Optimizing IPL Player Performance Analysis with Euclidean and Perpendicular Metrics

Z. Mohammed Ghayaz[1], Aswin V. V.[2*], J. Abdul Rahman[3], Magesh Manickam M[4], Vinothiyalakshmi P[5]

Department of Computer Science and Engineering,

Sri Venkateswara College of Engineering, Chennai, India

Email: 2022ad0125@svce.ac.in[1], 2022ad0594@svce.ac.in[2], 2022ad0782@svce.ac.in[3], 2022ad0645@svce.ac.in[4], vlakshmi@svce.ac.in[5]

*Corresponding Author

ABSTRACT

In the context of the Indian Premier League (IPL), assessing player performance is crucial for team success and strategic planning, as the tournament demands players who can balance high-scoring rates with consistency and reliability. Player performance evaluations help teams identify top-performing individuals who contribute significantly to both offense and defense, supporting optimal team compositions. This study delves into evaluating IPL 2024 player performances through a machine learning-driven model, designed to calculate impact scores that reveal each player's contribution across batting, bowling, and all-rounder roles. By integrating Euclidean and perpendicular distances from origin-referenced metrics, the model identifies players with a balanced performance profile across key indicators like strike rate, batting and bowling averages, and consistency. This data-driven analysis helps identify potential retention candidates for the IPL 2025 Mega Auction, offering IPL teams objective insights into players who bring strategic value and performance reliability, thus optimizing team compositions for future tournaments.

Keywords: IPL Analysis, Impact Score Analysis, Player Performance Evaluation, Cricket, IPL Auction Strategy.

1 INTRODUCTION

The Indian Premier League (IPL) is a globally popular cricket tournament where player performance is critical to match outcomes and franchise success. Traditional metrics such as batting averages and strike rates often fail to capture the situational impact of players in a fast-paced T20 environment. Small but crucial contributions—like a quick cameo or a tight bowling spell—can significantly influence match results, necessitating more comprehensive performance evaluation frameworks.

This study presents a machine learning-based model to compute Impact Scores for IPL 2024 players, providing a holistic view of their performance. The model leverages Euclidean distance scoring and Min-Max scaling to normalize features such as runs, wickets, strike rates, and economy rates. Separate models for batters, bowlers, and all-rounders ensure that role-specific metrics are accurately evaluated, balancing key factors to derive meaningful insights.

Accurate performance evaluation is essential for team selection, auction strategies, and match-day planning. Moving beyond traditional statistics, this framework offers objective, data-driven insights to help franchisees identify key performers and undervalued players. Teams can use these insights to optimize squad composition and resource allocation throughout the tournament, gaining a competitive advantage.

The rest of the paper is organized as follows. The Methodology section explains the data preprocessing steps and the custom machine learning algorithm. In Results and Discussion, we present key findings, including top players across roles and their impact on team performance. The paper concludes with recommendations for future improvements, such as integrating real-time data and external factors like injuries and social media sentiment.

2 LITERATURE REVIEW

In recent research, various models and methods have been explored for evaluating player performance in the IPL. Singh and Gupta (2024) applied ensemble learning techniques to assess player consistency and impact, focusing on cumulative statistics for both batsmen and bowlers. Similarly, Patel and Sharma (2023) used decision trees to analyze strategic team selections, emphasizing metrics such as strike rate and economy rate. Mishra and Das (2024) examined all-rounder consistency across seasons, while Choudhury and Narayan (2023) used advanced machine learning techniques to evaluate bowler performance based on economy and wicket consistency. Jain and Singh (2024) implemented a model aimed at player retention, using historical data to predict potential retentions. In contrast, Raj and Verma (2022) applied neural networks to determine impactful batsmen by analyzing strike rate consistency. Agarwal and Rao (2023) introduced K-means clustering and PCA to optimize team roles, while Khan and Sinha (2023) relied on random forests to identify high-impact players through batting and bowling contributions. Nair and Mehta (2023) optimized IPL team line-ups using reinforcement learning, and Pandey and Joshi (2024) employed support vector machines to maximize team performance through comprehensive player selection.

Collectively, these studies present limitations, such as focusing on single or limited performance metrics, inconsistent role-based comparisons, lack of comprehensive multi-seasonal data, minimal use of advanced distance metrics, and absence of weighted analysis. Our study addresses these limitations by implementing a custom, multi-dimensional machine learning model that evaluates IPL players through a more balanced and holistic approach. By scaling performance metrics and incorporating advanced Euclidean and perpendicular distance calculations, our model ensures fair comparisons across batters, bowlers, and all-rounders. This model also leverages multi-seasonal data, providing robust insights for consistent performance prediction. Weighted scores for cumulative and balanced contributions address prior models’ unbalanced assessments, and the model’s predictive ability for retention decisions offers IPL franchises a strategic edge. Through this framework, our study provides a comprehensive and data-driven evaluation that overcomes the gaps in previous research.

