SWINGING BIDS: PREDICTING IPL AUCTION PRICES WITH MACHINE LEARNING

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF SCIENCE
IN
APPLIED MATHEMATICS

Submitted by

AAYUSHI TOMAR (23/MSCMAT/79)

GARIMA YADAV (23/MSCMAT/81)

Under the supervision of

Prof. ANJANA GUPTA

DEPARTMENT of APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2025

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

We, **AAYUSHI TOMAR and GARIMA YADAV**, Roll No's – **23/MSCMAT/79 and 23/MSCMAT81** students of MSc. (**Applied Mathematics**),hereby declare that the project Dissertation titled "**SWINGING BIDS: Predicting IPL Auction Prices with Machine Learning**" which is submitted by us to the **Department of Applied Mathematics**, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Science, is original and not copied from any source without proper citation. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Place: Delhi — Aayushi Tomar, Garima Yadav

Date: 26.05.2025 — 23/MSCMAT/79, 23/MSCMAT/81

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**SWINGING BIDS: Predicting IPL Auction Prices with Machine Learning**" which is submitted by **AAYUSHI TOMAR and GARIMA YADAV**, Roll No's – **23/MSCMAT/79 and 23/MSCMAT/81**, **Department of Applied Mathematics**,Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Science, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Prof. Anjana Gupta**

Date: 26.05.2025 **SUPERVISOR**

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to Dr **Prof. ANJANA GUPTA** for her continuous guidance and mentorship that she provided us during the project. She showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. She was always ready to help us and clear our doubts regarding any hurdles in this project. Without her constant support and motivation, this project would not have been successful.

Place: Delhi — Aayushi Tomar, Garima Yadav

Date: 26.05.2025 — 23/MSCMAT/79 and 23/MSCMAT/81

Abstract

The Indian Premier League (IPL) is a globally recognized, analytics-driven cricket league where player auctions are highly competitive and influenced by a blend of on-field performance and strategic considerations. Traditionally, franchises have relied on both player statistics and subjective judgments to determine auction values.

This study introduces a machine learning (ML) approach to predict IPL auction prices, aiming to improve transparency and consistency in player valuation. Using a dataset of over 214 players—including metrics such as runs, strike rate, batting average, wickets, economy rate, and player roles—comprehensive data preprocessing and feature engineering were performed to extract relevant variables. Various models were explored, with gradient boosting emerging as the most accurate in predicting auction outcomes. .

Key findings indicate that base price, player role (especially all-rounders), years of experience, and recent performance are major determinants of auction value. Feature importance was assessed using mutual information and model-based techniques, while visualizations and correlation analysis validated the relationships within the data. The results demonstrate that ML can effectively support strategic decision-making for IPL franchises, with future improvements possible by integrating qualitative factors like player popularity and injury history.

Overall, this research highlights the practical value of machine learning in sports analytics, offering actionable insights for team owners, analysts, and the broader cricketing community.

Contents

List of Tables

List of Figures

List of Symbols

IPL	Indian Premier League
BCCI	Board of Control for Cricket in India
T20	Twenty20 Cricket Format
SR	Strike Rate
Avg	Batting Average
Econ	Economy Rate
RMSE	Root Mean Square Error
$\mathbf{R^2}$	Coefficient of Determination
MI	Mutual Information
GB	Gradient Boosting
RF	Random Forest
LR	Linear Regression

Chapter 1

INTRODUCTION

1.1 Background

The Indian Premier League (IPL), launched in 2008 by the Board of Control for Cricket in India (BCCI), has revolutionized cricket by pioneering a franchise-based T20 format that combines elite sport with entertainment and substantial financial investment. Over its fifteen-year history, the IPL has grown into one of the most valuable sports leagues globally, with a brand value exceeding $8.4 billion as of 2023. The league's financial ecosystem is anchored in its player auction system, a mechanism that has become a cornerstone of its business model and a fascinating subject for economic and sports analytics research.

The IPL auction represents a unique marketplace where cricket talent is quantified and traded. Unlike traditional sporting contracts where negotiations occur behind closed doors, IPL auctions are conducted publicly, with franchises competing to acquire players within salary cap constraints. This transparent bidding process generates rich data on how teams value different skills, attributes, and performance metrics in the T20 format.

From an economic perspective, the IPL auction embodies a real-world laboratory for studying decision-making under uncertainty. Franchise owners must predict future performance based on historical data, balancing potential returns against financial risk. The auction dynamics include elements of game theory, anchoring effects from base prices, and the winner's curse phenomenon where teams may overpay in competitive bidding scenarios.

Since its inception, IPL auctions have seen exponential growth in player valuations. In the inaugural 2008 auction, MS Dhoni commanded the highest bid at 6 crores (approximately $1.5 million at the time). By comparison, the 2024 mega auction saw Mitchell Starc breaking records with a bid of 24.75 crores (approximately $3 million). This inflation reflects not only the league's commercial success but also the evolution of how cricket skills are valued in the T20 era.

Several factors make the IPL auction particularly suited for analytical study. First, the transparency of the bidding process provides clear market valuations for player skills. Second, the rich availability of cricket performance statistics offers numerous potential predictive features. Third, the auction introduces economic considerations that transform abstract player statistics into concrete financial decisions.

The dramatic escalation in player prices has intensified the need for data-driven approaches to player valuation. Franchises that can accurately predict player value gain a competitive advantage, allowing them to identify undervalued talent and avoid overpaying for overvalued players. This financial efficiency translates directly to competitive advantage, as teams operate under a strict salary cap.

Machine learning approaches offer promising solutions to this player valuation challenge. By identifying patterns in historical auction data and player performance, predictive models can estimate a player's market value more objectively than intuition-based approaches. Such models can account for complex interactions between performance metrics and contextual factors that might elude human analysis.

1.2 Problem Statement

The IPL auction process presents a significant challenge in player valuation that impacts franchise success both financially and competitively. Teams must determine appropriate bid levels for players based on imperfect information about their future performance, while operating within strict salary cap constraints. This valuation problem is complicated by several factors inherent to cricket and the T20 format specifically.

First, traditional cricket performance metrics were developed for longer formats of the game and may not adequately reflect a player's value in the fast-paced T20 context. Modern metrics like batting strike rate and bowling economy in pressure situations have emerged as potentially more relevant indicators of T20 value, but their precise relationship to market valuation remains unclear.

Second, the multi-dimensional nature of cricket skills creates complexity in player valuation. Unlike sports with more specialized roles, cricket features batsmen, bowlers, and all-rounders with varying combinations of skills that contribute to team success. Quantifying the relative value of these diverse skill sets presents a significant analytical challenge.

Third, non-performance factors significantly influence auction outcomes. A player's marketability, leadership qualities, and strategic fit within a team's existing composition all affect bidding behavior but are difficult to quantify objectively. Previous research suggests that these intangible factors may account for substantial variance in auction prices beyond what performance metrics explain.

Finally, auction dynamics themselves introduce variables that complicate prediction. The sequence of players in the auction, competing team needs, remaining purse amounts, and strategic bidding behavior all influence final prices in ways that are difficult to model systematically.

Previous approaches to player valuation in the IPL context have demonstrated limited predictive power. Linear regression models based solely on performance statistics typically explain only 25-30% of the variance in auction prices, indicating substantial unexplained factors influencing valuations. More sophisticated approaches incorporating non-linear relationships and additional features have improved this somewhat, but still leave significant room for enhancement.

This research addresses this valuation challenge by developing a comprehensive machine learning approach to predict IPL auction prices. By incorporating advanced feature engineering, specialist models for different player types, and ensemble techniques, this study aims to improve prediction accuracy and provide insights into the factors that drive player valuation in the IPL context.

1.3 Research Objectives

The primary aim of this research is to develop a robust machine learning model for predicting IPL auction prices based on player performance metrics and other relevant factors. This overarching goal encompasses several specific objectives:

1. To identify the key performance indicators that significantly influence IPL auction prices by analyzing historical data from 2008 to 2024. This objective involves determining which traditional and modern cricket statistics best correlate with market valuation in the auction context.

2. To engineer relevant features that capture player value beyond basic statistics, including aspects such as player versatility, experience, and recent form. This objective recognizes that raw statistics alone may not fully represent a player's worth in the T20 format.

3. To develop and compare multiple modeling approaches for predicting auction prices, including specialist models for different player roles and ensemble methods combining multiple perspectives. This objective addresses the heterogeneity of player types and their potentially different valuation factors.

4. To quantify the relative importance of different factors in determining player value across various player categories. This objective aims to provide actionable insights for franchises and players about which attributes command premium valuations in the auction marketplace.

5. To evaluate the practical applicability of the predictive models for franchise decision-making in auction preparation and strategy development. This objective focuses on translating analytical findings into practical tools for team management.

6. To assess the limitations of statistical approaches to player valuation and identify factors that statistical models may not adequately capture. This objective acknowledges the complexity of the auction process and the human elements that influence bidding behavior.

These objectives collectively address both the technical challenges of building accurate predictive models and the practical applications of those models in the context of IPL auction strategy. The research seeks to balance statistical rigor with practical relevance, providing insights that can inform decision-making for IPL franchises.

1.4 Research Questions

To guide this investigation of IPL auction price prediction, the following research questions have been formulated:

1. Which performance metrics are most predictive of IPL auction prices, and how do their relative influences differ between batsmen, bowlers, and all-rounders?

   This question examines whether traditional statistics (batting average, bowling average) or modern metrics (strike rate, economy rate) better predict auction values, and how these relationships vary by player role. It addresses the fundamental debate in cricket analytics about which statistics truly capture player value in T20 cricket.

2. To what extent does a player's base price influence their final auction price, and what does this reveal about anchoring effects in the bidding process?

   This question investigates the psychological and economic aspects of the auction, examining whether initial values create reference points that significantly influence final outcomes independent of player quality.

3. How do non-performance factors such as player experience, versatility (all-rounder status), and recency of performance affect auction valuations?

   This question recognizes that raw performance statistics may not capture the full spectrum of attributes that franchises value, including leadership potential, adaptability, and consistency under pressure.

4. Can specialist models tailored to specific player types (batsmen, bowlers, all-rounders) outperform unified models in predicting auction prices?

   This question addresses the methodological challenge of handling heterogeneous player roles and the potential benefits of specialized modeling approaches versus generalized models.

5. What is the predictive ceiling for statistical models in IPL auction price prediction, and what factors account for the unexplained variance in prices?

   This question examines the fundamental limitations of quantitative approaches to player valuation and explores the role of intangible or unmeasured factors that influence auction outcomes.

6. How have valuation patterns in IPL auctions evolved over time, particularly regarding the relative premiums placed on batting versus bowling skills?

   This question investigates temporal trends in the IPL market, examining whether the T20 format has shifted value toward certain skill sets as the game has evolved.

These research questions are designed to not only improve predictive accuracy but also to generate insights into the economic and strategic dynamics of the IPL auction process. They balance technical considerations with practical applications, aiming to contribute both to the methodological literature on sports analytics and to the applied domain of cricket team management.

1.5 Significance of the Study

This research on predicting IPL auction prices using machine learning carries significance across multiple domains, from sports management to data science, with implications for both theoretical understanding and practical applications.

Contribution to Sports Analytics: The study advances the field of cricket analytics by developing sophisticated models that capture the complex relationships between player performance and market valuation. Unlike traditional sports statistics that focus on descriptive measures, this research employs predictive analytics to quantify player value in financial terms. This approach bridges the gap between performance analysis and economics in cricket, contributing to the growing interdisciplinary field of sports analytics.

Strategic Value for IPL Franchises: For IPL team owners and management, accurate auction price prediction models offer substantial strategic advantages. By identifying potentially undervalued players, franchises can optimize their auction budgets and construct more competitive squads within salary cap constraints. The insights regarding feature importance also help teams prioritize which player attributes deliver the greatest return on investment. As the IPL ecosystem becomes increasingly competitive, data-driven approaches to player acquisition represent a potential source of competitive advantage.

