5-2024

Teaming Strategy Optimization: An Analysis Of NBA Statistics, Shot Charts, And Constraints

Shaykh Siddique

TEAMING STRATEGY OPTIMIZATION: AN ANALYSIS OF NBA STATISTICS, SHOT CHARTS, AND CONSTRAINTS

A Thesis

by

SHAYKH SIDDIQUE

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2024

Major Subject: Computer Science

TEAMING STRATEGY OPTIMIZATION: AN ANALYSIS OF NBA STATISTICS, SHOT CHARTS, AND CONSTRAINTS

A Thesis

by

SHAYKH SIDDIQUE

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Lin Li Chair of Committee	Yonghui Wang Co-Chair of Committee
Sherri S. Frizell Committee Member	Na Li Committee Member
Yonggao Yang Department Head	Pamela H. Obiomon Dean, Roy G. Perry College of Engineering
Tyrone Tanner Dean, Graduate Studies	

May 2024

Major Subject: Computer Science

ABSTRACT

Teaming Strategy Optimization: An Analysis of NBA Statistics, Shot Charts, and Constraints

(May 2024)

Shaykh Siddique, B.S., East West University, Bangladesh Chair of

Advisory Committee: Dr. Lin Li

Co-Chair of Advisory Committee: Dr. Yonghui Wang

In the dynamic realm of NBA team management, balancing the intricacies of player performance and strategic signings present formidable challenges. Negotiating salary caps, player roles, on-court minutes, and contract durations requires a nuanced approach. This study comprehensively evaluated player efficiency and team performance, scrutinizing key statistical indicators like Points, Rebounds, Assists, Blocks, PER, RPM, Shot Charts, and others. Leveraging machine learning algorithms, including logistic, ridge, and lasso regressions, facilitated modeling the intricate relationship between player performance and team winning rates. Based on that, incorporating practical constraints yielded diverse and effective teaming strategies. Analyzing NBA player and game statistics from 2012 to 2022, the experimental findings underscore the accuracy of prediction models and the success of player selection strategies. This research provides actionable insights for NBA franchises seeking to streamline team operations and achieve triumphs on the court.

Index Terms: Court coverage, machine learning, NBA, performance evaluation, shot charts, sports analysis, teaming strategy.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Lin Li for his constant support and endless patience in completing my master's thesis. I would not have been able to complete my thesis without his guidance in each step from beginning to end. I would also like to thank my co-supervisor Dr. Yonghui Wang and committee members Dr. Sherri S. Frizell and Dr. Na Li for their valuable time and efforts in perfecting my thesis through suggestions and feedback. Besides, I would like to thank my professors in the Department of Computer Science of PVAMU who taught me over the past two years. Their knowledge, caring of students, and devotion to higher education will keep impacting me through my career pursuit.

Last but not least, I would like to acknowledge the unconditional love and support of my parents, who live overseas. I dedicate this thesis to my parents.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER 1

INTRODUCTION

The National Basketball Association (NBA) has made significant contributions to the worlds of sports, entertainment, and the sports marketplace, establishing itself as a key influencer in commercial sports. The NBA declared the salary cap for the 2023-24 season at $136.021 million [1]. This salary cap represented a noticeable increase compared to the previous season when the 2022-23 salary cap was fixed at $123.655 million. The rise in the salary cap reflected the NBA's adaptability in response to the ever-evolving financial dynamics of professional basketball. It is important to note that the salary cap serves as a crucial parameter, defining the upper limit for teams regarding player salary expenses for the upcoming season. Golden State Warriors Stephen Curry achieved a historic milestone to become the first player in NBA history to earn $50 million annually [2]. This remarkable salary marked a significant turning point in the league's compensation landscape, showcasing top-tier talent's growing value and recognition within the NBA. Securing potential players within a limited budget is always challenging, which requires planning, players' statistics analysis, and an understanding of NBA market dynamics.

1.1 Sports Statistics and Machine Learning

In this age of the digital revolution, a vast amount of sports data is generated by various advanced technologies. Machine learning techniques, powered by advanced algorithms and data analytics, are revolutionizing how coaches strategize, select players, train, and manage

---

This thesis follows the style of IEEE

teams. Dimitrije Cabarkapa examined NBA game-related statistics in both regular and post-season periods spanning 2016 to 2019. The research identified significant tactical shifts in the NBA's transition to a more conservative style during the post-season [3], characterized by reduced field goal attempts, assists, steals, turnovers, and points scored. Discriminant function analysis helps to tell when a team wins or loses. It shows that making more shots, getting more rebounds on defense, and shooting efficiencies are essential in the regular season and playoffs.

Xuanhao Feng's research focused on NBA player salaries, using data from 331 players in the 2020-2021 season to develop multiple linear regression models for salary analysis. The study employed various regression techniques and identified three-point shooting proficiency [4], defensive abilities, and teamwork skills that influence player compensation in line with the modern NBA playing style. The research's findings aimed to enhance the NBA's commercial value and offer valuable insights for league and team management, fostering growth and development.

