Crickonomics: The Moneyball Strategy for IPL

1. ABSTRACT

This project applies the Moneyball philosophy to the Indian Premier League (IPL) by using data analytics to replace intuition-driven player selection. Player performance and value were evaluated using statistical modeling, machine learning, and Monte Carlo simulations. The results show that a team built from undervalued players can achieve competitive performance at a significantly lower cost, demonstrating the effectiveness of data-driven decision-making in professional sports.

2. INTRODUCTION

In the Indian Premier League (IPL), players are acquired through a highly competitive auction system where franchises bid within a fixed salary cap. While this process aims to ensure fairness, player selection is often influenced by reputation, recent form, and bidding dynamics rather than objective performance value, leading to inefficient spending. This project addresses this problem by applying data analytics to evaluate player performance and cost-effectiveness, enabling the construction of a competitive IPL team built on undervalued talent rather than intuition.

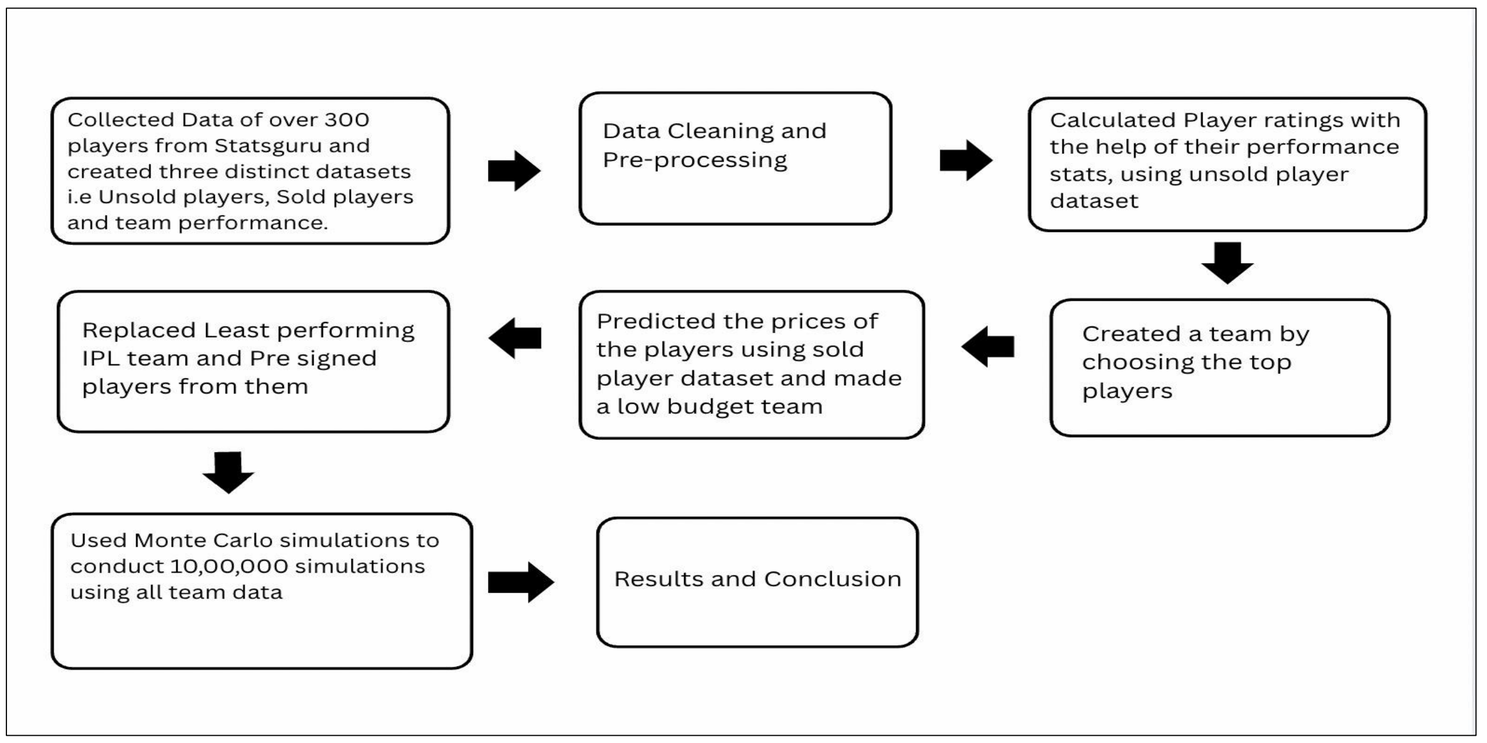

Image 01: Flow of the project

Role	Feature	Formula	PCA Weight
Batting	Hard Hitter	(Fours*4 + Sixes*6) / Balls_faced	0.167 (16.7%)
	Finisher	NO / Bat_inns	0.158 (15.8%)
	Fast Scorer	Strike Rate	0.173 (17.3%)
	Consistent	Batting Average	0.166 (16.6%)
	Boundary Frequency	(Fours + Sixes) / Balls_faced	0.168 (16.8%)
	Innings Builder	Runs_scored / Bat_inns	0.169 (16.9%)
Bowling	Economy	Econ (The raw economy rate)	0.173 (17.3%)
	Wicket Taker	1 / Bowling Avg	0.155 (15.5%)
	Strike Bowler	1 / Bowl_sr	0.193 (19.3%)
	Pressure Builder	1 / Econ	0.173 (17.3%)
	Consistent Bowler	Wickets / Bowl_inns	0.183 (18.3%)
	Big Wicket Potential	(4 + 5) / Bowl_inns	0.122 (12.2%)
Fielding	Dismissals Per Inning	Dismissal / Matches	0.588 (58.8%)
	Wicket Keeping Bonus	Stumps / Matches (for wicket-keepers)	0.412 (41.2%)

Image 02: PCA With Feature Engineering and its Weights