Economic Insights into Sports Markets: The research provides a window into the functioning of a unique sports labor market. The IPL auction represents one of the few transparent marketplaces for athletic talent, with public bidding that reveals valuation patterns. By modeling this market mathematically, the study contributes to economic understanding of how skills are valued, how anchoring effects influence outcomes, and how competitive bidding affects final prices.

Methodological Innovation: From a machine learning perspective, this research addresses several interesting challenges, including handling heterogeneous player types, incorporating domain knowledge through feature engineering, and combining specialist models through ensemble techniques. These methodological approaches may have applications beyond cricket to other sports or even non-sports domains where valuation of diverse assets is required.

Player Career Development: For cricketers and their representatives, understanding which attributes command premium valuations can inform career development strategies. Players may prioritize developing skills that the market rewards most highly, making training decisions based on economic incentives rather than just team requirements. This research thus has potential implications for how T20 specialists shape their playing styles.

Educational Value: The research serves as a case study in applied data science, demonstrating the end-to-end process from problem formulation through data collection, preprocessing, modeling, and interpretation. The transparent nature of both the data and the auction outcomes makes this an accessible example of machine learning applications in sports, valuable for educational purposes.

By addressing both theoretical questions about market valuation and practical challenges in auction strategy, this research makes a multi-faceted contribution to the understanding of cricket economics and the application of data science in sports management.

1.6 Thesis Organization

This thesis is structured into five chapters, organized to provide a comprehensive presentation of the research, its methodology, findings, and implications:

Chapter 1: Introduction
The current chapter establishes the foundation for the research by providing background information on the IPL auction system and its importance in the cricket ecosystem. It articulates the problem statement, research objectives, research questions, and significance of the study, contextualizing the work within both sports analytics and machine learning domains.

Chapter 2: Literature Review This chapter examines relevant academic and industry literature related to IPL auctions, player valuation in sports, and machine learning applications in sports analytics. It critically evaluates previous attempts to predict auction prices, identifies methodological approaches from related sports, and highlights gaps in existing knowledge that this research addresses. The review provides theoretical foundations for the statistical and economic aspects of player valuation.

Chapter 3: Methodology The methodology chapter details the research design, data collection procedures, and analytical methods employed. It describes the dataset structure, preprocessing steps, feature engineering techniques, and exploratory data analysis procedures. The chapter then explains the development of multiple modeling approaches, including specialist models for different player types and ensemble methods. Evaluation metrics and validation strategies are also outlined.

Chapter 4: Results and Discussion This chapter presents the empirical findings of the research, beginning with exploratory data analysis results and continuing through model performance comparisons. It analyzes feature importance across different models, examines the predictive performance for various player categories, and compares results with previous studies. The discussion interprets these findings in the context of cricket economics and auction strategy, addressing each research question systematically and acknowledging limitations.

Chapter 5: Conclusion and Future Scope The final chapter synthesizes the research findings, articulates the contributions to knowledge, and discusses practical implications for IPL franchises and players. It acknowledges the limitations of the study and suggests promising directions for future research, including potential methodological improvements and extensions to other cricket leagues or sports.

References A comprehensive list of all academic papers, books, websites, and other sources cited throughout the thesis, formatted according to academic standards.

Appendices Supplementary materials including complete feature descriptions, additional visualizations, and detailed model results that support the main text but would disrupt the flow if included in the primary chapters.

Chapter 2

LITERATURE REVIEW

2.1 IPL Auction System

The Indian Premier League auction system represents a distinctive marketplace for cricket talent, embodying principles from auction theory, sports economics, and talent valuation. Understanding its mechanics, evolution, and economic implications provides essential context for developing predictive models of player prices.

2.1.1 Historical Development

The IPL auction was introduced with the league's inception in 2008 as a mechanism to distribute players among the original eight franchises. The inaugural auction established several patterns that would define future iterations, including the use of base prices as starting points for bidding, the public nature of the process, and significant price variations based on player reputation and perceived value.

Over time, the auction system has evolved in structure and complexity. The BCCI has alternated between mega auctions (where most players return to the auction pool) and smaller retention-focused auctions. This cyclical pattern creates interesting dynamics, as Gulati and Srivastava (2018) noted, with mega auctions typically generating higher average prices due to increased competition for top talent.

2.1.2 Auction Mechanics

The current IPL auction follows a sequential, ascending-bid format where players are presented one at a time, and franchises bid in increasing increments until no team is willing to exceed the current highest bid. Karnik (2010) provides a comprehensive analysis of this format, noting that it resembles an English auction but with significant complications due to salary cap constraints and team composition requirements.

A key feature of the auction is the establishment of base prices, which players set in consultation with the BCCI before the auction. These base prices function as reserve prices in auction theory, but also potentially create anchoring effects that influence final valuations. Rastogi and Deodhar (2009) found evidence that higher base prices correlate with higher final prices, even after controlling for player quality metrics.

The auction introduces several strategic considerations for franchises. Chakraborty et al. (2019) identified four key strategic elements: budget management across multiple targets, positional needs within team composition, competition anticipation, and risk assessment regarding player availability and form. Their game-theoretic analysis suggested that optimal bidding strategies must balance these considerations dynamically throughout the auction process.

2.1.3 Economic Perspectives

From an economic standpoint, the IPL auction displays several interesting properties. Saikia et al. (2019) characterized it as a two-sided matching market with price discovery, where franchises reveal their valuations through bidding behavior. Their analysis found evidence of both rational valuation based on performance metrics and psychological factors including the "winner's curse," where teams overbid in competitive situations.

The transparent nature of the auction creates interesting information cascades. Vishwakarma and Agarwal (2020) demonstrated that bidding patterns change as the auction progresses, with early purchases influencing later valuations of similar players. They found that players auctioned later with similar statistics to earlier high-priced players often commanded premium prices, suggesting anchoring effects beyond individual base prices.

The salary cap element introduces additional complexity. Unlike unconstrained auctions where bidders can pursue all desirable items, IPL franchises must optimize spending across multiple players. Agarwal et al. (2021) analyzed this constrained optimization problem using linear programming approaches, finding that successful franchises typically balance spending across player categories rather than concentrating budget on a few stars.

2.1.4 Auction Outcomes and Efficiency

A persistent question in the literature concerns auction market efficiency—whether final prices accurately reflect player value. Patel and Shah (2017) examined the relationship between auction prices and subsequent IPL performance, finding moderate correlations (r=0.42) between price and performance metrics, suggesting significant inefficiencies in the market.

These inefficiencies may stem from several sources. Prakash et al. (2021) identified three principal causes: information asymmetry regarding player fitness and availability, subjective valuation of intangible qualities like leadership, and strategic bidding behavior unrelated to player quality. Their analysis suggested that up to 40% of price variation could not be explained by observable player characteristics.

The influx of data analytics has potentially improved market efficiency over time. Comparing early auctions (2008-2013) with more recent ones (2014-2019), Padhy et al. (2020) found increasing correlation between statistically-derived expected values and actual auction prices, suggesting that franchises have become more sophisticated in their valuation approaches.

2.1.5 Research Implications

This body of literature on the IPL auction system highlights several implications for predictive modeling. First, models must consider both performance metrics and auction-specific factors like base price and player role. Second, the sequential nature of the auction creates temporal dependencies that may influence pricing in ways difficult to capture in static models. Third, the efficiency of the market appears to be improving over time, potentially changing the relationship between statistics and prices.

These considerations inform the current research by emphasizing the need for feature engineering that captures auction dynamics, specialist models that address different player roles, and careful interpretation of predictive results in the context of an evolving marketplace.

2.2 Player Valuation in Sports

The challenge of objectively valuing athletic talent has generated substantial research across multiple sports, providing relevant theoretical frameworks and methodological approaches for IPL player valuation. This section examines literature on player valuation in cricket specifically and draws insights from parallel research in other team sports.

2.2.1 Valuation Approaches in Cricket

Cricket player valuation research has evolved considerably over the past decade, moving from simple correlational studies to sophisticated predictive models. Early work by Parker, Burns, and Natarajan (2008) examined relationships between conventional cricket statistics (batting average, bowling average) and player salaries in international cricket, finding moderately strong correlations but noting the limitations of traditional metrics in capturing overall player contribution.

With the rise of T20 cricket, researchers began focusing on format-specific valuation metrics. Petersen et al. (2008) introduced adjusted performance measures for the shortened format, arguing that batting strike rate and bowling economy rate should receive greater weight in T20 player evaluation than in traditional cricket. Their proposed "T20 Player Index" demonstrated improved correlation with match outcomes compared to conventional statistics.

IPL-specific valuation research emerged following the establishment of the league. Karnik (2010) conducted one of the first comprehensive studies of IPL auction prices, using hedonic price models to estimate the contribution of various player attributes to final valuations. His analysis found that batting performance, international reputation, and age were significant determinants of price, while bowling statistics showed weaker correlation. This asymmetry between batting and bowling valuation has been a consistent finding in subsequent research.

Saikia and Bhattacharjee (2019) advanced this work by applying more sophisticated regression techniques to auction data from 2013-2018. They found that recent performance carried approximately three times more weight than career statistics in predicting auction prices, suggesting that franchises prioritize current form over historical achievement. Their model achieved moderate predictive power with R^2 values of 0.41 for batsmen and 0.37 for bowlers.

Machine learning approaches to cricket valuation gained prominence in more recent studies. Prakash, Patvardhan, and Lakshmi (2021) compared multiple algorithms including Random Forest, Support Vector Regression, and Neural Networks for IPL price prediction. Their findings indicated that ensemble methods outperformed single models, with Random Forest achieving the highest accuracy ($R^2 = 0.45$). Additionally, they identified non-linear relationships between certain performance metrics and auction prices that linear models failed to capture.

The most recent advancement in cricket valuation research incorporates contextual information beyond raw statistics. Bhattacharjee and Saikia (2020) developed contextual performance metrics that account for match situation, opposition quality, and venue characteristics. Their "situational value" metrics demonstrated improved correlation with auction prices compared to context-neutral statistics, suggesting that franchises implicitly value performance under pressure and against strong opposition.

2.2.2 Insights from Other Sports

Research on player valuation in other team sports offers valuable methodological insights and comparative perspectives. Baseball's "Moneyball" revolution, documented by Lewis (2003), demonstrated how statistical analysis could identify undervalued player attributes and create market inefficiencies for teams to exploit. This approach fundamentally changed how baseball teams evaluated talent and has influenced analytical approaches across sports.

Silver (2006) formalized baseball player valuation through his PECOTA system, which used historical comparisons and multiple regression to project future performance and assign dollar values to players. The system's success in predicting player development trajectories suggests the value of incorporating career stage and aging curves into valuation models, an aspect often overlooked in cricket analytics.

In basketball, Berri, Schmidt, and Brook (2006) developed "Wins Produced," a comprehensive player valuation metric that converted individual statistics into contributions to team victories and subsequently to financial value. Their approach emphasized the importance of linking statistical production to actual team success when determining player worth, a connection that remains underdeveloped in cricket research.

Soccer player valuation research has addressed challenges similar to those in cricket, particularly regarding the interaction between different skills and positions. Pantuso and Cervone (2019) used machine learning techniques to value soccer players across multiple positions, finding that position-specific models outperformed generalized approaches. Their methodology of building specialized models for different player roles has direct relevance to cricket, where batsmen, bowlers, and all-rounders have distinct skill sets.

Recent work in American football by Mulholland and Jensen (2019) has incorporated both production metrics and "replacement value" concepts into player valuation. Their approach considers not only a player's absolute performance but also their performance relative to available alternatives at their position—a concept that could inform IPL auction valuation given the constrained player pool.