Luca Grassetti introduced a novel model-based approach to assess the effectiveness of five-person basketball lineups [5], departing from traditional adjusted plus-minus methods. Grassetti's innovation included replacing the conventional response variable with a multifaceted score derived from box score statistics and shifting the focus from individual players to evaluating entire lineups. This research provides a valuable tool for ranking players and lineups using Bayesian estimation, with applications in lineup management and real-time performance monitoring.

1.2 Research Objectives

The aim of this study was to analyze the NBA players' statistics comprehensively and maximize the NBA team performance by using effective NBA player selection.

The contributions can be defined more precisely as:

- Develop a machine learning algorithm for predicting the NBA team's winning rate and generate weights in a linear programming model.
- Utilize the weights generated by the machine learning model to develop linear programming models for optimized team player selection.
- Integrate court coverage and Non-Negative Matrix Factorization into the linear programming model for selecting optimized team players.
- Conduct a comparative analysis between static algorithms (that is, Pareto optimization) considering other relevant studies, and evaluate their effectiveness against our proposed models for optimized team selection in the NBA.

1.3 Programming Languages and Tools

Python was the primary programming language in this study, providing a robust and versatile data analysis and modeling platform. Web scraping techniques were employed for efficient data collection, enabling the acquisition of comprehensive and up-to-date NBA player statistics and relevant team performance metrics. Several essential Python packages were utilized to facilitate various stages of the research process. The scikit-learn library was leveraged for implementing machine learning algorithms, allowing for in-depth statistical analysis and predictive modeling of player performance based on historical data. Additionally, the Pulp library was utilized for linear programming tasks, enabling the optimization of salary cap allocation and player recruitment strategies. Furthermore, various plotting packages within the Python ecosystem were employed to visualize and interpret the findings, enhancing the clarity and effectiveness of the research outcomes.

1.4 Thesis Outline

The paper is organized into six main sections, providing a structured framework for presenting the research on sports statistics and machine learning. In Chapter 1 Introduction, the significance of sports statistics in the context of machine learning is outlined, along with the research objectives, choice of programming languages, and tools. Following this, the background and related work of Chapter 2 delves into the exploration of player statistics, including traditional metrics and advanced measures such as Defensive Rating (DEFRTG), Player Efficiency Rating (PER), Real Plus- Minus (RPM), and Sweet Spots. The third section, methodology, details the statistical analysis of team performance, winning rate prediction through various regression methods, and team optimization and player selection strategies. The subsequent section, experiment, analysis, and model evaluation, utilized datasets to predict winning rates based on different attributes, evaluated the trained regression models, and analyzed team optimization and player selection strategies. Chapter 5 delves into team optimization with court coverage, incorporating player shot charts, Gaussian smoothing, non-negative matrix factorization, and experimentation with base vectors. Finally, the paper concludes with a summary of findings and suggestions for future work in Chapter 6.

CHAPTER 2

BACKGROUND AND RELATED WORK

In professional basketball, classical player statistics like points (PTS), rebounds (REB), and assists (AST) have been essential for assessing player capabilities and showcasing offensive and defensive skills. Advanced metrics such as Player Efficiency Rating (PER), Real Plus-Minus (RPM), and Defensive Rating (DEF) provide comprehensive assessments of player impact on team performance. PER, created by John Hollinger, evaluates overall efficiency, while RPM measures a player's influence on both offense and defense. Defensive rating tracks points allowed per 100 possessions to gauge defensive contributions. The concept of sweet spots emphasizes specific court areas where players excel, highlighting spatial analysis for optimizing team strategies. Understanding these metrics is crucial for comprehensive player evaluation and effective team optimization in the evolving landscape of the NBA.

2.1 Exploring Player Statistics

2.1.1 Traditional Statistics

Having long functioned as fundamental tools for evaluating player performance and their contribution to team success, classical player statistics have played an indispensable role in the National Basketball Association context. Table 2.1 presents a list of metrics that are frequently used to measure a player's performance. Points per game (PPG), rebounds per game (RPG), and assists per game (APG) have been pivotal in assessing the offensive and defensive capabilities of players, providing valuable insights into their scoring ability, rebounding proficiency, and playmaking skills [6]. Additionally, advanced metrics like assists (AST), steals (STL), blocks (BLK), field goals made (FGM), field goals attempted (FGA), three-pointers made (3PM), three-pointers attempted (3PA), free throws made (FTM), turnovers (TO), total rebounds (REB), points scored (PTS), games played (GP), minutes played (MIN), pace (PACE), and games started (GS) have gained prominence for their comprehensive evaluation of player impact on team performance [7]. These classical and advanced statistics have become essential for analyzing player proficiency and impact within the dynamic context of NBA gameplay by offering tangible measures of player effectiveness on the court.