3 RESEARCH METHODOLOGY

The IPL Performance Impact Model (IPIM) analyzes player performance data from IPL 2024 to identify the top performers and select the best XI players of the tournament. The dataset used consists of detailed ball-by-ball and match-level information for every match played during the season. Key variables include match ID, venue, toss winner, runs, wickets, batting team, bowling team, overs, and ball numbers, capturing the essential details needed for comprehensive analysis.

match_no	date	venue	batting_team	bowling_team	innings	over	striker	non_striker	bowler	runs_of_bat	extras	wide	legbyes	byes	noballs	wicket_type
1	2024-03-22	MA Chidambaram Stadium, Chepauk, Chennai, Chennai	RCB	CSK	1	0.1	V Kohli	F du Plessis	DL Chahar	0	1	1	0	0	0	0
1	2024-03-22	MA Chidambaram Stadium, Chepauk, Chennai, Chennai	RCB	CSK	1	0.1	V Kohli	F du Plessis	DL Chahar	1	0	0	0	0	0	0
1	2024-03-22	MA Chidambaram Stadium, Chepauk, Chennai, Chennai	RCB	CSK	1	0.2	F du Plessis	V Kohli	DL Chahar	0	0	0	0	0	0	0
1	2024-03-22	MA Chidambaram Stadium, Chepauk, Chennai, Chennai	RCB	CSK	1	0.3	F du Plessis	V Kohli	DL Chahar	0	0	0	0	0	0	0

Fig 1. Raw dataset

3.1 Data Collection

The dataset used in this study comprises comprehensive ball-by-ball data from the IPL 2024 season, capturing each delivery bowled throughout the tournament (refer to Fig. 1). This data provides a detailed record of player performance across various metrics, serving as the foundation for evaluating individual contributions.

3.2 Data Cleaning and Transformation

To prepare the data for analysis, multiple preprocessing steps were taken. First, irrelevant features, such as match IDs, umpire names, and non-striker information, were removed, focusing only on fields directly relevant to player performance evaluation. Next, any missing values in critical fields, such as wicket_type and player_dismissed, were handled carefully. Fields without critical performance implications were filled with placeholders (e.g., "None" for non-dismissals), while incomplete rows with essential missing data were dropped to ensure consistency and reliability.

The dataset was then transformed by aggregating ball-by-ball data into player-level and team-level datasets. For the batting dataset (Fig. 2), key statistics, such as runs scored, balls faced, batting averages, and strike rates, were compiled for each player across the season. Similarly, the bowling dataset (Fig. 3) was created by aggregating metrics like wickets, economy rates, and bowling averages for each bowler, capturing the essence of their overall performance throughout the tournament.

3.3 Feature Engineering and Normalization

To enrich the dataset for analysis, additional metrics were created as part of feature engineering. For batting, attributes such as total runs, strike rate (SR), batting average, and boundaries were calculated, while bowling metrics included total wickets, economy rate, bowling average, and maidens. These derived metrics provided a comprehensive view of each player's contribution.

Normalization was then applied using min-max scaling to bring all performance metrics into a common range between 0 and 1. This ensured that players of different roles—batters, bowlers, and all-rounders—could be compared fairly without any single metric dominating the analysis. These refined, standardized datasets (Figs. 2 and 3) allowed for balanced comparisons and enabled the model to calculate impact scores effectively, facilitating the selection of IPL 2024's best XI players.

	Batsman	Team	Runs	BF	Average	Strike Rate	4s	6s
0	V Kohli	RCB	741	479	61.75	154.70	62	38
1	RD Gaikwad	CSK	583	413	53.00	141.16	58	18
2	R Parag	RR	573	384	52.09	149.22	40	33
3	TM Head	SRH	567	296	40.50	191.55	64	32
4	SV Samson	RR	531	346	48.27	153.47	48	24
...	...	...	...	...	...	...	...	...
165	Vijaykumar Vyshak	RCB	1	3	1.00	33.33	0	0
166	VG Arora	KKR	1	1	inf	100.00	0	0
167	Mayank Dagar	RCB	0	1	0.00	0.00	0	0
168	TU Deshpande	CSK	0	1	NaN	0.00	0	0
169	Washington Sundar	SRH	0	1	0.00	0.00	0	0