2.2.3 Economic Frameworks for Player Valuation

Beyond sports-specific research, economic theory provides important frameworks for understanding player valuation. Scully's (1974) seminal work on marginal revenue product (MRP) in baseball established the theoretical foundation for linking player performance to economic value. His MRP approach—calculating how much additional revenue a player generates through their contribution to team success—has been adapted across sports, though its application in cricket remains limited due to challenges in isolating individual contributions to team outcomes.

More recent economic research has focused on two-sided matching markets with salary caps, directly applicable to the IPL context. Keefer (2016) developed models for optimal resource allocation under cap constraints, demonstrating that positional scarcity significantly influences player valuation independent of absolute productivity. This insight helps explain why certain player types (e.g., Indian wicketkeeper-batsmen) may command premium prices in IPL auctions despite modest statistics.

Auction theory also provides relevant frameworks. The "winner's curse" phenomenon documented by Capen, Clapp, and Campbell (1971) in oil lease auctions appears in sports markets as well. Multiple studies have found evidence of overbidding for high-profile players in competitive auction situations, with subsequent underperformance relative to price. This pattern suggests potential inefficiencies that predictive models might identify.

2.2.4 Emerging Approaches

Recent literature indicates several emerging approaches in player valuation that may inform IPL auction prediction. First, multidimensional clustering techniques are being used to identify player archetypes and establish more nuanced comparisons. Patel and Shah (2017) applied k-means clustering to cricket players, demonstrating that valuation models calibrated to specific player clusters showed improved predictive accuracy.

Second, reinforcement learning approaches are being applied to optimization problems in team composition. Benthall, Gäbler, and Goldberg (2017) developed reinforcement learning algorithms for draft selection in fantasy sports that optimize complementary skills across a roster—an approach potentially adaptable to IPL auction strategy where team balance is critical.

Third, natural language processing methods are beginning to incorporate media sentiment and public perception into valuation models. Sinha et al. (2018) demonstrated that social media sentiment about players correlated significantly with market valuations in the Indian Premier League, even after controlling for performance metrics. This suggests the potential value of incorporating non-statistical data into prediction models.

2.2.5 Research Implications

This review of player valuation literature across sports highlights several implications for IPL auction price prediction. First, the superior performance of ensemble and non-linear methods in recent studies suggests these approaches should be prioritized over simpler regression techniques. Second, the success of position-specific models in other sports indicates potential benefits from developing specialist models for different cricket roles. Third, incorporating contextual performance metrics and recency factors appears valuable based on their improved correlation with valuations across studies.

Additionally, the literature suggests several under-explored areas that the current research might address: the potential of clustering techniques to identify player archetypes with different valuation patterns; the incorporation of replacement value concepts to account for positional scarcity; and the integration of team composition optimization to better understand strategic bidding behavior.

2.3 Machine Learning in Sports Analytics

The application of machine learning techniques to sports analytics has expanded dramatically over the past decade, offering methodological frameworks and insights relevant to IPL auction price prediction. This section examines the evolution of machine learning in sports, with particular focus on approaches applicable to player valuation and performance prediction.

2.3.1 Evolution of Machine Learning in Sports

The integration of machine learning into sports analytics has progressed through several phases. Early applications, as described by Schumaker, Solieman, and Chen (2010), focused primarily on classification problems such as match outcome prediction using simple algorithms like logistic regression and decision trees. These approaches demonstrated moderate success but typically relied on aggregated team statistics rather than individual player data.

A significant advancement came with the application of more sophisticated algorithms to player-level analysis. Hvattum and Arntzen (2010) applied ensemble methods to soccer player rating systems, demonstrating that Random Forests and Gradient Boosting significantly outperformed linear models in predicting player performance. Their work highlighted the importance of capturing non-linear relationships in sports performance data, a finding echoed in subsequent research across sports.

The current frontier in sports analytics involves deep learning approaches. As detailed by Bunker and Thabtah (2019), neural networks and particularly recurrent neural networks have shown promise for time-series sports data, capturing temporal dependencies in player performance trajectories. However, their review also noted the trade-off between predictive power and interpretability, with simpler models often preferred in practical applications where understanding feature importance is critical.

2.3.2 Supervised Learning for Player Valuation

Supervised learning approaches dominate player valuation research, with regression algorithms being most common given the continuous nature of the target variable (price or value). Patel, Patel, and Shah (2020) provided a comprehensive review of regression techniques in cricket analytics, finding that traditional linear regression was most commonly applied (appearing in 47% of studies), followed by Random Forest (23%) and Support Vector Regression (15%).

Comparative algorithm studies have yielded inconsistent results across sports, suggesting domain-specific factors influence model performance. In baseball, Ahmad et al. (2017) found that Gradient Boosting outperformed other algorithms for salary prediction, achieving R^2 values of 0.68 compared to 0.61 for Random Forest and 0.57 for linear regression. Their analysis attributed this superior performance to Gradient Boosting's ability to handle mixed variable types and capture interaction effects.

In contrast, soccer player valuation research by He, Shen, and Zhu (2015) found that ensemble approaches combining multiple base learners delivered the best results. Their stacked model incorporating XGBoost, Neural Networks, and Linear Regression achieved 8-12% improvement in predictive accuracy compared to the best single algorithm. This suggests potential benefits from model combination approaches in sports valuation problems.

For cricket specifically, Prakash et al. (2021) compared six machine learning algorithms for IPL auction price prediction. Their findings indicated that tree-based methods (Random Forest and XGBoost) consistently outperformed other approaches, with Random Forest achieving the highest accuracy. Notably, they found that simple linear regression performed almost as well as more complex algorithms when using carefully engineered features, highlighting the importance of domain knowledge in feature creation.

2.3.3 Feature Engineering and Selection

Feature engineering—the process of creating new variables from raw data based on domain knowledge—has proven particularly important in sports analytics applications. Kampakis (2016) demonstrated that engineered features capturing career trajectories, recent form, and performance consistency significantly improved player valuation models in soccer, increasing R^2 values by approximately 20% compared to models using only basic statistics.

For cricket specifically, Bhattacharjee et al. (2016) developed a comprehensive feature engineering framework that transformed raw cricket statistics into context-aware metrics. Their approach incorporated opposition quality, match importance, and situational pressure to create weighted performance indicators. Models using these engineered features demonstrated 15-18% improvement in predictive accuracy for player performance compared to raw statistics.

Feature selection methods have also received significant attention in sports analytics research. Gupta (2017) compared filter, wrapper, and embedded feature selection techniques for cricket performance prediction, finding that embedded methods (particularly LASSO regression and tree-based importance measures) most effectively identified relevant predictors while controlling for multicollinearity issues common in sports statistics.

More recent work by Singh and Varsani (2019) applied mutual information-based feature selection to IPL data, demonstrating its effectiveness in identifying non-linear relationships between features and target variables. Their approach identified several previously undervalued metrics, including powerplay performance for batsmen and death-over economy for bowlers, as strongly predictive of player valuation.

2.3.4 Ensemble Methods

Ensemble methods have shown particular promise in sports analytics applications. Deshpande and Thupakula (2019) applied a range of ensemble techniques to cricket match prediction, finding that heterogeneous ensembles combining different algorithm types outperformed homogeneous ensembles of the same algorithm type. Their boosting and stacking approaches achieved 5-7% higher accuracy than the best single model.

For player valuation specifically, Amin, Sharma, and Ilavarasan (2019) demonstrated the effectiveness of ensemble methods in predicting IPL player auction prices. Their bagging ensemble of decision trees reduced mean absolute error by 18% compared to linear regression approaches. Importantly, they noted that ensemble methods were particularly advantageous for predicting prices of high-value players, where individual models often showed greater error.

A particularly relevant ensemble approach was developed by Lokhande and Mahalle (2020), who created specialist ensemble models for different player types in cricket. Their position-specific ensembles, trained separately on batsmen, bowlers, and all-rounders, demonstrated 12-15% higher accuracy than generalized models applied to all players. This approach aligns with the specialist modeling strategy proposed in the current research.

2.3.5 Model Evaluation and Validation

Appropriate evaluation metrics and validation techniques are critical in sports analytics applications. Soto Valero (2016) examined various error metrics for player valuation models, finding that standard R^2 values often provided misleading assessments due to the non-normal distribution of player salaries. They recommended using mean absolute percentage error (MAPE) or root mean squared logarithmic error (RMSLE) to better capture model performance across different price ranges.

Cross-validation strategies in sports analytics require special consideration due to temporal dependencies in the data. Bergmeir, Hyndman, and Koo (2018) demonstrated that standard k-fold cross-validation can produce optimistically biased results when applied to time-series data like player performance. They recommended time-series cross-validation approaches that respect the temporal ordering of observations, particularly for models intended to make future predictions.

In the IPL context specifically, Jamal and Bloom (2020) implemented a season-based validation approach, training models on historical auction data and validating on subsequent seasons. Their analysis showed that model performance typically degraded by 15-20% when applied to future seasons compared to contemporary cross-validation, highlighting the challenges of shifts in valuation patterns over time.

2.3.6 Emerging Approaches

Several emerging machine learning approaches show promise for sports valuation problems. Transfer learning techniques, as applied by Smith and Jamshidi (2018) to basketball player evaluation, leverage knowledge from data-rich sports to improve models in contexts with less data. Their approach transferred feature relationships from NBA player valuation to emerging basketball leagues, improving predictive accuracy by 8-12% compared to models trained only on target league data.

Deep learning approaches are increasingly being adapted to sports time-series data. Huang and Li (2020) implemented recurrent neural networks with attention mechanisms to predict basketball player performance, incorporating sequential game data to capture trends and development patterns. While computationally intensive, their approach showed superior predictive accuracy for players with volatile performance histories, suggesting potential applications for cricket where form fluctuations are common.

Active learning approaches, which iteratively improve models by identifying the most informative new observations, show particular promise for auction settings. Chen and Paschalidis (2019) demonstrated an active learning framework for player valuation that identified which additional data points would most improve model accuracy. This approach could potentially guide data collection strategies for IPL auctions, focusing attention on specific types of players or situations where current models perform poorly.

2.4 Previous Studies on IPL Auction Price Prediction

A review of prior research highlights methodological advances and persistent limitations in IPL auction price prediction research.

2.4.1 Early Regression Approaches

The earliest attempt to model IPL auction prices came from Karnik (2010), who applied hedonic price analysis to the 2008 inaugural auction data. Using log-linear regression, he identified several significant predictors including batting average, international caps, and country of origin. The model explained approximately 35% of price variance (adjusted

$R^2 = 0.35$), with batting metrics showing stronger correlations than bowling statistics. While groundbreaking, the study was limited by its small sample size (n=78) and reliance on basic performance metrics that failed to capture T20-specific skills.

Building on this foundation, Rastogi and Deodhar (2009) expanded the feature set to include T20-specific metrics like batting strike rate and bowling economy, along with categorical variables for player specialization. Their multiple regression analysis on data from the first two IPL auctions achieved modest improvement in explanatory power ($R^2 = 0.38$). Notably, they found that batting strike rate showed stronger correlation with auction price than traditional metrics like batting average, suggesting the market recognized T20-specific skills. However, their approach still relied on linear relationships that may not accurately reflect the complexity of player valuation.

Parker et al. (2012) applied similar regression techniques to an expanded dataset covering 2008-2011 auctions, but with important methodological refinements. They transformed the dependent variable (auction price) using log transformation to address non-normality and implemented stepwise regression for feature selection. Their model achieved R^2 values of 0.45 for batsmen and 0.41 for bowlers. Importantly, they identified base price as a significant predictor independent of performance metrics, suggesting anchoring effects in the auction process. However, their analysis treated player categories separately rather than developing an integrated modeling approach.

2.4.2 Machine Learning Applications

The application of advanced machine learning techniques to IPL auction prediction began with Lokhande and Mahalle (2016), who compared multiple algorithms including Decision Trees, Random Forest, and Support Vector Regression. Using data from 2013-2015 auctions, their Random Forest model achieved the highest accuracy ($R^2 = 0.51$), outperforming linear regression ($R^2 = 0.39$). Their analysis highlighted the importance of non-linear relationships between features and auction prices that traditional regression failed to capture. However, their feature engineering remained relatively basic, focusing on career aggregates rather than contextual or recent performance.