TABLE 2.1
STATISTIC NOTATIONS AND DEFINITIONS

Notation	Definition
AST	Assists
STL	Steals
BLK	Blocks
FGM	Field Goal Made
FGA	Free Throw Attempted
3PM	Three-point Field Goal Made
3PA	Three-point Field Goal Attempted
TO	Turnovers
PF	Personal Fouls
OREB	Offense Rebounds
DREB	Defense Rebounds
REB	Total Rebounds
PTS	Points
GP	Games Played
GS	Game Starter
Min	Minutes Played
PACE	Possessions in a game

2.1.2 Comprehensive One-Number Stats

2.1.2.1 Defensive Rating (DEFRTG)

NBA players' Defensive rating likely refers to a statistical metric used to assess the defensive performance of NBA teams in individual games. Defensive rating is a common advanced statistic in basketball that quantifies a team's defensive effectiveness by measuring the number of points they allow per 100 possessions. It provides insight into a team's ability to prevent their opponents from scoring. A high defensive rating indicates a less effective defense, while a lower rating signifies a more robust defensive performance. This statistic is valuable for evaluating a team's defensive prowess and comparing different teams in the NBA.

$$\text{DEFRTG} = 100 \times \frac{\text{Opp Points}}{\text{Opp Poss}} \tag{2.1}$$

Equation 2.1 is the formula for calculating a player's defensive rating. Where *OppPoints* is the total number of points scored by the opposing team, and *OppPoss* is the total number of possessions the team's opponents had during the game.

2.1.2.2 Player Efficiency Rating (PER)

The NBA Player Efficiency Rating (PER) is a statistic that John Hollinger [6] developed to provide a single number to summarize a player's overall performance in a basketball game. PER considers a player's positive contributions, such as points, assists, rebounds, steals, and blocks, and their negative contributions, such as turnovers and missed field goals [8]. It is a popular advanced statistic used to evaluate and compare the efficiency and impact of NBA players.

$$factor = \frac{2}{3} - (0.50 \times \frac{l_g AST}{l_g FGM}) \div (2 \times \frac{l_g FG}{l_g FTM}) \tag{2.2}$$

$$VOP = \frac{lgPTS}{lgFGA - lgORB + lgTO + 0.44 \times lgFTA} \tag{2.3}$$

$$DRBP = \frac{lgTRB - lgORB}{lgTRB} \tag{2.4}$$

$$\begin{aligned} uPER = \frac{1}{\text{min}} \times \Bigg(3P - \frac{PF \times \text{lgFT}}{\text{lgPF}} + \left[\frac{FT}{2} \times \left(2 - \frac{\text{tmAST}}{3 \times \text{tmFG}}\right)\right] \\ + \left[FG \times \left(2 - \frac{\text{factor} \times \text{tmAST}}{\text{tmFG}}\right)\right] + \frac{2 \times \text{AST}}{3} \\ + \text{VOP} \times \Bigg[\text{DRBP} \times \Bigg(2 \times \text{ORB} + \text{BLK} - 0.2464 \times [FTA - FT] \\ - [FGA - FG] - \text{TRB}] + \frac{0.44 \times \text{lgFTA} \times \text{PF}}{\text{lgPF}} - (TO + \text{ORB}) + \text{STL} \\ + \text{TRB} - 0.1936(FTA - FT)\Bigg]\Bigg) \end{aligned} \tag{2.5}$$

In equation 2.5, the notation tm denotes a team prefix, while lg represents a league prefix, distinguishing them from player-specific indicators. Key metrics include min for minutes played, 3P for three-point field goals, FG for field goals, and FT for free throws. Additionally, VOP quantifies possession value, specifically within the league context. Rebound statistics comprise RB for overall rebounds, divided into ORB for offensive rebounds, DRB for defensive rebounds, and TRB for the combined total of offensive and defensive rebounds. Finally, RBP denotes the percentage of offensive or defensive rebounds.

2.1.2.3 Real Plus-Minus (RPM)

Real Plus-Minus (RPM) is an advanced NBA player evaluation metric developed by Jeremias Engelmann [9] and presented by ESPN. It employs a proprietary formula and utilizes an extensive dataset to isolate a player's unique on-court impact while considering the influence of teammates and opponents. RPM assesses a player's contribution by estimating how many points they add or subtract per 100 possessions, offering separate ratings for offensive (ORPM) and defensive (DRPM) impact. Although the exact formula is undisclosed, RPM is recognized for providing a comprehensive view of a player's value and impact in the NBA. To access RPM ratings, individuals often turn to sources like ESPN, which regularly publishes and updates these ratings.

2.1.3 Shot Charts and Sweet Spots

In professional basketball, a player's sweet spot refers to a specific court location (X, Y) where they demonstrate high shooting efficiency and comfort. These areas, characterized by optimal angles and familiar positioning, are crucial for determining a player's offensive capabilities and strategic significance within the team. Utilizing these sweet spots effectively allows players to maximize their scoring potential and contribute to the team's offensive strategy. Analyzing these preferred shooting areas provides valuable insights into player performance dynamics, enabling coaches and analysts to design gameplay strategies based on individual players' strengths and shooting preferences.