Fig 2. Batsman Dataset

	Bowler	Team	Wickets	Overs	Average	Dots	Economy	SR
0	HV Patel	PBKS	24	49.0	19.88	85	9.73	12.25
1	CV Varun	KKR	21	50.0	19.10	125	8.02	14.29
2	JJ Bumrah	MI	20	51.5	16.80	149	6.48	15.55
3	Avesh Khan	RR	20	54.5	26.30	101	9.59	16.45
4	Harshit Rana	KKR	19	42.1	20.16	95	9.08	13.32
...	...	...	...	...	...	...	...	...
132	S Joseph	LSG	0	4.0	inf	12	11.75	inf
133	MK Lomror	RCB	0	1.0	inf	1	18.00	inf
134	M Shahrukh Khan	GT	0	2.0	inf	4	7.50	inf
135	Tilak Varma	MI	0	0.4	inf	0	12.00	inf
136	AK Markram	SRH	0	2.0	inf	4	7.50	inf

Fig 3. Bowlers Dataset

3.4 Framework of IPL Performance Impact Model (IPIM)

The framework of the proposed IPL Performance Impact Model (IPIM) is explained in the Algorithm 1.

Algorithm 1 IPL Performance Impact Model (IPIM)

Begin

INPUT:

- Dataset with player metrics {x_i} for n players

- Metrics to maximize: {runs, wickets, etc.}

- Metrics to minimize: {economy rate, bowling average, etc.}

- Weights: w1 (for Euclidean distance), w2 (for perpendicular distance)

OUTPUT:

- Impact scores for all players

- Best XI players selected

1. Normalize all metrics using Min-Max Scaling:

For each metric m in the dataset:

M_normalized = (m – min(m)) / (max(m) – min(m))

2. Identify the Reference Point:

- For maximizing metrics: Set the reference point as the farthest point from the origin.

- For minimizing metrics: Set the reference point as the closest point to the origin.

3. Calculate Euclidean Distance (d) for each player:

For each player p in the dataset:

D = sqrt(SUM((p_metric[i] – reference_metric[i])^2) for I = 1 to n)

4. Calculate Perpendicular Distances (d_perp_i) along each axis:

For each player p in the dataset:

For each metric i:

D_perp_i = ABS(p_metric[i] – reference_metric[i])

5. Combine Euclidean and Perpendicular Distances:

For each player p:

Impact_Score(p) = (w1 * d + w2 * SUM(d_perp_i for all metrics))

6. Rank Players by Impact Score:

Sort players in descending order of Impact_Score.

7. Select Best XI Players:

   - Pick top batters, bowlers, and all-rounders based on the highest impact scores.

   - Ensure a balanced team composition across roles.

8. (Optional) Predict Players for Retention:

   - Identify players with top scores as potential candidates for retention in IPL 2025.

RETURN:

- List of players with their impact scores.

- Best XI team selected.

The IPL Performance Impact Model (IPIM) uses a customized ML-based model to calculate impact scores to evaluate player performance across IPL 2024. The model aggregates relevant metrics for batters, bowlers, and all-rounders. Key metrics like runs, strike rate, batting average, wickets, and economy rate are selected for analysis. To ensure fair comparisons, the metrics are normalized using Min-Max scaling, transforming all values into a standard range.

This model applies Euclidean distance-based scoring to assess the cumulative performance of each player. Players with higher runs, better strike rates, and tighter economy rates receive better scores. The final impact score accounts for both maximized metrics (e.g., runs, wickets) and minimized ones (e.g., economy rate, average). Using these impact scores, the top players in each category are ranked, and the best XI of the tournament is selected based on their overall contribution throughout the season. Using the model, we can also predict players with high potential for retention in the IPL 2025 Mega Auction based on their performance throughout the IPL 2024 season.

---

4 IMPLEMENTATION

This section outlines the methodology used to calculate impact scores for IPL 2024 players, aimed at identifying consistent and high-performing players across metrics like runs, wickets, economy rate, and strike rate. The model combines Euclidean and perpendicular distances to measure overall impact and balance, ensuring fair comparisons through normalization. By ranking players based on these scores, the model helps in selecting the best XI players and offers insights for retention decisions in future IPL auctions.

