vs Net Batting Performance

The first graph shows the relationship between the squad's total cost and net batting performance. A generally positive trend can be observed, with higher batting performances correlating with an increase in squad cost. This suggests that boosting batting strength typically requires a greater investment, as high-performing batters often command higher auction prices. However, some lower-cost squads achieve reasonable batting performances, indicating that effective batters can occasionally be acquired without exceeding budget constraints.



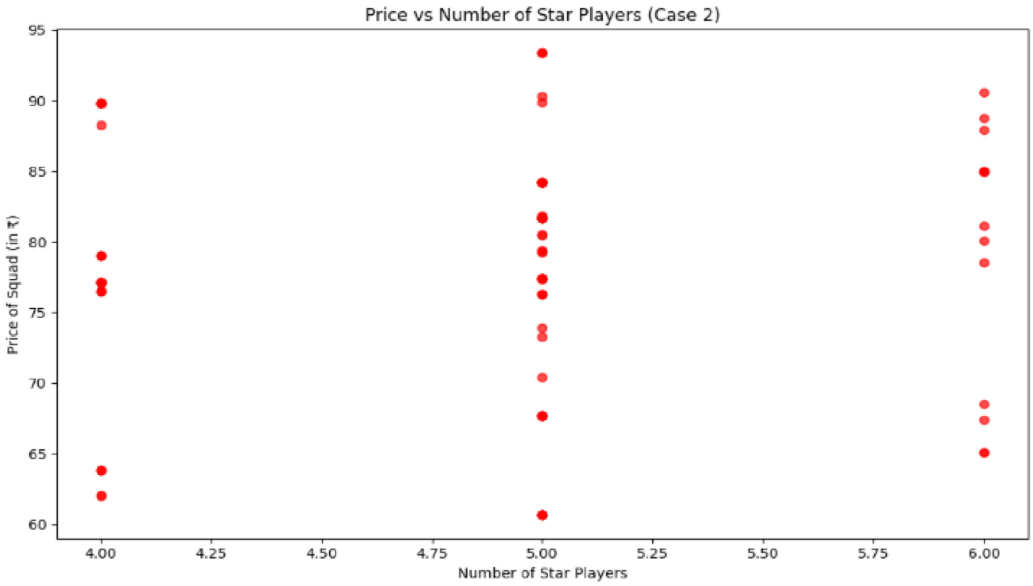
4.4.2 Price vs Net Bowling Performance

The second graph explores the link between squad cost and net bowling performance. Here, a more dispersed pattern is evident, with squads across various price points achieving comparable levels of bowling strength. This spread indicates that strong bowling lineups can be formed with both lower and higher budgets, providing flexibility in squad selection. Additionally, some high-cost squads achieve only moderate bowling performance, suggesting that high investment does not always equate to superior bowling capabilities.



4.4.3 Price vs Number of Star Players

The third graph examines the relationship between squad cost and the number of star players. The results indicate that as the number of star players increases, the squad cost tends to rise, with squads featuring more star players generally occupying higher price ranges. However, squads with four or five star players can still be found at moderate price points, allowing teams to maintain star quality while staying within budget. This distribution shows that while star players can enhance team performance, a balanced squad with fewer star players may still deliver strong results at a lower cost.



4.4.4 Player's frequency in Trade-Off Squads:

The trade-off squads generated using the BNSGA-II algorithm's run with maximum hypervolume comprise a total of 54 players in various combinations, as shown in Table 3-14. Each player's frequency indicates the number of occurrences across all possible trade-off squads, highlighting key players with higher frequencies. Calculating the frequency percentage of each player, using below equation, allows us to identify top-performing players essential for constructing an optimized squad.

$$f(\%) = \frac{\text{Frequency}}{\text{Total number of squads}} \times 100$$

Out of the 54 players, 4 players—Travis Head, Tristan Stubbs, Devdutt Padikkal, and Vaibhav Arora—have the highest frequency percentage of 100, marking them as pivotal squad members in all possible trade-offs. Additionally, several players, including MS Dhoni, Ruturaj Gaikwad, and Ravindra Jadeja, exhibit near-100 frequencies, further indicating their reliability and significance in maintaining team balance.

Interestingly, some players with lower frequencies (e.g., Prithvi Shaw and Jason Behrendorff with 2 each) contribute to specific squad compositions despite limited occurrences, showing that niche players still have value in specialized contexts.