2.2 Related Works

Prior studies in analyzing the performance of NBA players have emphasized the importance of statistical measures and sophisticated analytics in assessing how players contribute to the overall success of their teams. Wong-Toi's research employed a pioneering Bayesian nonparametric mixture model [10] to classify players into distinct shooting behavior clusters across twelve front-court regions. The analysis of NBA data from the 2018–2019 regular season and the 2019–2020 bubble season revealed 13 player clusters and highlights their diversity. The study's implications for strategic planning, shot selection, and recruitment make it an invaluable resource for basketball coaches, analysts, and scholars. Guanyu Hu introduced a zero-inflated Poisson model [11] with clustered regression coefficients for analyzing shooting patterns in basketball, emphasizing the significance of static information in fast-paced sports. The model effectively captures player shooting habits and court-specific tendencies, accounting for regions with a high prevalence of zero field goal attempts. The empirical validation using NBA data from the 2017–2018 regular season demonstrates its practical utility and offers valuable insights for players, coaches, and team managers.

Ramya Nagarajan addressed the challenge of optimizing player selection and team performance [12] in the NBA, emphasizing the importance of maximizing efficiency within the salary cap constraints. The study identified 14 key statistics, including Points, Rebounds, Assists, Blocks, Steals, and Defense Rating, to construct a robust model for predicting team winning rates. Through rigorous testing, player selection strategies are presented, tailored to specific objectives and constraints, demonstrating the practical value of the proposed approach in the context of NBA contract negotiations and team management. Stephanie A. Kovalchik highlighted the growing availability of player tracking data [13], which has led to significant advancements in sports analytics. These data enable precise player performance evaluation, including metrics like distance traveled and expected points for specific plays or actions. The review emphasized the role of statistical and machine learning techniques in leveraging player tracking data and its impact on diverse domains. It also acknowledged ongoing methodological challenges in the field, providing a comprehensive view of the evolving landscape of sports analytics. Cem Osken identified the challenge of predicting basketball game winners using a unique approach based on player clustering and prototype heuristics. Their model [14] achieved a remarkable 76% accuracy over five NBA seasons and maintains 71% accuracy on unseen seasons, outperforming human experts. This innovative method highlighted the effectiveness of using player stereotypes extracted from individual statistics for game outcome prediction, contributing significantly to the field.

Benjamin S. Baumer provided a comprehensive overview of sports analytics, focusing on [15] improving athletic performance through data analysis. It discussed four key concepts applicable across sports, emphasizing the role of statistical techniques and data sources. The paper highlighted the fusion of data, models, tools, and sports knowledge to yield actionable insights, making it a valuable resource for those interested in sports analytics. It refrained from detailed player evaluations but offered insights into the field's growth and critical principles.

B. Jay Coleman studied a crucial gap in sports analytics research extensively analyzing the scope and scale of published refereed articles in this field. While exploring 1146 articles from 140 journals across multiple disciplines [16], the study unveiled the size, nature, and critical contributors in sports analytics. By illuminating the growth and prominence of sports analytics within the academic literature, this work offered valuable insights into the parameters and trajectory of this burgeoning sub-discipline.

Tomislav Horvat presented a data-driven model employing machine learning techniques for predicting outcomes in NBA and other basketball league games. Building on a solid mathematical foundation for basketball statistics and performance indicators [17], the model integrated the extended team efficiency index, combining individual player efficiency metrics with a comparative element to reward teams for surpassing opponents in specific areas. They calculated symmetrically predicted indices for upcoming games by analyzing historical data, resulting in a win function for outcome prediction. Considering an optimal time window for training data, the model achieved an average prediction accuracy of approximately 66% and a maximum accuracy of around 78%, establishing its significance in sports analytics research.

CHAPTER 3

METHODOLOGY

The methodology for this research involved implementing two distinct techniques to optimize player selection and team composition, with a primary objective of maximizing the team's winning rate while minimizing the associated salary expenditure. The first approach was a static method, utilizing the skyline algorithm with dynamic programming principles to strategically choose players and construct an optimal team roster. This method facilitated the identification of players who offered the most significant contributions to the team's success while considering budgetary constraints. The second technique involved the integration of a machine learning model coupled with linear programming, allowing for a data-driven approach to player selection and team formation. By leveraging historical and real-time data, this model aided in predicting player performance and facilitated informed decision-making to balance performance optimization and financial efficiency. Utilizing both these methodologies enabled a comprehensive analysis of the intricate dynamics of selecting and managing a competitive NBA team.

3.1 Statistical Analysis of Team Performance

Evaluating team performance requires a comprehensive review of various statistical measures. These measures include assessing the points scored by each player, the frequency of their assists, rebounds, and steals, and their effectiveness in blocking opponent shots. The efficiency of players' shooting abilities, both near the hoop and from a distance, was gauged using metrics like field goal percentages and three-point shooting accuracy. Furthermore, it included the success rates of players making free throws, offering crucial insights into their overall performance. By scrutinizing these statistics meticulously, researchers can identify the strengths and weaknesses of a team, thereby aiding coaches and players in devising strategies to enhance their gameplay and achieve better outcomes on the court.