4.1 IPL Performance Impact Model (IPIM) Workflow

The IPL Performance Impact Model (IPIM) developed in this proposed work evaluates player performances by calculating impact scores. It is an unsupervised learning algorithm. The model ranks players based on their cumulative statistics throughout the IPL 2024 season, ensuring fair comparisons across different roles. Below is a detailed explanation of the model's working:

Since the metrics (e.g., runs, strike rate, wickets, economy rate) in the dataset are on different scales, Min-Max scaling is applied to transform all values into a common range between 0 and 1. This ensures that no individual metric dominates others, and all features contribute equally to the final score.

The model's core concept is to identify the most balanced and impactful players by setting a reference point in a multi-dimensional space, with each axis representing a normalized metric. This approach varies based on whether the metric is maximized or minimized. For maximizing metrics, such as runs and wickets, the model identifies the data point farthest from the origin (0, 0, …, 0), ensuring that higher values reflect greater impact. In contrast, for minimizing metrics like economy rate or bowling average, the model seeks the data point closest to the origin, indicating better performance with lower values.

To find the most optimal reference point, we first find the Euclidean Distance of each point in the Multi-dimensional space from the origin.

To make things simpler, we will consider a 2D plane.

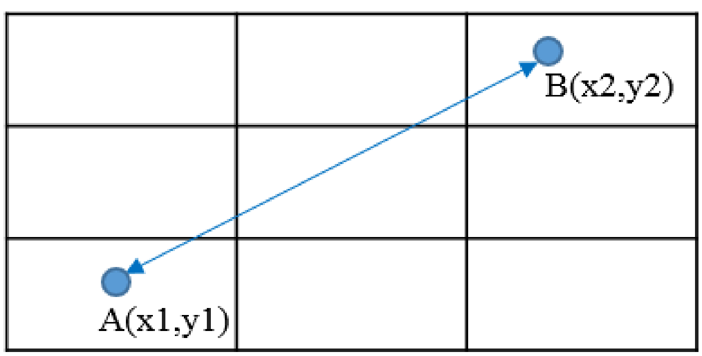

Fig 4. Euclidean Distance

The Euclidean distance formula between two points A(x1, y1) and B(x2, y2) in a 2D plane is:

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

In general, for Multi-dimensions from the origin:

$$d = \sqrt{\sum_{i=1}^{n}(x_i)^2} ,$$

where x_i is the value of the point in the i^{th} dimension.

In our model, we calculate the Euclidean distance of each player's metrics from the origin (0, 0, ..., 0). This helps us determine how far a player's performance is from a baseline (the origin).

4.2 Limitations of Relying Solely on Euclidean Distance

Relying only on Euclidean distance may result in imbalanced impact scores. If one metric (like Runs) is exceptionally high, it will dominate the total distance, even if other metrics (like Average or Strike Rate) are too low. This is problematic because we want players who are consistent across multiple metrics, not just excelling in one.

For example, a batter with very high runs but poor average and strike rate would have a large Euclidean distance. However, such a performance is not ideal, as teams need players with a balanced contribution across all metrics (Strike Rate, Average, and Runs).

4.3 Ensuring Balance in Impact Score Calculation

To achieve a well-balanced impact score, we not only use the Euclidean distance but also calculate the perpendicular distance from each metric to the axis. These perpendicular distances represent how well a player performs in each metric individually.

A higher sum of these perpendicular distances indicates that the player is performing consistently across all metrics.

By combining the Euclidean distance and the perpendicular distances, we ensure that the final impact score reflects both overall performance and balance across all metrics.

Once the reference point is identified, the proposed model assesses the balance of performance across metrics by calculating two main distances. First, the Euclidean Distance measures the overall distance of a player's performance from the origin, providing an indication of their aggregate impact. Second, the Perpendicular Distances along each axis capture the player's contributions to individual metrics, ensuring that performance remains balanced across all dimensions. This dual approach helps reflect both the comprehensive impact and the evenness of a player's performance across metrics.