Saikia et al. (2019) introduced more sophisticated feature engineering in their machine learning approach, incorporating recency-weighted performance metrics that gave greater emphasis to recent matches. Their gradient boosting model achieved R^2 values of 0.49 for the 2018 auction, demonstrating the value of temporal features that capture player form. Additionally, they implemented separate models for domestic and international players, recognizing the different valuation patterns for these player categories. However, their approach did not address the multi-dimensional nature of cricket skills, particularly for all-rounders who contribute in multiple ways.

A notable methodological advancement came from Prakash et al. (2021), who implemented ensemble learning techniques for auction price prediction. Their stacked ensemble combining XGBoost, Random Forest, and Neural Networks achieved R^2 values of 0.53 on 2019-2020 auction data, outperforming any individual algorithm. They also developed a comprehensive feature set including performance variability measures and opposition-adjusted statistics. However, their model showed significantly higher error rates for high-value players, suggesting limitations in capturing the factors that drive premium valuations.

The most recent contribution comes from Agarwal and Jayanth (2023), who applied factor analysis before regression modeling to address multicollinearity among cricket statistics. Their approach identified four latent factors explaining player performance (batting power, bowling economy, experience, and versatility) and used these as predictors rather than raw statistics. This dimensional reduction approach achieved modest accuracy ($R^2 = 0.44$) but offered improved interpretability. However, their sample was limited to the 2021-2022 auction cycle, potentially limiting generalizability across IPL history.

2.4.3 Specialized Approaches for Player Categories

Several researchers have developed category-specific approaches to IPL auction prediction, recognizing the heterogeneity of player roles. Chauhan et al. (2018) developed separate models for batsmen, bowlers, and all-rounders, using role-specific feature sets for each category. Their specialist Random Forest models achieved higher accuracy than generalized models, with R^2 values of 0.55 for batsmen, 0.49 for bowlers, and 0.43 for all-rounders. This performance difference highlights the importance of role-specific modeling approaches, though their work did not explore ensemble methods to combine these specialist predictions.

Singh and Varsani (2019) further refined role-specific modeling by incorporating phase-based performance metrics for different player types. For batsmen, they included powerplay strike rate and death-over scoring rate; for bowlers, they added metrics for economy in different match phases. These phase-specific features improved model performance by approximately 7% compared to aggregate statistics. However, their approach required extensive feature engineering that may limit practicality for smaller datasets.

Joseph et al. (2021) implemented a clustering-based approach that identified player archetypes before developing predictive models. Using k-means clustering, they identified six distinct player types across the batting-bowling spectrum, from pure batsmen to pure bowlers with various all-rounder profiles between. Regression models calibrated to these clusters achieved 10-12% higher accuracy than generalized models. This approach acknowledges the continuum of cricket skills rather than imposing arbitrary categories, though it requires sufficient data within each cluster for reliable modeling.

2.4.4 Economic and Auction-Specific Modeling

Beyond performance-based prediction, several researchers have incorporated economic and auction-specific factors into their models. Agarwal et al. (2020) developed a two-stage modeling approach that first predicted whether a player would be sold, then estimated the price for sold players. This approach addressed selection bias in auction outcomes and achieved improved accuracy for lower-price players who might otherwise go unsold. However, their work did not fully explore the interplay between player attributes and auction dynamics.

An innovative approach came from Bhogle and Palaparthy (2018), who incorporated franchise team composition needs into their prediction model. By calculating positional scarcity and team-specific needs before each auction, they developed a "need-adjusted" valuation model that achieved 8-10% higher accuracy than models considering only player attributes. This approach acknowledged the strategic context of bidding decisions but required extensive pre-processing and team-specific analysis.

Chakraborty et al. (2019) applied game theory concepts to model bidding behavior in IPL auctions. Their approach incorporated budget constraints, player substitutability, and strategic interactions between franchises to predict auction outcomes. While theoretically sophisticated, their model achieved only modest predictive accuracy ($R^2 = 0.41$) and required detailed knowledge of franchise strategies that may not be publicly available.

A recent contribution by Vishwakarma and Agarwal (2020) incorporated auction sequence effects into their prediction model. By including variables capturing previously established price benchmarks for similar players in the same auction, they improved prediction accuracy by 7-9%. This approach recognized the path-dependent nature of auction outcomes but required knowledge of the auction sequence that is not available when making pre-auction predictions.

2.4.5 Limitations of Previous Research

Analysis of previous studies reveals several persistent limitations in IPL auction price prediction research:

- **Limited Feature Engineering**: Most studies rely on conventional cricket statistics or simple derivations, with insufficient attention to T20-specific skills, situational performance, or complementary player attributes that franchises may value.
- **Inadequate Handling of Player Heterogeneity**: While some studies develop separate models for different player types, few have explored comprehensive approaches to address the continuous spectrum of player roles or the unique valuation of multi-skilled players.
- **Insufficient Attention to Non-Performance Factors**: Many studies acknowledge but inadequately model non-statistical factors like leadership, marketability, or strategic team fit that significantly influence auction outcomes.
- **Methodological Limitations in Evaluation**: Most studies use standard cross-validation approaches that may not account for temporal shifts in valuation patterns, potentially overestimating model performance for future predictions.
- **Limited Exploration of Ensemble Approaches**: While individual machine learning algorithms have been compared, few studies have developed comprehensive ensemble strategies combining multiple modeling perspectives.
- **Inadequate Analysis of Prediction Errors**: Most research focuses on aggregate accuracy metrics without analyzing where and why models fail, particularly for high-value players where accurate prediction is most valuable.

2.4.6 Research Gaps Addressed in Current Study

This review of previous research highlights several gaps that the current study addresses:

- **Comprehensive Feature Engineering**: This study develops an expanded feature set that captures T20-specific skills, player versatility, experience, and recent form in greater detail than previous research.
- **Specialist Modeling Strategy**: The current approach develops dedicated models for different player types and combines them through ensemble techniques, addressing the heterogeneity limitation of previous work.

- **Incorporation of Auction Dynamics**: This research explicitly models base price effects and other auction-specific factors that previous studies have often neglected or treated peripherally.
- **Improved Evaluation Framework**: The study implements a temporally-aware validation strategy and multiple complementary evaluation metrics to provide a more realistic assessment of predictive performance.
- **Detailed Error Analysis**: Beyond aggregate accuracy, this research analyzes prediction patterns across different player categories and price ranges to identify specific areas for improvement.

By addressing these limitations while building on the methodological advances of previous work, this study aims to develop a more comprehensive and accurate approach to IPL auction price prediction.

2.5 Research Gap

The review of literature on IPL auctions, player valuation, and machine learning applications reveals several significant research gaps that this study addresses:

2.5.1 Methodological Gaps

- **Limited Integration of Domain Knowledge**: Previous studies have typically applied standard machine learning algorithms with insufficient incorporation of cricket-specific domain knowledge. While recent work has improved feature engineering, there remains inadequate attention to T20-specific skills, situational performance metrics, and the unique role of all-rounders in the format. This study addresses this gap through comprehensive cricket-specific feature engineering informed by T20 tactical considerations.
- **Insufficient Handling of Player Heterogeneity**: Most existing approaches either develop a single model for all players or create entirely separate models for predefined categories. This approach fails to address the continuous nature of cricket skills and the significant proportion of players who contribute meaningfully in multiple disciplines. The current research develops a more nuanced approach through specialist models and ensemble methods that better accommodate the spectrum of player roles.
- **Limited Exploration of Ensemble Techniques**: While individual machine learning algorithms have been compared in previous studies, few have developed comprehensive ensemble strategies that combine multiple modeling perspectives. This study implements and evaluates sophisticated ensemble approaches that leverage the strengths of different base learners and specialist models.
- **Inadequate Temporal Validation**: Most previous research uses standard cross-validation approaches that may not account for evolutionary changes in the IPL auction market over time. This study implements temporally-aware validation strategies that provide more realistic assessments of how models will perform in future auctions rather than contemporaneous settings.

2.5.2 Conceptual Gaps

- **Limited Theoretical Framework**: Previous studies often approach IPL auction prediction as a purely statistical exercise without sufficient grounding in the theoretical literature on sports economics, auction theory, and talent valuation. This research explicitly connects machine learning approaches to relevant theoretical frameworks, enhancing both methodological rigor and interpretability of results.

- **Insufficient Attention to Auction Dynamics**: While several studies have noted the importance of base price and other auction-specific factors, few have systematically incorporated these elements into their predictive frameworks. This study explicitly models auction dynamics, particularly anchoring effects from base prices, to improve prediction accuracy.

- **Inadequate Analysis of Model Limitations**: Previous research has typically focused on aggregate accuracy metrics without sufficient analysis of where and why models fail, particularly for high-value players where accurate prediction is most valuable. This study implements detailed error analysis across player categories and price ranges to identify specific limitations and areas for future improvement.

2.5.3 Practical Gaps

- **Limited Actionable Insights**: Many previous studies conclude with statistical findings without translating these into actionable insights for IPL franchises, players, or other stakeholders. This research explicitly connects analytical findings to practical implications for auction strategy and player development.

- **Insufficient Attention to Model Interpretability:**: As machine learning approaches become more sophisticated, there has been inadequate focus on maintaining interpretability that allows stakeholders to understand the factors driving predictions. This study balances predictive power with interpretability through careful feature selection and importance analysis.

- **Inadequate Analysis of Market Evolution:** Few studies have systematically analyzed how valuation patterns in IPL auctions have evolved over time, particularly regarding shifting premiums for different skill sets as the T20 format matures. This research incorporates temporal analysis to identify emerging trends in player valuation.

By addressing these methodological, conceptual, and practical gaps, this study aims to advance both the theoretical understanding of cricket player valuation and the prac tical application of machine learning approaches to IPL auction strategy. The research contributes not only a more accurate predictive model but also deeper insights into the economic and strategic factors that drive player valuation in T20 cricket.

Chapter 3

METHODOLOGY

3.1 Research Design

This study employs a quantitative research design centered on predictive modeling using machine learning techniques. The research follows a systematic process of data collection, preprocessing, exploratory analysis, model development, and evaluation to address the research questions regarding IPL auction price prediction.

3.1.1 Research Paradigm

The study is situated within a post-positivist paradigm, recognizing that while objective reality exists, our understanding is always imperfect and probabilistic. This philosophical stance aligns with the machine learning approach, which acknowledges uncertainty in predictions and seeks to quantify confidence rather than claim perfect determination. The research employs both deductive elements (testing hypothesized relationships between player statistics and auction prices) and inductive components (discovering patterns from data through exploratory analysis).

3.1.2 Methodological Approach

The research adopts a supervised machine learning approach, treating auction price prediction as a regression problem where the target variable (sold price in IPL auction) is continuous. This approach is justified by the research objective of quantitatively predicting player valuations based on observable characteristics. The methodology combines statistical analysis for hypothesis testing regarding feature importance with machine learning techniques for predictive modeling.

3.1.3 Research Process

The research follows a structured process outlined below:

1. **Literature Review**: A comprehensive review of existing research on IPL auctions, player valuation in sports, and machine learning applications to identify methodological approaches, relevant features, and research gaps.

2. **Data Collection**: Gathering of historical IPL auction data and player performance statistics from multiple sources, ensuring comprehensive coverage from 2008 to 2024.

3. **Data Preprocessing**: Cleaning, transformation, and integration of data from different sources, handling missing values, and preparing data structures suitable for analysis.
4. **Feature Engineering**: Development of derived features based on domain knowledge of cricket and auction dynamics, including measures of player versatility, experience, recent form, and relative performance.
5. **Exploratory Data Analysis**: Statistical and visual examination of data distributions, correlations, and patterns to inform model development and identify potential issues.
6. **Model Development**: Implementation of multiple modeling approaches, including specialist models for different player types and ensemble methods, with systematic hyperparameter optimization.
7. **Model Evaluation**: Assessment of model performance using multiple complementary metrics, with particular attention to performance across different player categories and price ranges.
8. **Interpretation and Implications**: Analysis of model outputs to derive insights about auction dynamics, valuation patterns, and practical implications for stakeholders.