3.1.1 Player Types

A thorough examination of NBA team performance involves analyzing several essential elements. The configuration of each team, including the count of players and their specific positions, significantly influences the team's dynamics. During NBA games, a group of five players, called starters, assumes distinct roles like point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C), working together to score points and defend against the opposing team. Additionally, the presence of backup players provides strategic options for substitutions and added support during gameplay. NBA games consist of 48 minutes, divided into four quarters of 12 minutes each. There are other backup players to replace those five players. In this study, five players were used as starters and five as bench players, a total of ten players.

Table 3.1 offers a systematic categorization of NBA players, assigning them to ten distinctive groups based on their respective positions and whether they primarily function as starters (S) or bench players (B). The indexing scheme ranges from 1 to 10n, where n signifies the count of players within each group. This indexing system streamlines the pre-processing of the collected NBA data, enabling a structured and efficient approach to data organization and analysis for subsequent research and statistical evaluation.

TABLE 3.1
PLAYER INDEX, POSITION, WITH STARTER STATUS

Player Index	Position	S/B
1 to n	C	S
n + 1 to 2n	SF	S
2n + 1 to 3n	PF	S
3n + 1 to 4n	PG	S
4n + 1 to 5n	SG	S
5n + 1 to 6n	C	B
6n + 1 to 7n	SF	B
7n + 1 to 8n	PF	B
8n + 1 to 9n	PG	B
9n + 1 to 10n	SG	B

3.1.2 Players Attributes

Player statistics were used as features for the model predictions. For model design and evaluation, two sets of player attributes were used.

- Set of 15 Attributes mentioned in Table 3.2
- Set of 3 Attributes mentioned in Table 3.3

TABLE 3.2
ATTRIBUTE NOTATIONS OF 15 PARAMETERS

Attributes	Notation	Definition
A_1	3PA	Three-point Field Goal Attempted
A_2	3PM	Three-point Field Goal Made
A_3	AST	Assists
A_4	BLK	Blocks
A_5	DEF	Defensive Rating
A_6	DREB	Defensive Rebounds
A_7	FGA	Field Goals Attempted
A_8	FGM	Field Goals Made
A_9	FTA	Free Throws Attempted
A_{10}	FTM	Free Throws Made
A_{11}	OREB	Offensive Rebounds
A_{12}	PF	Personal Fouls
A_{13}	STL	Steals
A_{14}	TO	Turnovers
A_{15}	RPM	Real Plus-Minus

TABLE 3.3
ATTRIBUTE NOTATIONS OF THREE PARAMETERS

Attributes	Notation	Definition
A_1	PER	Three-point Field Goal Attempted
A_2	DEF	Three-point Field Goal Made
A_3	RPM	Assists

3.1.3 Variables and Constants

Table 3.4 presents the key variables and constants essential for optimizing the selection of NBA players for a team. The $Salary_i$ variable signifies the salary of the specific player, while P_{ij} represents the statistical value of the player in a particular attribute. The S_i variable denotes the selection status of the player. It is a Boolean type whether a player is selected or not. The w_i coefficients were derived from regression analysis and held significant weight in the optimization model. Moreover, the SalaryCap constant established the upper limit for the team's budget for each NBA team. Lastly, the WinThreshold constant set the minimum benchmark for the team's winning rate, serving as a performance indicator. Understanding these variables and constants is crucial for effective player selection and team optimization in the NBA.

TABLE 3.4
VARIABLES AND CONSTANTS FOR OPTIMIZATION

Variables	Description	Type
$Salary_i$	Salary of the ith player	real
P_{ij}	ith player's value in statistics attribute j	real
S_i	Selection status of the ith player	bool
w_i	Weight coefficients learned from regression	real
t_i	Minutes played by the ith player	real
SalaryCap	Upper bound of team budget	real
WinThreshold	Minimum of team winning rate	real
TimeLimit	Upper bound of team's minutes played	real

3.2 Winning Rate Prediction

This study was characterized by a dual objective centered around predicting the winning rate and selecting optimal players for the team. The first aspect involved utilizing the teams' historical game statistics to forecast their future winning rates, allowing for a comprehensive understanding of their performance trajectory. This approach aimed to develop an accurate predictive model for estimating the teams' potential success rates by analyzing appropriate metrics and performance indicators from previous games. The second facet focused on selecting the most suitable players for the team, emphasizing the identification of individuals whose skill sets and contributions aligned with the team's strategic objectives. Integrating a meticulous evaluation of players' performance records and attributes aimed to optimize the team's overall composition, fostering enhanced gameplay and increased chances of success in future games.

3.2.1 Linear Regression

Linear regression is a fundamental statistical technique to model the relationship between a dependent variable and one or more independent variables.

$$Y = w_0 + w_1X_1 + w_2X_2 + ... + w_nX_n \tag{3.1}$$

From equation 3.1, Y is the dependent variable, X_1, X_2, ..., X_n are the independent variables, w_0, w_1, ..., w_n are the regression coefficients representing the parameters. By estimating the values of the coefficients, the model provides insights into the relationship and the impact of the independent variables on the dependent variable. It serves as a powerful tool in various research domains, aiding in exploring causal relationships and predicting outcomes based on the observed data. The simplicity and interpretability of the linear regression model make it a widely utilized technique for analyzing datasets and uncovering underlying patterns and trends within the data.