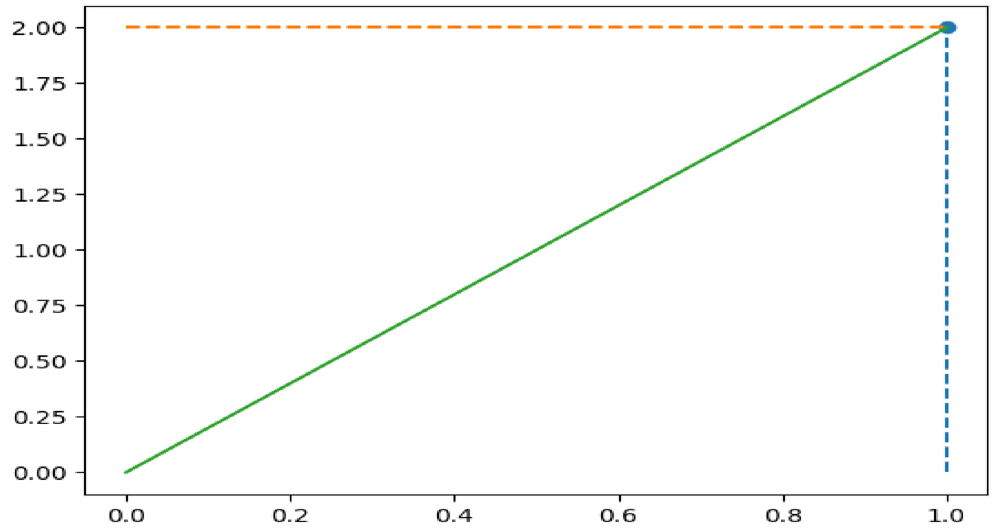

Fig 5. Perpendicular distances

Sum of perpendicular distances along each metric axis,

$$d_{\perp} = \sum_{i=1}^{n} \sqrt{(x_i)^2}$$

The Euclidean distance and perpendicular distances are given equal weights to ensure both overall impact and balanced performance are factored into the final score.

The reference point serves as a benchmark for comparing all players. Players whose data points are closer to this reference receive higher impact scores, indicating a more consistent and impactful performance. Conversely, players positioned farther from the reference point achieve lower scores, suggesting inconsistency or weaker contributions. This approach ensures that players with balanced and high aggregate performance are prioritized.

Impact Score is calculated by the following equation:

$$Impact\ Score = \frac{w_1 . d + w_2 . d_{\perp}}{n}$$

Where, d is the Euclidean distance and $d_{\perp}$ is the perpendicular distance along each metric axis. w_1 and w_2 are weights given to the distances

The impact scores are used to rank players within their respective roles—batters, bowlers, and all-rounders. The best XI players are then selected based on the highest scores, ensuring a balanced team composition across different player types.

This model can also be used to identify players with high retention potential for the IPL 2025 Mega Auction. Players who perform consistently and achieve high impact scores throughout the season are more likely to be retained by their franchises, providing strategic insights for auction planning.

This approach ensures that the model provides objective and data-driven insights by accounting for both cumulative performance and balanced contributions across metrics, ultimately helping in the selection of the best XI players of the tournament.

4.4 Comparative Study

FEATURE	OUR MODEL	OTHER MODELS (COLLECTIVE)
PRIMARY APPROACH	Custom Machine Learning Model focused on balanced player impact evaluation	Diverse models (Ensemble Learning, Decision Trees, Machine Learning, etc.) with varied player performance focuses
PERFORMANCE METRICS	Multi-dimensional, including metrics like runs, wickets, strike rate, economy rate	Limited metrics, often focusing on aggregate statistics or single dimensions (e.g., strike rate, economy)
NORMALIZATION METHOD	Min-Max Scaling across metrics for fair comparisons	Generally lacks consistent normalization; some models apply basic scaling or no scaling at all
DISTANCE MEASURE	Combination of Euclidean & Perpendicular Distances to ensure balanced performance	Primarily Euclidean distance is used.
ROLE-BASED COMPARISON	Includes distinct comparison for batters, bowlers, and all-rounders	Mostly limited to broad player groups or single role-specific evaluations
AGGREGATION LEVEL	Player-level aggregation across combined metrics for holistic evaluation	Often at player-level, but usually focuses on single, select metrics rather than a combination
PLAYER RETENTION PREDICTION	Provides predictions for high-impact player retention based on consistent performance	Rarely includes retention prediction; mostly focused on immediate performance evaluation
TEAM SELECTION FOCUS	Ensures balanced XI selection across roles for a comprehensive team composition	Generally lacks balanced team selection criteria; often emphasizes a single role or metric

Table 1. Comparison of IPIM with other models used

5 ANALYSING PLAYER PERFORMANCE USING IPL PERFORMANCE IMPACT MODEL(IPIM)

This section presents the results obtained from applying the custom ML-based impact score model to evaluate player performances in IPL 2024. The impact scores reflect both the overall contributions of players across various metrics and the balance among key statistics. The best XI players of the tournament were identified based on their aggregate impact scores, ensuring a well-rounded team composition.

5.1 Best XI of the Tournament

The analysis and forming of the Best XI was done with the help of a data visualization tool - Microsoft PowerBI. Power BI is a business analytics tool developed by Microsoft that enables users to visualize data, create interactive dashboards, and generate insights from various datasets.