```
==========================================
FINAL COMPARISON
==========================================
Model          R² Score    RMSE
------------------------------------------
XGBoost        0.4486      ₹15,806,301.64
KNN            0.5310      ₹14,577,825.24
RandomForest   0.4505      ₹15,779,684.44

🎯 BEST MODEL: KNN (R²: 0.5310)
```

Image 03: Comparison of Price Estimation Algorithms

3. METHOD / PROCESS

- **Player Rating:** Batting, bowling, and fielding data were processed to engineer role-specific T20 performance features, and Principal Component Analysis (PCA) was applied to generate balanced player ratings while minimizing bias from volume-based statistics.
- **Price Estimation Pre-Auction & :** Historical auction data was modeled using machine-learning techniques to predict expected player prices and identify undervalued, cost-efficient talent. Three core players were pre-signed with pre-determined prices as per IPL rules.
- **Monte Carlo Simulation:** The final squad was evaluated using a Monte Carlo simulation of one million IPL seasons to estimate probabilistic outcomes such as wins, playoff qualification, and championship success.

	Player Name	Role	Predicted Price
1	Devdutt Padikkal	Batter	₹ 5,42,08,226
2	Prithvi Shaw	Batter	₹ 1,19,76,736
3	Mayank Agarwal	Batter	₹ 3,85,99,829
4	Anmolpreet Singh	Batter	₹ 2,39,41,040
5	Daryl Mitchell	All-Rounder	₹ 4,49,10,976
6	K.S Bharat	Wicket-Keeper	₹ 1,21,04,808
7	Dewald Brevis	Batter	₹ 83,47,995
8	Shardul Thakur	Bowler	₹ 8,36,31,187
9	Krishnappa Gowtham	Bowler	₹ 5,54,14,217
10	Ashton Turner	All-Rounder	₹ 4,75,20,384
11	Tom Curran	All-Rounder	₹ 3,90,48,130
12	Utkarsh Singh	Bowler	₹ 45,22,789
13	Yash Dhull	Batter	₹ 96,69,612
14	Luvnith Sisodia	Wicket-Keeper	₹ 53,29,149
15	Matthew Short	Batter	₹ 1,01,32,302
16	Jimmy Neesham	Bowler	₹ 2,28,16,917
17	Keemo Paul	Bowler	₹ 3,43,16,591
18	Jofra Archer	Bowler	₹ 12,00,00,000
19	Riyan Parag	All-Rounder	₹ 12,00,00,000
20	Yashasvi jaiswal	Batter	₹ 20,00,00,000
		Total Price =	₹ 94,64,90,888