Player Name	**(f%)**	**Player Name**	**(f%)**
Travis Head	100.0	Gus Atkinson	96.0
Tristan Stubbs	100.0	Tilak Varma	96.0
Devdutt Padikkal	100.0	Yash Thakur	96.0
Vaibhav Arora	100.0	Vidwath Kaverappa	95.0
MS Dhoni	98.0	Wanindu Hasaranga	91.0
Ruturaj Gaikwad	98.0	Yudhvir Singh	90.0
Ayush Badoni	98.0	Jitesh Sharma	90.0
Rinku Singh	98.0	Sai Sudharsan	90.0
Ravindra Jadeja	98.0	Liam Livingstone	90.0
Abdul Samad	97.0	Khaleel Ahmed	89.0
Jonny Bairstow	88.0	Rahmanullah Gurbaz	88.0
Rilee Rossouw	83.0	Nehal Wadhera	83.0
Akash Deep	82.0	Karn Sharma	18.0

Player Name	(f%)	**Player Name**	(f%)
Rahul Tewatia	18.0	Rasikh Dar	17.0
Simarjeet Singh	16.0	Harshit Rana	15.0
Nathan Ellis	10.0	Ravichandran Ashwin	9.0
Mukesh Choudhary	7.0	Angkrish Raghuvanshi	5.0
Abhishek Sharma	5.0	Harpreet Brar	5.0
Glenn Phillips	4.0	Kuldeep Sen	4.0
Vyshak Vijay Kumar	3.0	Anmolpreet Singh	2.0
Mujeeb Ur Rahman	2.0	Rishi Dhawan	2.0
Akash Madhwal	2.0	Mayank Yadav	2.0
Mohammad Nabi	2.0	Prithvi Shaw	2.0
Krunal Pandya	2.0	Joshua Little	2.0
Rajvardhan Hangargekar	2.0	Washington Sundar	2.0
Jason Behrendorff	2.0	Will Jacks	2.0
Prasidh Krishna	2.0	Heinrich Klaasen	1.0
Shivam Singh	1.0		

Chapter 5

Conclusion

The analysis of IPL squad formation using the BNSGA-II algorithm for Case 2 provides valuable insights into the trade-offs between squad cost, performance, and the inclusion of star players. The findings reveal that while higher costs generally correlate with improved batting performance, bowling performance does not strictly depend on higher expenditure. This suggests that it is possible to achieve a strong bowling lineup within a modest budget, whereas substantial investment may be required to boost batting strength.

Additionally, the influence of star players on squad cost and performance demonstrates the significance of these players in driving up total expenses. However, balanced squads with fewer star players still maintain competitive performance, indicating that a well-rounded team with cost-effective selections can perform comparably to those with numerous high-profile players. This balance enables franchises to optimize both cost-efficiency and on-field effectiveness.

Overall, this study underscores the importance of strategic resource allocation in IPL squad selection. By leveraging multi-objective optimization and balancing investments across batting, bowling, and star players, teams can form competitive, cost-effective squads capable of delivering strong performances. The BNSGA-II approach provides a robust framework for navigating the complex trade-offs inherent in constructing an IPL squad within financial and performance constraints.

Bibliography

[1] S. Verma, *Enhancing NSGA-II for Multi-Objective Combinatorial Optimization Problems*, Ph.D. Thesis, Department of Applied Mathematics and Scientific Computing, Indian Institute of Technology Roorkee, Roorkee, India, April 2024.

[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[3] V. Lunawat, *Sport and development in India: Professional sport league systems*, Routledge, 2021.

[4] N. Riquelme, C. Von Lücken, and B. Barán, "Performance metrics in multi-objective optimization," in *2015 Latin American Computing Conference (CLEI)*, 2015, pp. 1–11.

[5] F. Ahmed, K. Deb, and A. Jindal, "Multi-objective optimization and decision making approaches to cricket team selection," *Appl. Soft Comput. J.*, vol. 13, no. 1, pp. 402–414, Jan. 2013.

[6] S. Chand, H. K. Singh, and T. Ray, "Team Selection Using Multi-/Many-Objective Optimization with Integer Linear Programming," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018, pp. 1–8.

[7] G. D. I. Barr and B. S. Kantor, "A criterion for comparing and selecting batsmen in limited overs cricket," *J. Oper. Res. Soc.*, vol. 55, no. 12, pp. 1266–1274, Dec. 2004.

[8] H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," *South African J. Res. Sport. Phys. Educ. Recreat.*, vol. 24, no. 2, pp. 37–44, Feb. 2002.

[9] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley Sons, 1990.

[10] S. S. Sathya and M. S. Jamal, "Applying genetic algorithm to select an optimal cricket team," in *International Conference on Advances in Computing, Communication and Control (ICAC3'09)*, 2009, pp. 43–47.

[11] D. Bhattacharjee and H. Saikia, "An objective approach of balanced cricket team selection using binary integer programming method," *OPSEARCH*, vol. 53, no. 2, pp. 225–247, Oct. 2016.