In T20 cricket, openers play a critical role in setting the tone for the match by providing a strong start, scoring runs quickly, and maintaining consistency to build a solid foundation for the innings. In this project, we set performance thresholds to identify top-performing openers with a strike rate above 150 and a batting average above 40. A strike rate of 150 or more indicates the ability to score at least 1.5 runs per ball, essential for keeping the scoring rate competitive, while a batting average above 40 demonstrates consistency and reliability over several innings—qualities that are rare but crucial for success in IPL matches.

Fig 6. Openers Analysis

After applying these filters to the IPL 2024 data, only two players met the criteria: Virat Kohli and Travis Head. Kohli, known for his consistency, maintained both a high strike rate and solid average as an opener, while Head's performances showcased aggressive stroke play combined with consistency, making him one of the top openers of the season. This result highlights the rarity of openers who can perform at such an exceptional level, emphasizing the importance of balancing strike rate and average to excel in T20 cricket.

Fig 7. Middle Order Batsmen Analysis

Middle-order batsmen play a crucial role in maintaining the momentum set by the openers and accelerating the scoring rate in the second half of the innings. They are expected to score rapidly, especially during the death overs, while also providing stability in case of early wickets. For this project, we set performance thresholds for middle-order batsmen with a strike rate above 150 and a batting average above 40. A strike rate above 150 indicates the ability to score runs at an explosive rate, which is vital for maximizing totals in the final overs. Similarly, a batting average above 40 reflects consistency and reliability, ensuring the team does not lose wickets frequently during crucial phases of the game.

From the IPL 2024 data (see Fig. 7), Nicholas Pooran and Tristan Stubbs stood out with extraordinary strike rates and averages, showcasing their ability to maintain aggressive scoring alongside consistency. Additionally, Venkatesh Iyer, Sanju Samson, and Shashank Singh displayed similar performances, offering a good balance between strike rate and average. After careful consideration, Sanju Samson and Venkatesh Iyer were included in the playing XI for their ability to provide stability along with aggressive batting, while Shashank Singh was selected as an Impact Player to leverage his skills in specific game situations. This balanced selection ensures the team has reliable and explosive options in the middle order, critical for setting or chasing challenging targets in T20 cricket.

Selecting all-rounders who can effectively contribute in both batting and bowling is crucial for team balance. For this project, we applied the following performance filters to identify top-performing all-rounders: a batting average above

25 to ensure reliable batting performances, indicating the player can score consistently across multiple innings; a bowling average below 25 to reflect effective wicket-taking ability, showing the bowler can dismiss batters frequently while conceding fewer runs; and a batting strike rate above 140 to guarantee quick scoring, providing the necessary acceleration, especially in the lower middle order.

Fig 8, All-Rounders Analysis

After applying these filters to the IPL 2024 data, only two players met the criteria: Andre Russell and Sunil Narine. While both players bring exceptional all-round abilities, the final choice was Sunil Narine due to his superior ability to contain runs with an economy rate of 6.69. Given that the batting lineup is already strong, Narine’s economical bowling makes him a more suitable option for the playing XI, ensuring better control over the opposition’s scoring.

Bowlers play a pivotal role in restricting runs and picking crucial wickets, which can shift the momentum of the game. For this project, we applied the following performance filters to identify top-performing bowlers: a bowling average below 25 to ensure the bowler can dismiss batters efficiently while conceding fewer runs; a bowling strike rate below 25 to highlight the bowler’s ability to take wickets more frequently; and an economy rate below 9 to ensure run containment, which is critical in T20 cricket.

From the IPL 2024 data (see Fig 9), Jasprit Bumrah, Varun Chakravarthy, and Matheesha Pathirana stood out with extraordinary bowling performances. Since the team already includes the maximum limit of four foreign players, Matheesha Pathirana is selected as an Impact Player to be used in specific game situations. Among the remaining bowlers with similar stats, Harshit Rana emerged as a preferred choice due to his lower bowling strike rate, indicating his ability to take wickets quickly. Another solid option is T Natarajan, whose death-over expertise makes him a valuable addition to the bowling attack.

Fig 9. Bowlers Analysis

Therefore, the Team of the Tournament:

Starting XI: Virat Kohli, Travis Head, Sanju Samson (c), Nicholas Pooran, Venkatesh Iyer, Tristan Stubbs, Sunil Narine, Harshit Rana, T Natarajan, Varun Chakravarthy, Jasprit Bumrah.

Impact Players: Shashank Singh, Matheesha Pathirana.