3.2.2 Lasso Regression

Lasso regression introduced by Tibshirani [18], is a valuable technique in statistical modeling and machine learning. It operates as a form of regression analysis that applies a shrinkage penalty to the regression coefficients, inducing some of the coefficients to be precisely zero. Doing so encourages sparsity and performs automatic feature selection, making it particularly useful for handling high-dimensional data and addressing overfitting problems. The Lasso method has found widespread application in various research domains due to its ability to manage multicollinearity and enhance the interpretability of models, providing researchers with an effective tool for navigating complex datasets and improving the robustness of their analyses.

3.2.3 Ridge Regression

Ridge regression [19] is a robust technique utilized in statistical modeling to address the issue of multicollinearity and stabilize the estimates in regression analysis. This method introduces an L2 regularization term to the ordinary least squares method, thereby constraining the magnitude of the coefficients. By adding a penalty term equivalent to the square of the coefficients, ridge regression effectively reduces the impact of irrelevant variables, ultimately preventing overfitting. Widely applicable in various research domains, ridge regression has proven to be an indispensable tool for improving the reliability and performance of predictive models, ensuring greater robustness and accuracy in data analysis.

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{3.2}$$

In equation 3.2, β^ridge ridge represents the ridge regression estimate, y_i is the observed response for the ith observation, β_0 is the intercept, β_j are the regression coefficients, x_{ij} is the value of the jth predictor for the ith observation, λ is the regularization parameter, and n is the number of observations in the dataset.

3.2.4 Logistic Regression

Logistic regression, a prominent statistical method introduced by David Cox [20], is widely employed for modeling the probability of a binary response based on one or more predictor variables. The logistic regression model transforms the linear regression equation using the logistic function to produce a sigmoidal curve, providing outputs in the range of [0,1] representing the probability of a binary outcome.

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}} \tag{3.3}$$

$$\text{Win\%} = 1/(1 + e^{-y}) \tag{3.4}$$

Equation 3.3 denotes the probability of the event Y = 1 given the predictor variables X, and β_0, β_1,..., β_p are the regression coefficients to be estimated. Logistic regression finds extensive application in various research domains, including epidemiology, healthcare. Equation 3.4 evaluates the winning rate for logistic regression.

3.3 Players Selection and Team Optimization

3.3.1 Skyline Problem and Dynamic Programming

3.3.1.1 Skyline Problem

The skyline problem involves determining the outline formed by the highest points of a collection of buildings or structures when viewed from a particular perspective, typically from the side. Given a set of buildings, each characterized by its position along a horizontal axis, width, and height, the goal is to compute the skyline silhouette, representing the contour created by the tallest points of these structures without any overlaps or obstructions. This problem arises in various fields, including urban planning, computer graphics, and geographical information systems, where understanding the visual profile of structures is crucial for visualization, design, and analysis purposes.

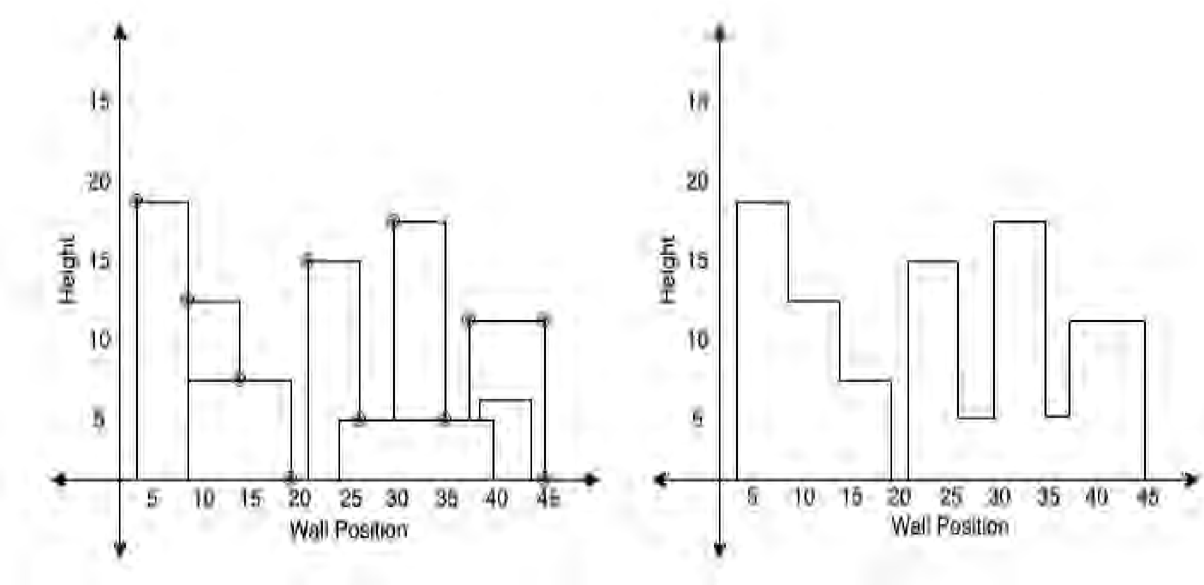


Fig. 3.1. Skyline Example.