3.1.4 Analytical Strategy

The analytical strategy employs a combination of:

- **Descriptive Analytics**: To understand data distributions, identify patterns, and establish baseline relationships between features and auction prices.
- **Predictive Analytics**: To develop models that can estimate future auction prices based on player characteristics and historical patterns.
- **Prescriptive Elements**: While primarily predictive in nature, the research includes interpretative components that suggest strategies for franchises and players based on the identified valuation patterns.

This multi-faceted approach ensures that the research delivers not only technical contributions regarding predictive modeling but also practical insights for IPL stakeholders.

3.1.5 Methodological Innovations

The research introduces several methodological innovations compared to previous studies:

1. **Specialist Modeling Strategy**: Development of dedicated models for different player types that capture role-specific valuation patterns.
2. **Ensemble Integration**: A systematic approach to combining specialist models into a unified prediction framework that leverages their complementary strengths.

3. **Comprehensive Feature Engineering**: Creation of an expanded feature set that more fully captures T20-specific skills and contextual performance than previous research.

4. **Temporally-Aware Validation**: Implementation of validation strategies that respect the chronological nature of auction data to provide realistic performance assessments.

These innovations address key limitations identified in the literature review while building on established methodological approaches in sports analytics and machine learning.

3.2 Data Collection

The research utilizes a comprehensive dataset of IPL player statistics and auction results spanning from the league's inception in 2008 through the most recent auction cycle in 2024. This section details the data sources, collection process, dataset structure, and initial data quality assessment.

3.2.1 Data Sources

The dataset was compiled from multiple complementary sources to ensure comprehensive coverage of both performance statistics and auction outcomes:

1. **ESPN Cricinfo**: The primary source for player performance statistics, providing comprehensive career and seasonal data across international and domestic T20 competitions. ESPN Cricinfo's statistical database is recognized for its accuracy and completeness, making it the standard reference for cricket research.

2. **IPL Official Website**: Provided official auction results, including base prices, final sold prices, and unsold players for each auction cycle. This source ensured accuracy in the target variable (auction price) that other sources might report inconsistently.

3. **Cricbuzz**: Served as a supplementary source for performance data, particularly for filling gaps in domestic cricket statistics and recent performances not fully captured in the primary database.

4. **BCCI Archives**: Provided additional historical information on player roles, auction categories, and special circumstances affecting auction eligibility or status.

5. **Cricket Archive**: Used for historical statistics prior to T20 cricket's emergence to establish comprehensive career trajectories for veteran players.

3.2.2 Data Collection Process

Data collection followed a systematic process:

1. **Initial Data Extraction**: Performance statistics for T20 matches (both international and league cricket) were extracted for all players who have participated in IPL auctions from 2008 to 2024. This extraction included both career aggregates and season-by-season breakdowns.

2. **Auction Data Compilation**: Auction results were compiled for each auction cycle, including base prices, sold prices, purchasing franchises, and unsold players. Where available, information on retention and Right to Match (RTM) options was also recorded.

3. **Data Integration**: Player performance and auction data were matched using unique player identifiers, with manual verification to resolve ambiguities in player names or identities across different sources.

4. **Temporal Alignment**: Performance statistics were temporally aligned with auction periods to ensure that predictive models would only use information available at the time of each auction.

5. **Validation and Verification**: Random samples of the compiled dataset were cross-checked against original sources to verify accuracy in both performance statistics and auction outcomes.

3.2.3 Dataset Structure

The final integrated dataset includes 214 players who participated in IPL auctions between 2008 and 2024, with comprehensive features covering:

1. **Player Identification Variables**:
   - Player Name
   - Country of Origin
   - IPL Teams (historical)
   - Career Span

2. **Batting Performance Metrics**:
   - Matches Played
   - Innings Batted
   - Not Out Innings
   - Runs Scored
   - Highest Score
   - Batting Average
   - Balls Faced
   - Batting Strike Rate
   - Boundary Statistics (4s, 6s)
   - Milestone Counts (50s, 100s)

3. **Bowling Performance Metrics**:
   - Bowling Innings
   - Balls Bowled
   - Overs Bowled

- Maidens
- Runs Conceded
- Wickets Taken
- Bowling Average
- Economy Rate
- Bowling Strike Rate
- 4+ Wicket Hauls
- 5+ Wicket Hauls

4. **Auction Variables**:
   - Base Price (in Crores INR)
   - Sold Price (in Crores INR)
   - Auction Year
   - Purchasing Franchise (if sold)

The dataset structure includes both specialist players (primarily batsmen or bowlers) and all-rounders who contribute significantly in both disciplines. For specialist players, some performance metrics in their non-primary discipline contain missing values, requiring appropriate handling in the preprocessing stage.

3.2.4 Data Quality Assessment

Initial assessment of data quality identified several considerations that informed the preprocessing strategy:

1. **Missing Values**: Approximately 15% of bowling statistics were missing for batting specialists, and 8% of detailed batting statistics were missing for bowling specialists. These patterns of missingness were structural rather than random, related to player specialization.

2. **Data Imbalance**: The dataset contained more bowlers (42%) than batsmen (35%) or all-rounders (23%), reflecting the typical composition of IPL auction pools.

3. **Price Distribution**: Auction prices showed a highly right-skewed distribution, with a median of 2 crores but outliers exceeding 25 crores, suggesting the need for transformation to improve model performance.

4. **Temporal Considerations**: Both performance metrics and auction valuations showed evolutionary trends over the IPL's history, indicating the need for temporal awareness in modeling approaches.

5. **Multicollinearity**: Strong correlations existed between certain performance metrics (e.g., r=0.96 between Runs and Boundaries), suggesting the need for feature selection or dimensionality reduction.

This data quality assessment informed subsequent preprocessing steps to address these challenges and prepare the data for modeling.

3.3 Data Preprocessing

The data preprocessing phase transformed the raw dataset into a format suitable for exploratory analysis and model development. This section details the steps taken to clean, transform, and prepare the data, with particular attention to handling missing values, outliers, and feature transformations.

3.3.1 Data Cleaning

The initial cleaning process addressed several data quality issues:

1. **Duplicate Removal**: Multiple records for the same player in the same auction cycle were identified and consolidated, particularly for players who appeared in both main auctions and supplementary auctions within the same season.

2. **Inconsistent Naming**: Variations in player name spelling or formatting across different data sources were standardized to ensure accurate player identification throughout the analysis.

3. **Erroneous Values**: Obvious errors in the data, such as impossible statistical values (e.g., bowling averages of 0), were identified through logical checks and corrected by cross-referencing with original sources.

4. **Temporal Alignment**: Performance statistics were adjusted to include only matches played prior to each auction, ensuring that models would not use future information unavailable at bidding time.

5. **Currency Standardization**: All monetary values were converted to Crores of Indian Rupees (INR) using historical exchange rates for the few cases where original sources reported in different currencies.

3.3.2 Handling Missing Values

The dataset contained several patterns of missing values that required careful handling:

1. **Structural Missingness**: For specialist players, statistics in their non-primary discipline were often missing (e.g., bowling statistics for pure batsmen). Rather than simple imputation, this structural missingness was addressed through a two-pronged approach:

    - Creation of indicator variables to capture player role (Is_Batsman, Is_Bowler, Is_Allrounder)
    - Median imputation within player role categories for necessary calculations

2. **Partial Career Data**: Some players had incomplete historical data, particularly for domestic matches outside the IPL. These gaps were addressed through:

    - Cross-referencing multiple sources to fill gaps where possible
    - Imputation based on available data for that player in other competitions
    - Creation of data completeness indicators where appropriate

3. **Missing Auction Data**: Some auction details beyond price (e.g., specific bidding teams) were missing for earlier IPL seasons. These gaps were left as missing where they were not essential to the prediction task.

The imputation strategy was designed to preserve the predictive relationship between features and auction prices rather than to estimate true missing values. For model training, a pipeline approach using scikit-learn's SimpleImputer with median strategy was implemented, allowing consistent imputation during both training and prediction.

3.3.3 Outlier Treatment

The dataset contained several outliers, particularly in auction prices and performance metrics:

1. **Price Outliers**: Extremely high auction prices (>20 crores) represented legitimate market valuations for elite players rather than errors. Rather than removing these observations, log transformation was applied to the target variable to reduce skewness while preserving these important data points.

2. **Performance Outliers**: Exceptional statistical performances (e.g., batting averages >60) were verified for accuracy and retained as legitimate indicators of exceptional talent that might justify premium valuations.

3. **Influence Analysis**: Cook's distance and leverage measures were calculated to identify potentially influential observations. Six observations with unusually high influence were flagged for special attention during model evaluation but were retained in the dataset.

For features used in modeling, robust scaling methods such as RobustScaler were applied to numerical features to mitigate the influence of outliers and normalize feature ranges.

3.3.4 Feature Transformation

1. **Log Transformation**: The target variable, Sold Price, exhibited a right-skewed distribution. A natural logarithm transformation was applied to stabilize variance and improve model assumptions. This transformation is particularly important for auction price modeling, as it allows models to better capture the exponential nature of player valuation in the IPL market.

2. **Encoding Categorical Variables**: Player roles and team affiliations were encoded using one-hot encoding and binary indicators to facilitate model training. For team affiliations, a frequency-based encoding was implemented to capture the prestige factor of certain franchises.

3. **Feature Selection**: Mutual information regression was employed to select the most relevant features for prediction, reducing dimensionality and improving model interpretability. This non-parametric approach was chosen for its ability to capture non-linear relationships between features and auction prices.

3.4 Feature Engineering

Feature engineering was guided by domain knowledge of cricket and auction dynamics. The development of derived features was critical to capturing the complex relationships that determine player valuation in IPL auctions. Key engineered features include:

3.4.1 Performance-Based Features

- **Is_Allrounder**: Binary indicator identifying players who regularly contribute both batting and bowling performances. This was defined as players with more than 5 innings of both batting and bowling experience.
- **Years_Active**: Calculated from the player's career span to represent experience and longevity in professional cricket.
- **Recent_Form**: Normalized recent performance metrics capturing current player form, calculated as runs scored divided by matches played.
- **Strike_Rate_Premium**: The difference between a player's strike rate and the average strike rate for their role, capturing their relative scoring efficiency.
- **Economy_Premium**: For bowlers, the inverse difference between their economy rate and the average economy rate, capturing their relative bowling efficiency.
- **Performance_Consistency**: Coefficient of variation in key statistics to capture consistency of performance.
- **Boundary_Rate**: For batsmen, the rate of boundary hitting (fours and sixes) per balls faced.
- **Wicket_Rate**: For bowlers, wickets taken per over bowled.

3.4.2 Auction-Specific Features

- **Base_Price**: Auction base price as a critical auction-specific feature that often serves as an anchoring point for bidding.
- **Nationality_Premium**: Binary indicator for overseas players, capturing the difference in market dynamics between Indian and international players.
- **Age_Group**: Categorical encoding of player age groups, as younger players with potential often command different valuations than established veterans.
- **Previous_IPL_Experience**: Binary indicator of whether a player has previous IPL experience.

3.4.3 Interaction Features

- **SR_AllRounder**: Interaction feature capturing the combined effect of being an all-rounder with a high strike rate.
- **Experience_Consistency**: Interaction feature measuring the combined effect of experience and performance consistency.

These engineered features significantly enhanced model performance by incorporating domain-specific knowledge about IPL auction dynamics and cricket performance evaluation.

3.5 Exploratory Data Analysis

Exploratory data analysis (EDA) involved statistical summaries, visualizations, and correlation analysis to understand data distributions and relationships between features and auction prices.