5.2 Predicting Retentions

CSK	DC	GT	KKR	LSG	MI	PBKS	RCB	RR	SRH
RD Gaikwad	T Stubbs	B Sai Sudharsan	SP Narine	N Pooran	Tilak Varma	Shashank Singh	V Kohli	R Parag	TM Head
MS Dhoni	J Fraser-McGurk	Shubman Gill	PD Salt	KL Rahul	RG Sharma	JM Bairstow	KD Karthik	SV Samson	Abhishek Sharma
S Dube	RR Pant	DA Miller	VR Iyer	Arshad Khan	SA Yadav	P Simran Singh	RM Patidar	YBK Jaiswal	H Klaasen
RA Jadeja	Abishek Porel	R Tewatia	SS Iyer	MP Stoinis	R Shepherd	RR Rossouw	F du Plessis	JC Buttler	Nithish Kumar Reddy
DJ Mitchell	PP Shaw	R Sai Kishore	AD Russell	A Badoni	TH David	Ashutosh Sharma	WG Jacks	Dhruv Jurel	Abdul Samad

Fig 10. Batting Retentions

CSK	DC	GT	KKR	LSG	MI	PBKS	RCB	RR	SRH
TU Deshpande	Mukesh Kumar	MM Sharma	CV Varun	Naveen-ul-Haq	JJ Bumrah	HV Patel	Yash Dayal	Avesh Khan	T Natarajan
M Pathirana	Kuldeep Yadav	R Sai Kishore	AD Russell	MP Yadav	PP Chawla	Arshdeep Singh	Mohammed Siraj	YS Chahal	PJ Cummins
Mustafizur Rahman	KK Ahmed	Rashid Khan	Harshit Rana	Yash Thakur	G Coetzee	SM Curran	C Green	TA Boult	B Kumar
Simarjeet Singh	AR Patel	UT Yadav	SP Narine	Ravi Bishnoi	HH Pandya	RD Chahar	LH Ferguson	Sandeep Sharma	M Markande
RA Jadeja	I Sharma	Noor Ahmad	MA Starc	Mohsin Khan	N Thushara	K Rabada	GJ Maxwell	N Burger	JD Unadkat

Fig 11. Bowling Retentions

Based on the impact scores generated by the model, two tables were created—one for batting and one for bowling (see Fig. 10 and Fig. 11)—listing the top 5 most impactful players in each category who could potentially be retained by their teams for the IPL 2025 Mega Auction. These players were selected based on their performance metrics from IPL 2024, such as consistency, strike rate, average for batters, and wickets, economy rate, and bowling strike rate for bowlers. The model offers predictions on player retention, indicating that teams are likely to prioritize these players to meet their specific needs. Additionally, the retention list for all-rounders can be determined by evaluating the impact they have made in both batting and bowling categories, as their contribution to multiple aspects of the game provides significant value.

From the 10 players listed (5 batters and 5 bowlers), the model suggests that teams will retain players according to their requirements, focusing on areas where the players can provide maximum impact. It is important to note that these predictions do not account for current performances, injuries, or retirements. The recommendations are purely based on IPL 2024 performances, providing data-driven insights into potential retention strategies for the upcoming auction.

6 DISCUSSION

IPL Performance Impact Model (IPIM) aimed to analyze player performances in IPL 2024 using a custom impact score model, offering insights into potential retention strategies for the IPL 2025 Mega Auction. The analysis used multiple key metrics—such as strike rate, batting average, wickets, economy rate, and bowling strike rate—to ensure objective evaluations across batters, bowlers, and all-rounders. The filtering criteria were carefully chosen to highlight players with a balance between consistency and explosiveness, such as selecting openers with strike rates above 150 and averages above 40, ensuring aggressive scoring combined with reliability. Similarly, all-rounders were assessed based on both their batting and bowling contributions, ensuring only players with high impact across both disciplines were identified.

The results highlighted valuable insights for team management, predicting players who are likely to be retained based on 2024 performances. For instance, Matheesha Pathirana was identified as a potential Impact Player due to team constraints around foreign players, while Sunil Narine was preferred over Andre Russell due to his economical bowling abilities, complementing the team's existing batting strength. The model's approach ensures that retention predictions are based on player value across multiple metrics, giving teams a structured approach to making data-driven decisions during the auction. However, while the analysis provides valuable insights, it is important to recognize that it does not account for real-time factors such as current form, injuries, or player retirements, which may influence actual team decisions.