Fig. 3.1 presents an example of the skyline problem. One common solution for the skyline problem involves using a sweep line algorithm. This algorithm iterates through the buildings while scanning from left to right along the x-axis. As it encounters each building, it updates a data structure, often a priority queue or a balanced binary search tree, to keep track of the heights of active buildings at that point.

3.3.1.2 Dynamic Programming for Player Selection

Algorithm 1 efficiently identifies an optimal player combination, considering financial and time constraints.

Algorithm 1: Dynamic Programming Approach with Skyline Dominates

```
Input: List of current selection players (cSelect), current budget
       (cBudget), current time (cTime)
Result: Best selection based on skyline dominance
function findBestCombination(cSelect, cBudget, cTime)
    if (tuple(cSelect), cBudget, cTime) is in memo then
        return memo[(tuple(cSelect), cBudget, cTime)]
    end
    for player in allplayers do
        if player["Time"] ≤ cTime and player["Salary"] ≤ cBudget
         then
            newSelect ← cSelect + [player];
            memo[(tuple(newSelect), cBudget, cTime)] ←
             findBestCombination( newSelect, cBudget -
             player["Salary"], cTime - player["Time"])
            if skylineDominates(newSelect, bestSelection) then
                bestSelection ← newSelect
            end
        end
    end
    return memo[(tuple(newSelect), cBudget, cTime)]
end
```

Through the recursive function *findBestCombination*, the algorithm systematically evaluated each player's compatibility with the available budget and time. Players attributes, such as salary and playing time, ensured that selected players fit within the allocated resources. The algorithm further employed memorization to store computed values, expedited decision-making, and avoidance of redundant computations. When determining the best player combination, the algorithm employed the skylineDominates function in Algorithm 2, which compared the ongoing player selection with the current best selection. For example, if the current selection was more efficient than the previous best selection, it updated the best selection to reflect the current one, ensuring the most dominant player combination is chosen.

Algorithm 2: Skyline Dominance Function

```
Input: Two sets of players, bestSelection and newSelect
Result: Boolean indicating whether newSelect skyline dominates
        bestSelection
function skylineDominates(newSelect, bestSelection)
  for each attribute a in newSelect[0] and bestSelection[0] do
    if max(newSelect[:, a]) ≥ max(bestSelection[:, a]) then
      return True;
    end
  end
  return False;
end
```

Players attributes, such as salary and playing time, ensured that selected players fit within the allocated resources. The algorithm further employs memoization to store computed values, expediting decision-making and avoiding redundant computations. When determining the best player combination, the algorithm employs the skylineDominates function, which compares the ongoing player selection with the current best selection. If the current selection is more efficient than the previous best selection, it updates the best selection to reflect the current one, ensuring the most dominant player combination is chosen.

Table 3.5 shows a dummy example of players to be selected. Assume each team can have three players, and each player is evaluated with two performance indicators PER and RPM. Given the salary information of each player, and with a salary cap constraint of 7, the objective is to select players to maximize team performance (skyline dominance).

TABLE 3.5
EXAMPLE PLAYERS DATA WITH BUDGET, PER, AND RPM

Player	Salary	PER	RPM
1	2	8	6
2	1	7	6
3	3	4	8
4	2	9	7
5	4	7	8

Equation 3.5 shows the player selection criteria used in the dynamic programming algorithm for player selection, where max refers to skylineDominates that maximize between two values. For i^{th} player, represented with i and j represents the budget, and w_i is the salary for the corresponding player. v_i is the attribute for i^{th} players. As there are multiple attributes, the tuple is used here.

$$F(Player_i, Budget_j) = \begin{cases} max\{F(i-1,\ j),\ v_i + F(i-1,\ j-w_i)\} & \text{if } j - w_i \geq 0 \\ F(i-1,\ j) & \text{if } j - w_i < 0 \end{cases} \tag{3.5}$$

Table 3.6 presents the player selection procedure following the dynamic programming algorithm based on the data presented in Table 3.5.

TABLE 3.6

EXAMPLE OF SOLVING AN INSTANCE BY DYNAMIC PROGRAMMING ALGORITHM

	Budget$_j$							
Player$_i$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	0	(8, 6)	(8, 6)	(8, 6)	(8, 6)	(8, 6)	(8, 6)
2	0	(7, 6)	(8, 6)	(15, 12)	(15, 12)	(15, 12)	(15, 12)	(15, 12)
3	0	(7, 6)	(8, 6)	(15, 12)	(15, 12)	(15, 12)	(19, 20)	(19, 20)
4	0	(7, 6)	(9, 7)	(16, 13)	(17, 13)	(24, 19)	(24, 19)	(24, 19)
5	0	(7, 6)	(9, 7)	(16, 13)	(17, 13)	(24, 19)	(24, 19)	(24, 19)

Table 3.7 presents a simple demonstration of utilizing the memory function with the hash table. In this example, two constraints, namely budget and time, were employed. Additionally, as measured by the PER attribute, player performance was considered for evaluation. The algorithm utilized a hash table to optimize the search for valid combinations, as it checked whether a new player, combined with previous selections, formed an already calculated combination. The hash table used player names as keys for efficient retrieval.