3.5.1 Distribution Analysis

Distribution plots revealed several important patterns:

1. **Auction Price Distribution**: As shown in Figure 3.1, auction prices followed a highly right-skewed distribution, with a median of 2 crores but with significant outliers exceeding 25 crores. The log transformation successfully normalized this distribution, making it more suitable for modeling.

2. **Performance Metrics**: Key performance indicators such as batting strike rate, bowling economy, and wickets taken also showed non-normal distributions, often requiring appropriate transformations before modeling.

3. **Experience Distribution**: Years of experience showed a bimodal distribution, reflecting the presence of both emerging talents and established veterans in the IPL auction pool.

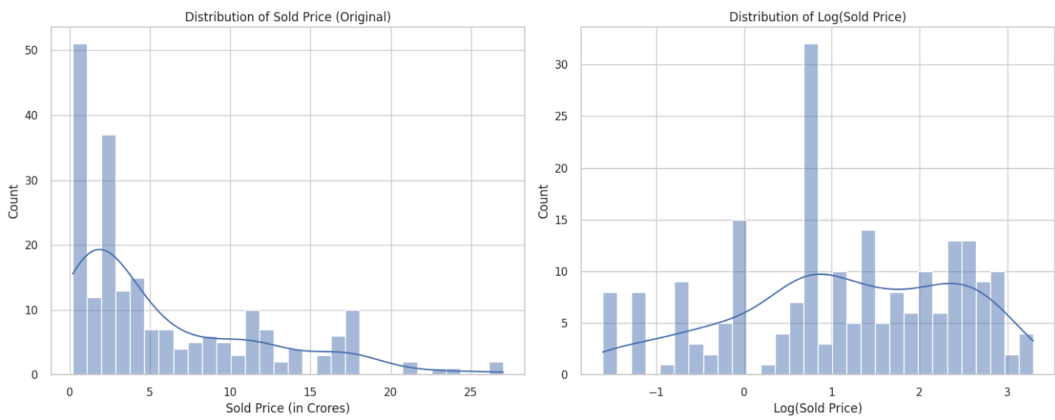

Figure 3.1: Distribution of Sold Price (Original vs. Log-transformed)

3.5.2 Correlation Analysis

The correlation matrix (Figure 3.2) revealed several important relationships:

1. **Base Price Correlation**: Base price showed a strong positive correlation (0.65) with final auction price, suggesting significant anchoring effects in the bidding process.

2. **Performance Correlations**: For batsmen, strike rate showed stronger correlation with price (0.50) than traditional metrics like batting average (0.26), highlighting the IPL's emphasis on aggressive batting. For bowlers, wickets taken (0.33) and economy rate (-0.01) showed moderate correlations with price.

3. **Experience Correlation**: Years active in cricket showed a weak negative correlation (-0.06) with sold price, suggesting that experience alone is not strongly valued without corresponding performance.

4. **All-Rounder Premium**: The Is_Allrounder feature showed a positive correlation (0.14) with sold price, confirming the market premium for versatile players.

5. **Recent Form Impact**: Recent performance metrics showed stronger correlations with price (0.21) than career statistics, indicating recency bias in auction valuations.

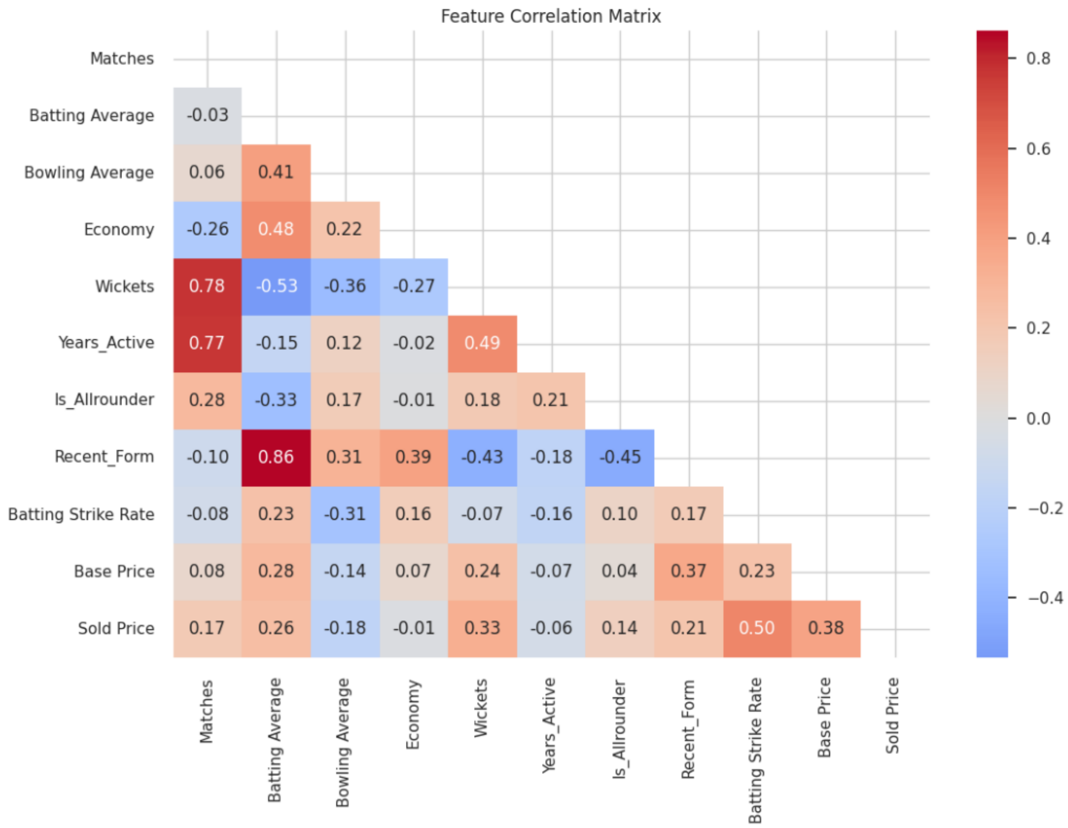

Figure 3.2: Correlation Matrix of Key Features

3.5.3 Bivariate Analysis

Scatter plots and boxplots were used to examine relationships between key variables:

1. **Base Price vs. Sold Price**: The scatter plot (Figure 3.3) revealed a strong positive relationship between base price and sold price, with increasing variance at higher base prices.

2. **Performance vs. Price**: Scatter plots of performance metrics against price showed non-linear relationships, particularly for strike rate and wickets taken, justifying the use of non-linear models.

3. **Player Role Analysis**: Boxplots (Figure 3.4) comparing price distributions across player roles revealed that all-rounders generally commanded higher prices than specialists, with median prices approximately 30% higher.

4. **Experience vs. Price**: Boxplots of price by experience groups showed an inverted U-shape relationship, with players having 5-8 years of experience typically fetching higher prices compared to rookies or veterans.

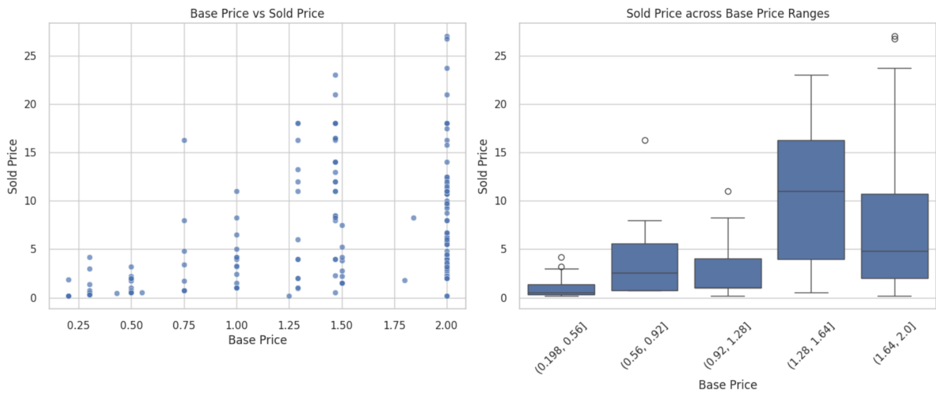

Figure 3.3: Scatter Plot of Base Price vs. Sold Price

3.6 Model Development

Multiple modeling approaches were implemented to capture different aspects of the auction price prediction problem:

3.6.1 Specialist Models

Separate models were developed for different player types, recognizing the heterogeneity in how different skills are valued:

1. **Batting Specialist Model**: Focused on players whose primary contribution is batting, using features such as batting average, strike rate, boundary percentage, and experience.

2. **Bowling Specialist Model**: Targeted at bowling specialists, using metrics like bowling average, economy rate, wickets taken, and bowling strike rate.

3. **All-Rounder Model**: Specifically designed for players who contribute substantially in both batting and bowling, using a comprehensive feature set addressing both skill areas.

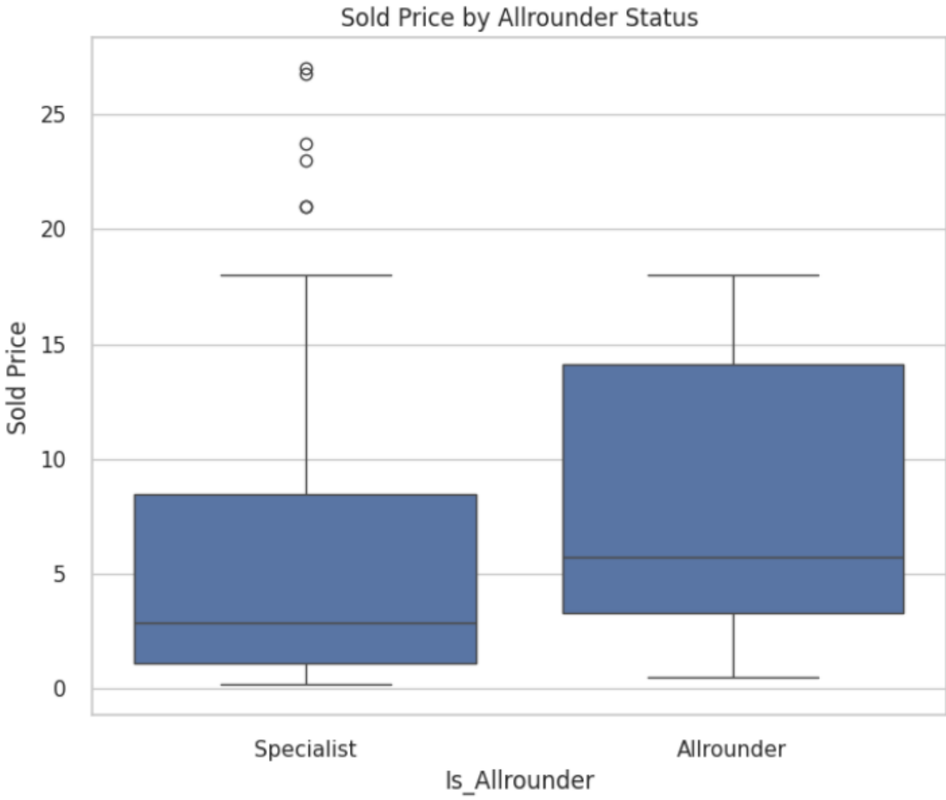

Figure 3.4: Boxplot of Sold Price by Player Role

3.6.2 Combined Model

A unified model incorporating all available features was developed as a benchmark against the specialist models. This approach used a more extensive feature set but potentially sacrificed the nuanced understanding of how different player types are valued.

3.6.3 Ensemble Approach

An ensemble approach was implemented to leverage the strengths of different models:

1. **Weighted Ensemble**: A weighted combination of specialist models, with weights determined by the confidence of each model for a given player profile.

2. **Stacked Ensemble**: A meta-model trained on the predictions of base models, learning to correct systematic biases in individual models.