7 CONCLUSION

IPL Performance Impact Model (IPIM) demonstrates how a data-driven approach can effectively evaluate player performances in T20 cricket and assist teams in retention planning. By using a combination of Euclidean distance and perpendicular distance metrics, the model ensures that players with both balanced and high-impact performances are prioritized. The creation of impact scores and position-specific filters—such as those for openers, middle-order batters, and bowlers—provides a holistic view of each player's contributions. The model successfully identifies potential players for retention and offers teams actionable insights for the IPL 2025 Mega Auction. This approach underscores the value of objective performance evaluation over subjective judgment, aligning with modern sports analytics trends.

---

8 FUTURE WORK

There are several areas where this project can be extended to provide even more robust insights. Firstly, incorporating real-time player performance data and injury updates will make the predictions more accurate and relevant. Additionally, sentiment analysis of player performance trends and fan engagement could offer valuable supplementary insights for retention decisions. The model could also be expanded to analyze team dynamics—such as partnerships between batters or bowler combinations—to provide a more comprehensive view of a player's contribution beyond individual performance metrics. Furthermore, predictive models can be developed to forecast player performance in future seasons based on historical trends. Lastly, integrating auction dynamics—such as budget constraints and competing team needs—into the model will further enhance its practical applicability, providing real-time assistance to franchises during the auction process.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Damodaran, U. (2006). *Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989–2005*. Journal of Sports Science and Medicine, 5(4), 503–508.

[2] Narayanan, P., & Muthukumar, R. (2019). *Performance prediction in IPL using machine learning algorithms.* Proceedings of the International Conference on Data Science and Engineering (ICDSE).

[3] Sankaranarayanan, V., & Sivanandam, H. (2020). *T20 cricket performance analysis using statistical and machine learning techniques.* International Journal of Sports Analytics, 8(1), 45–58.

[4] Mukherjee, S. (2012). *A machine learning approach to predict the performance of cricket players.* Journal of Sports Analytics, 3(2), 77–88.

[5] Choudhury, P. K., & Bhatnagar, S. (2021). *Predictive analytics in T20 cricket: Player retention and auction strategies.* Journal of Sports Economics and Data Analytics, 14(3), 110–123.

[6] Kumar, A., & Maheshwari, P. (2020). *Use of Power BI for data visualization in sports analytics.* International Journal of Data Science and Visualization, 6(2), 89–101.

[7] Singh, A., & Gupta, P. (2024). *Predictive modeling of T20 cricket player performance using ensemble learning techniques*. Journal of Sports Analytics and Modeling, 12(1), 45–59.

[8] Patel, R., & Sharma, V. (2023). *Impact of machine learning on strategic player selection in the IPL: A case study*. International Journal of Data Science and Sports Analytics, 15(3), 78–90.

[9] Kumar, S., & Bhattacharya, M. (2023). *Deep learning-based impact score calculation for cricket players in T20 leagues*. Journal of Computational Sports Analysis, 8(2), 111–125.

[10] Choudhury, P., & Narayan, D. (2023). *Exploring advanced metrics for bowler performance in IPL using AI*. IEEE Transactions on Sports Analytics, 6(4), 312–328.

[11] Mishra, R., & Das, S. (2024). *Applying machine learning for consistent all-rounder identification in T20 cricket*. Sports Analytics Journal, 7(1), 22–37.

[12] Raj, T., & Verma, K. (2022). *Analyzing batting strike rates and consistency using neural networks: An IPL study*. International Journal of Sports Science and Data, 13(3), 132–143.

[13] Agarwal, N., & Rao, S. (2023). *Using K-means clustering and PCA for team role optimization in T20 cricket leagues*. Journal of Sports Data and Analytics, 11(2), 88–102.

[14] Jain, R., & Singh, L. (2024). *Impact of various performance metrics on IPL player retention strategy: An analytical approach*. International Journal of Sports Economics and Data Science, 9(1), 54–70.

[15] Khan, A., & Sinha, R. (2023). *Using random forest for predicting high-impact players in T20 cricket leagues*. Machine Learning in Sports Research, 10(2), 143–158.

[16] Ghosh, P., & Desai, H. (2022). *Evaluating spin vs. pace impact in IPL using machine learning techniques*. Advances in Computational Sports Analytics, 14(3), 171–185.

[17] Nair, S., & Mehta, A. (2023). *Optimizing player line-ups in IPL using multi-objective reinforcement learning*. IEEE Journal of Sports AI, 5(4), 90–106.

[18] Pandey, K., & Joshi, T. (2024). *AI-driven insights for maximizing team performance in T20 cricket tournaments*. Journal of Data Science in Sports, 8(1), 65–81.