Image 04: Final Team Draft After Price Estimation

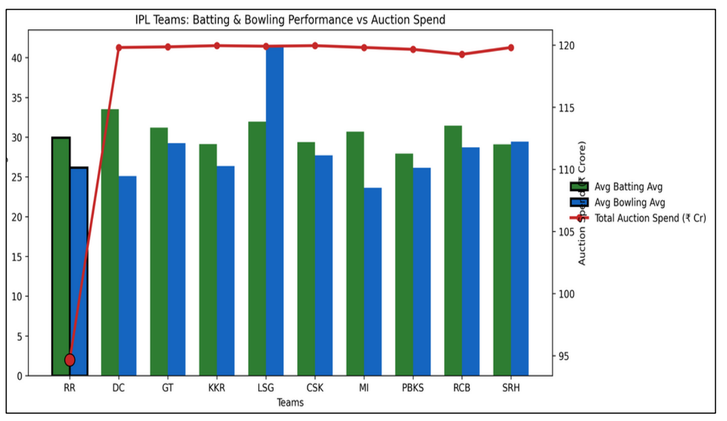

Image 05: Team Performance V/S Auction Spend

```
CHAMPIONSHIP PROBABILITIES:
----------------------------------
Team Championship Wins  Win %
  MI              137,254 13.73%
  DC              134,391 13.44%
 LSG              110,233 11.02%
  RR              104,401 10.44%
  GT              103,636 10.36%
PBKS               92,918  9.29%
 CSK               90,912  9.09%
 KKR               88,654  8.87%
 RCB               80,210  8.02%
 SRH               57,391  5.74%
```

Image 06: Championship Probabilities

4. RESULTS

- **Team Composition:** The final squad was assembled at an estimated total cost of ₹94.65 Crores, remaining well below the IPL salary cap of ₹120 Crores, while maintaining balanced coverage across all playing roles.
- **Simulation Outcomes:** Monte Carlo simulations of 1,000,000 IPL seasons showed that the data-driven team achieved an average of 14.6 league points, with a 39.57% probability of playoff qualification and a 10.44% probability of winning the championship.
- **Performance Insight:** Despite being composed largely of undervalued and unsold players, the analytically selected team performed on par with established franchises, outperforming several traditional teams in both playoff qualification and title probability.

5. CONCLUSION

This project demonstrates that a data-driven, Moneyball-inspired approach can improve IPL team selection by replacing intuition with objective analytics. Using engineered performance metrics, PCA-based player ratings, and machine-learning price estimation, the framework identifies undervalued players who deliver strong impact at lower cost. A Monte Carlo simulation of one million IPL seasons shows that the analytically constructed team, built at approximately ₹94.65 Cr, achieves competitive performance with playoff and championship probabilities comparable to established franchises, highlighting the practical value of analytics in professional sports management.

REFERENCES

[1] Joshi, Vilas, et al. "IPL data analysis and prediction using machine learning." Int J Sci Res Comput Sci Eng Inform Technol. https://doi.org/10.32628/cse it2390359 (2023).

[2] Prakash, C. Deep, C. Patvardhan, and Sushobhit Singh. "A new machine learning based deep performance index for ranking IPL T20 cricketers." International journal of computer applications 137.10 (2016): 42-49.

[3] Herath, Erandi K., and Udaya Wijenayake. "Statistical and Exploratory Data Analysis on Indian Premier League." Proceedings of Conference on Transdisciplinary Research in Engineering. Vol. 1. No. 1. 2024.

86032300016 | Shitiz Kr. Gupta
86032300043 | Keegan Nunes
86032300049 | Shrey Agarwal
86032300020 | Shashvath Arun

BSc Data Science
Mentor: Dr. Pradnya Khandepakar

SVKM'S NMIMS Deemed-to-be UNIVERSITY
NILKAMAL SCHOOL OF MATHEMATICS, APPLIED STATISTICS & ANALYTICS