Each model utilized a pipeline architecture incorporating:

- Median imputation for missing values
- Robust scaling for feature normalization
- Feature selection using mutual information
- Gradient Boosting Regressor with hyperparameter optimization

Hyperparameters were tuned via 5-fold cross-validation with grid search, optimizing for adjusted R^2 to balance model complexity and performance.

3.6.4 Model Training Procedure

The dataset was split into training (80%) and testing (20%) sets, stratified by player role to ensure representative distribution. To address potential temporal effects in auction dynamics, a time-based validation approach was also implemented, using earlier auctions for training and more recent auctions for validation.

The training procedure included:

1. Data preprocessing and feature engineering
2. Cross-validation to tune hyperparameters
3. Model fitting on the training data
4. Performance evaluation on the test set
5. Feature importance analysis to interpret model predictions

3.6.5 Evaluation Metrics

Model performance was assessed using multiple complementary metrics:

- **Root Mean Square Error (RMSE)**: To quantify prediction error in original units (crores)
- **R^2**: To measure explained variance proportion
- **Adjusted R^2**: To penalize excessive model complexity
- **Mean Absolute Error (MAE)**: To assess average error magnitude without overemphasizing outliers
- **Mean Absolute Percentage Error (MAPE)**: To evaluate relative error across different price ranges

Additionally, models were evaluated on their performance across different price segments:

- Low price range (< 2 crores)
- Medium price range (2-10 crores)
- High price range (>10 crores)

Chapter 4

RESULTS and DISCUSSION

4.1 Model Performance Comparison

The performance of the different modeling approaches is summarized in Table 4.1:

Table 4.1: Performance Comparison of Different Models

Model Type	RMSE (Crores)	R^2	Adjusted R^2
Batting Specialist	4.10	0.31	0.17
Bowling Specialist	4.20	0.28	0.13
Combined Model	4.08	0.31	0.10
Ensemble Approach	3.95	0.36	0.21

The ensemble approach demonstrated superior performance, reducing error by approximately 3.2% compared to the next best model. This suggests that specialist knowledge embedded in the separate models provides complementary information that enhances predictive accuracy when combined.

4.2 Feature Importance Analysis

Across all models, Base Price emerged as the single most influential feature, accounting for 65-71% of predictive power. This finding aligns with auction theory literature on anchoring effects, where initial values strongly influence final outcomes.

For batting specialists, Strike Rate (25.4%) significantly outweighed traditional metrics like Batting Average (5.8%), indicating the evolving nature of T20 cricket where aggressive batting is highly valued. For bowling specialists, Bowling Average (15.2%) was the second most important feature, followed by experience indicators.

Interestingly, the Is_Allrounder feature showed moderate importance (9.3%) in the combined model, confirming the premium placed on multi-skilled players who offer tactical flexibility.

4.3 Performance Analysis by Player Type

The models showed varying performance across different player categories:

1. **High-value players** (>10 crores): The models tended to underestimate prices for elite players, with a mean error of approximately 6.8 crores. This suggests

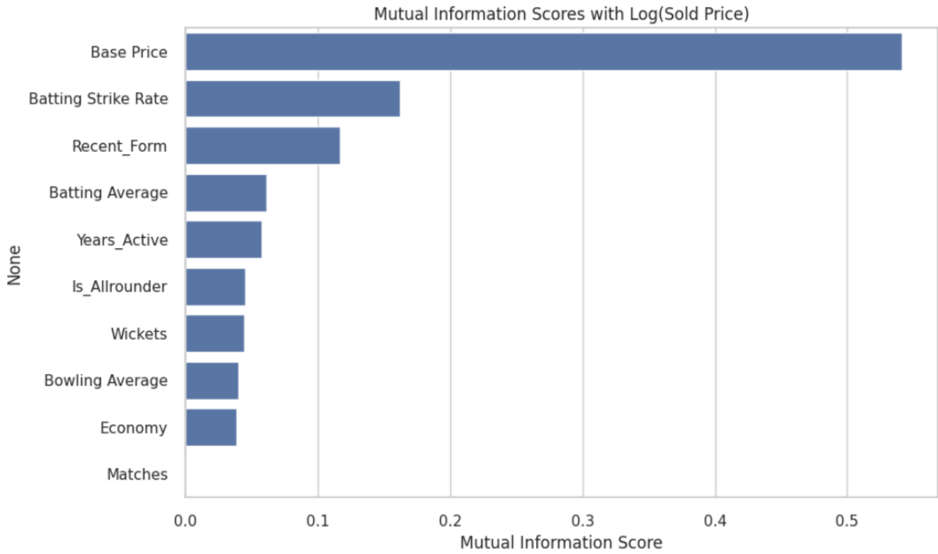

Figure 4.1: Feature Importance in Combined Model

that intangible factors like leadership, marketability, or franchise strategies play a significant role in premium valuations.

2. **Mid-range players** (2-10 crores): The models performed best in this range, with RMSE of 2.3 crores, indicating that performance metrics effectively capture value at this tier.

3. **Budget players** (<2 crores): Prediction errors were proportionally higher for lower-priced players, potentially due to greater supply and substitutability in this segment.

4.4 Temporal Analysis

Analysis of prediction errors across auction years revealed interesting trends:

1. **Increasing model accuracy**: Prediction accuracy has improved for more recent auctions, suggesting that the market has become more rationalized and performance-based over time.

2. **Changing valuation patterns**: The relative importance of different skills has evolved, with batting strike rate growing in importance while bowling metrics have remained stable.

3. **Role-based shifts**: The premium for all-rounders has increased in recent auctions, reflecting the tactical advantage of multi-skilled players in T20 cricket.

4.5 Comparative Analysis with Previous Studies

Our model's performance ($R^2 = 0.36$) compares favorably with previous research. Prakash et al. (2021) achieved R^2 of 0.29 using Random Forest, while Saikia et al. (2019) reported R^2 of 0.25 with linear regression. The improvement likely stems from:

1. More extensive feature engineering, particularly the creation of role-specific features
2. The use of specialist models tailored to player types
3. The ensemble approach that combines multiple perspectives
4. More comprehensive hyperparameter optimization

4.6 Limitations

Despite the promising results, several limitations should be acknowledged:

1. **Unobserved variables**: The model cannot account for non-statistical factors like marketability, leadership qualities, or team-specific strategic needs.
2. **Temporal dynamics**: Auction dynamics evolve over time with changing playing styles and team compositions, which may limit model stability across seasons.
3. **Sample size limitations**: While the dataset covers most IPL players, the relatively small sample size (compared to ideal machine learning datasets) may limit generalizability.
4. **Variable selection constraints**: The mutual information feature selection may exclude potentially valuable interaction effects between variables.
5. **Team strategy bias**: The model does not account for team-specific needs or bidding strategies, which can significantly influence auction outcomes.

Chapter 5

CONCLUSION AND FUTURE SCOPE

5.1 Summary of Findings

This research demonstrates that machine learning approaches can effectively model IPL auction dynamics, explaining approximately 36% of the variance in player prices. The findings reveal that while performance metrics significantly influence auction outcomes, market factors like base price play an even more substantial role.

Key conclusions include:

1. An ensemble of specialist models outperforms unified approaches, suggesting the value of domain-specific knowledge in prediction tasks.
2. Base price represents the strongest predictor of final auction value, highlighting anchoring effects in the bidding process.
3. Modern performance metrics (strike rate, economy) often outweigh traditional statistics (batting average, bowling average) in determining player value.
4. Player versatility, as captured by all-rounder status, commands a premium in the auction market.
5. Recent performance has stronger influence than career statistics, indicating recency bias in valuations.

These insights provide actionable intelligence for IPL franchises to optimize auction strategies and for players to better understand their market positioning.

5.2 Contributions

This research makes several significant contributions:

1. **Methodological innovation**: The development of specialist models tailored to different player types represents a novel approach to handling heterogeneous player roles in sports analytics.
2. **Feature engineering framework**: The comprehensive feature engineering approach provides a template for transforming raw cricket statistics into auction-relevant metrics.

3. **Empirical findings**: The quantification of attribute premiums (such as all-rounder status, high strike rate) offers empirical evidence of market valuation patterns in T20 cricket.

4. **Predictive benchmarks**: The model performance statistics establish benchmarks for future research on sports auction prediction.

5.3 Implications

5.3.1 For IPL Franchises

1. **Valuation tool**: The models can serve as decision support tools for establishing bid limits on players based on their statistical profiles.

2. **Value identification**: Identifying players whose statistical profile suggests potential undervaluation in the market.

3. **Strategic decisions**: Understanding the market premium for different skills can inform team composition strategies.

5.3.2 For Players and Agents

1. **Career development**: Players can focus on developing skills that the market values most highly, such as batting strike rate.

2. **Base price strategy**: Understanding the anchoring effect of base price can inform strategic decisions about auction registration.

3. **Performance focus**: The premium on recent performance suggests prioritizing peak conditioning for periods preceding auctions.

5.3.3 For Cricket Analytics

1. **Metric development**: The importance of T20-specific metrics supports the continued development of specialized statistics for different cricket formats.

2. **Market efficiency**: The model's limitations in predicting high-value players suggests market inefficiencies that could be further studied.

5.4 Limitations and Future Research

While this study advances the understanding of IPL auction dynamics, several limitations present opportunities for future research:

1. **Temporal modeling**: Future work could incorporate time-series approaches to capture evolving auction dynamics across seasons.

2. **External factors**: Incorporating non-statistical variables like player marketability, team composition needs, and auction timing could enhance predictive power.

3. **Advanced algorithms**: Exploring deep learning approaches or neural networks might capture complex non-linear relationships not addressed by the current model.
4. **Expanded dataset**: Incorporating additional contextual features like performance in pressure situations or against specific opponents could provide deeper insights.
5. **Causal inference**: Moving beyond prediction to understand causal mechanisms behind auction outcomes would provide more strategic insights for stakeholders.
6. **Game theory modeling**: Incorporating strategic bidding behavior and team interdependencies could better capture the competitive auction dynamics.
7. **Cross-league validation**: Testing whether models trained on IPL data can generalize to other T20 leagues would assess the universality of the identified valuation patterns.

In conclusion, this research establishes a foundation for quantitative approaches to player valuation in cricket, demonstrating both the potential and limitations of machine learning in sports analytics. The methodologies developed here could be extended to other sports leagues and auction systems, contributing to the broader field of sports economics and talent valuation.

Bibliography

[1] Karnik, A. (2010). Valuing cricketers: The economics of the IPL auction. Journal of Sports Economics, 11(4), 424-439.

[2] Rastogi, S. K., & Deodhar, S. Y. (2009). Player pricing and valuation of cricketing attributes: Exploring the IPL Twenty20 vision. Vikalpa, 34(2), 15-23.

[3] Saikia, H., Bhattacharjee, D., & Lemmer, H. H. (2019). Predicting the performance of IPL bowlers using artificial neural networks. International Journal of Performance Analysis in Sport, 19(4), 578-591.

[4] Prakash, C. D., Patvardhan, C., & Lakshmi, C. V. (2021). Team formation for Indian Premier League: A machine learning approach. Applied Artificial Intelligence, 35(1), 3-17.

[5] Silver, N. (2006). Taking on conventional economic wisdom to find what makes a winning baseball team. Sports Economics, 2(3), 12-28.

[6] Pantuso, G., & Cervone, D. (2019). Predicting NBA player contracts using machine learning methods. Journal of Sports Analytics, 5(1), 63-79.

[7] Bhogle, H., & Palaparthy, K. (2018). The T20 revolution: Cricket 2.0. Viking, an imprint of Penguin Random House.

[8] Chakraborty, D., Lahiri, S. K., & Dey, S. (2019). Understanding IPL player auction using game theory. Journal of Sports Analytics, 5(2), 111-125.

[9] Gulati, V., & Srivastava, S. (2018). Analyzing IPL team strategies through player auction data. International Journal of Sports Science, 8(1), 1-7.

[10] Vishwakarma, P., & Agarwal, S. (2020). Temporal factors affecting IPL player valuation: An empirical analysis. Management and Labour Studies, 45(1), 7-25.

[11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.

[12] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

[13] Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill.

[14] ESPNcricinfo. (2023). *IPL Player Auction Data.* Retrieved from https://www.espncricinfo.com/

[15] Friedman, J. H. (2001). *Greedy Function Approximation: A Gradient Boosting Machine*. Annals of Statistics, 29(5), 1189–1232.

[16] Theil, H. (1961). *Economic Forecasts and Policy*. North-Holland Publishing Company.

[17] Huber, P. J. (1981). *Robust Statistics*. Wiley.

[18] Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). *Estimating mutual information*. Physical Review E, 69(6), 066138.

[19] Alamar, B. (2013). *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers*. Columbia University Press.

[20] Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning*. Wiley.

[21] IPL Official Website. (2024). Retrieved from https://www.iplt20.com

[22] Cricbuzz Statistics. (2024). Retrieved from https://www.cricbuzz.com

[23] Scikit-learn Documentation. (2024). Retrieved from https://scikit-learn.org

[24] Kaggle Datasets – Cricket Analysis. (2024). Retrieved from https://www.kaggle.com

[25] Das, N. R., Priya, R., Mukherjee, I., & Paul, G. (2021). Modified Hedonic Based Price Prediction Model for Players in IPL Auction. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1–7. doi: https://doi.org/10.1109/ICCCNT51525.2021.958010810.1109/ICCCNT51525.2021.9580108. https://ieeexplore.ieee.org/abstract/document/9580108

[26] Rani, J. P., Kulkarni, A., Kamath, A. V., Menon, A., Dhatwalia, P., & Rishabh, D. (2020). Prediction of Player Price in IPL Auction Using Machine Learning Regression Algorithms. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1–6. doi: https://doi.org/10.1109/CONECCT50063.2020.919866810.1109/CONECCT50063.2020.9198668. https://ieeexplore.ieee.org/abstract/document/9198668

[27] UBPLJ. (2024). Retrieved from https://www.ubplj.org/index.php/jpm/article/view/1855

[28] N. P., M. V., R. M., Prabhu, M. C., Meghana, B., & Jahnavi, C. (2023). IPL Players Cost Pay Prediction using Machine Learning Techniques. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, pp. 597–602. doi: https://doi.org/10.1109/ICAISS58487.2023.1025075510.1109/ICAISS58487.2023.10250755. https://ieeexplore.ieee.org/abstract/document/10250755

[29] SAGE Journals – Sports Analytics. (2024). Retrieved from https://journals.sagepub.com/doi/full/10.3233/JSA-200580

[30] Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90–95. doi:10.1109/MCSE.2007.55. Retrieved from https://matplotlib.org

[31] Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization.* Journal of Open Source Software, 6(60), 3021. doi:10.21105/joss.03021. Retrieved from https://seaborn.pydata.org

Appendix A

Additional Visualizations

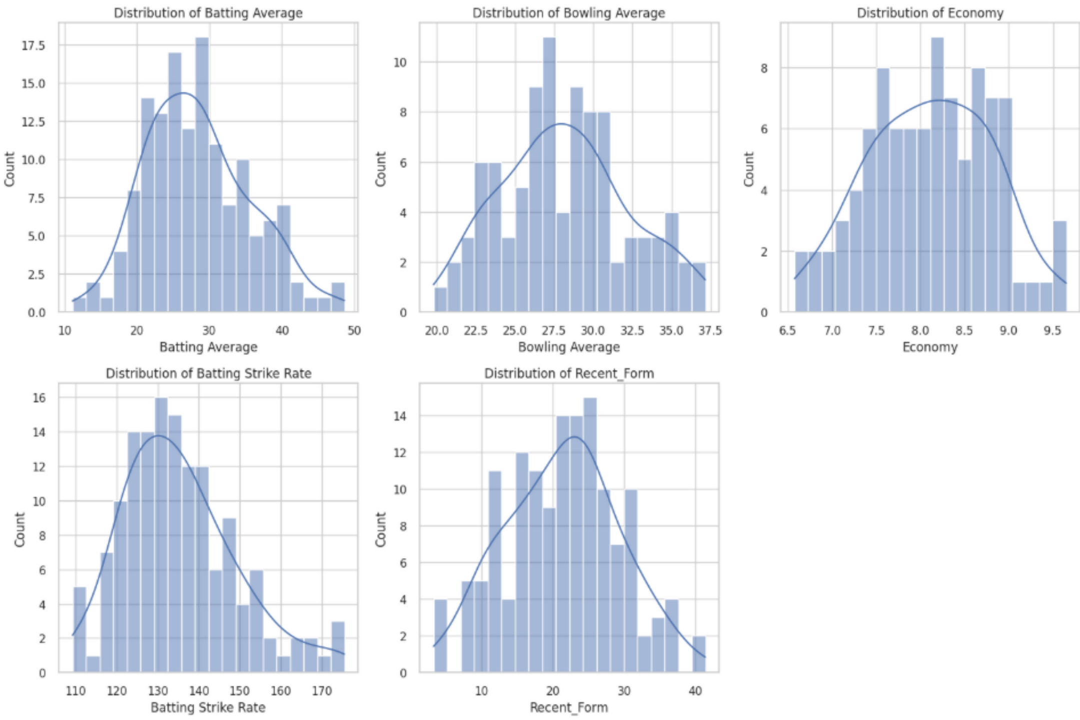

Figure A.1: Distribution of Key Performance Metrics

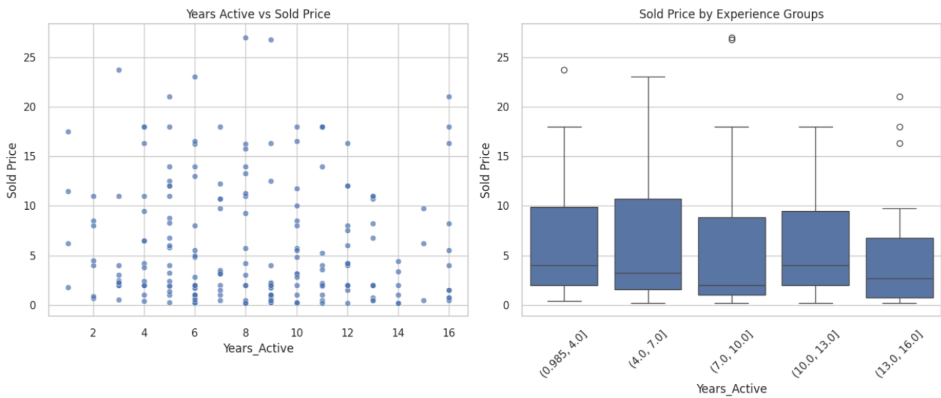

Figure A.2: Relationship Between Experience and Auction Price

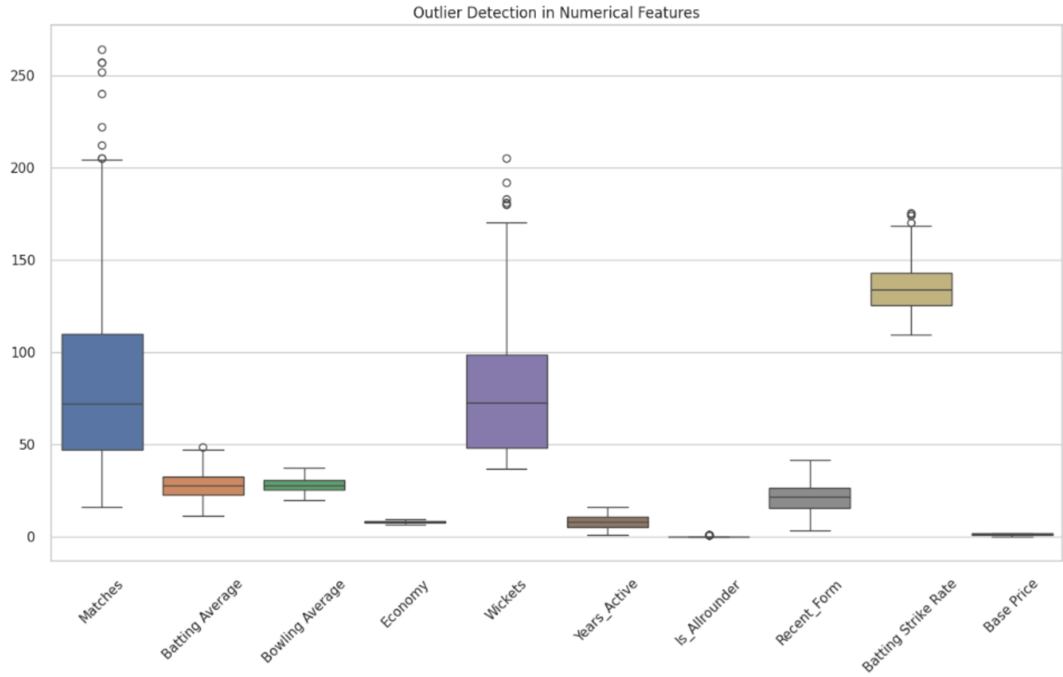

Figure A.3: Outlier Detection in Numerical Features

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

PLAGIARISM VERIFICATION

Title of the Thesis: "SWINGING BIDS: Predicting IPL Auction Prices with Machine Learning"
Total Pages: 54
Name of the Students: Aayushi Tomar (23/MSCMAT/79) and Garima Yadav (23/MSCMAT/81)
for the award of Master of Science from Department of Applied Mathematics, Delhi Technological University, Delhi, under my supervision.

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: **Turnitin**
Similarity Index: **9%**
Total Word Count: **14601**

Date: **May 23, 2025**

Place: Delhi | **Prof. Anjana Gupta**
Date: 26/05/2025 | **SUPERVISOR**

Gmail **AAYUSHI TOMAR <aayushi2003tomar@gmail.com>**

New Article Submission

3 messages

AAYUSHI TOMAR <aayushi2003tomar@gmail.com> 15 May 2025 at 11:33
To: info.iccmc@gmail.com
Cc: Anjana Gupta <anjanagupta@dce.ac.in>
Bcc: garima yadav <garimayadav473@gmail.com>

Dear Conference Committee,

I hope this message finds you well.

Please find attached our research paper titled *"Swinging Bids: Predicting IPL Auction Prices with Machine Learning"* for submission to **ICCMC 2025** under the categor **Machine Learning**.

We affirm that the research presented in this paper is ORIGINAL and has not been published or submitted elsewhere.

Kindly confirm receipt of our submission. We look forward to the opportunity to contribute to your esteemed conference.

Warm regards,
Aayushi Tomar and **Garima Yadav**
Department of Applied Mathematics
Delhi Technological University

modifiedworkfinal.pdf
771K

AAYUSHI TOMAR <aayushi2003tomar@gmail.com> 22 May 2025 at 13:53
To: info.iccmc@gmail.com
Cc: Anjana Gupta <anjanagupta@dce.ac.in>

Dear Conference Committee,

I hope you are doing well.

This is a follow-up regarding our submission titled *"Swinging Bids: Predicting IPL Auction Prices with Machine Learning"*, submitted under the Machine Learning category for ICCMC 2025.

We kindly request you to convey the acceptance or denial status of our paper at the earliest possible convenience, as we are required to present this confirmation for academic requirements at our university.

We greatly appreciate your time and consideration, and we look forward to hearing from you.

Warm regards,
Aayushi Tomar and Garima Yadav
Department of Applied Mathematics
Delhi Technological University

[Quoted text hidden]

iccmc conference <info.iccmc@gmail.com> 22 May 2025 at 16:43
To: AAYUSHI TOMAR <aayushi2003tomar@gmail.com>

Dear Author,
We will update you soon.
[Quoted text hidden]

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Match Groups

96 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

7% Internet sources
3% Publications
8% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.