TABLE 3.7

SIMPLE EXAMPLE OF SELECTING PLAYERS AND MEMORY FUNCTION

Players Tuple	Budget	Time	PER
Nikola Jokic	31579390	33.5	11.78
Nikola Jokic, Jayson Tatum	59682890	69.4	20.74
Nikola Jokic, Jayson Tatum, Luka Doncic	69857281	104.8	27.24
Nikola Jokic, Jayson Tatum, Desmond Bane	61716050	99.2	26.17
Nikola Jokic, Jayson Tatum, Bobby Portis	64030490	97.6	23.08
.	.	.	.
Jayson Tatum, Luka Doncic, Bobby Portis	42625491	99.5	17.8

For example, there is a budget constraint of \$62 million and a time constraint of 100 minutes. For simplicity of understanding, three players will be selected, and that maximizes efficiency based on PER (Player Efficiency Rating). For five players- Nikola Jokic, Jayson Tatum, Luka Doncic, Desmond Bane, and Bobby Portis, the selection process involved checking skyline dominance with the dynamic programming algorithm.

- Initialization of bestSelection with a random player. Nikola Jokic is chosen.
- The current selection of players, Nikola Jokic, with the current budget of \$31579390, current time 33.5 min, and PER of 11.78.
- So, the remaining budget is $(62000000 - 31579390) = 30420610$ and the remaining time is $(100 - 33.5) = 66.5$ which are greater than zero. Nikola Jokic can be a valid selection.
- Iterating through all the players and choosing the next player, Jayson Tatum.
- Consider the players one by one, ensuring that their salary and playing time fit within the budget and time constraints.
- The current selection of players, Nikola Jokic and Jayson Tatum, with the current budget of \$59682890, current time 69.4 min, and PER of 20.74.
- The remaining budget is $(62000000 - 59682890) = 2317110$ and the remaining time is $(100 - 69.4) = 30.6$ which are greater than zero. Nikola Jokic and Jayson Tatum can be a valid selection.
- Continue iterating through all the players and choosing the next player, Luka Doncic.

- The current selection of players, Nikola Jokic, Jayson Tatum, and Luka Doncic, with the current budget of \$69857281, current time 104.8 min, and PER of 27.24.
- The remaining budget is $(62000000-69857281) = -7857281$ and the remaining time is $(100 - 104.8) = -4.8$ which are not greater than zero. Nikola Jokic, Jayson Tatum, and Luka Doncic cannot be a valid selection (return).
- Iterating through all the players and choosing the next player, Desmond Bane.
- The current selection of players, Nikola Jokic, Jayson Tatum, and Desmond Bane, has a budget of \$61716050, current time 99.2 min, and PER of 26.17.
- To optimize efficiency, the algorithm checks if a new player combined with previous selections forms a valid combination already calculated (memorized) and returns the precomputed information, avoiding redundant calculations.
- The remaining budget is $(62000000 - 61716050) = 283950$ and the remaining time is $(100 - 99.2) = 0.8$ which are greater than zero. Nikola Jokic, Jayson Tatum, and Desmond Bane are valid selections (return).
- Check if this selection skyline dominates the current best selection. If yes, update the best selection.
- Repeat this process for each player in the table, updating the best selection whenever a skyline-dominating combination is found.

The time complexity of the provided algorithm can be analyzed by considering the key operations in the findBestCombination function and the skylineDominates function. The outer loop iterates through each player in the players list. In the worst case, this loop will be executed for each player, contributing a factor of O(N), where N is the number of players. Inside the loop was a recursive call to findBestCombination with reduced parameters, contributing to the overall time complexity. The number of recursive calls was affected by the time and budget constraints, leading to a time complexity of $O(T \cdot B)$, where T is the maximum time available, and B is the budget constraint.

The memorization mechanism reduced redundant computations by storing and reusing previously computed results. The memorization table ensured that each unique combination of parameters was computed only once. Therefore, the memorization had a time complexity proportional to the number of unique subproblems encountered during the recursive exploration.

The skylineDominates function iterated through each attribute of players in newSelect and bestSelection. In the worst case, this loop would be executed for each attribute, contributing a factor of $O(A \cdot P)$, where A is the number of attributes, and P is the number of players in each team. Considering this adjustment, the overall time complexity of the algorithm would be $O(N \cdot T \cdot B) \times O(A \cdot P)$.

3.3.1.3 Categorized player selection

In NBA games, players are categorized into five positions: Point Guard (PG), Shooting Guard (SG), Small Forward (SF), Power Forward (PF), and Center (C), each specializing in different aspects of the game such as ball handling, scoring, and defensive play. These positions help teams organize their strategies, utilize player strengths effectively, and maintain balance on the court.

Algorithm 1 has been adapted to enhance player selection efficiency by accounting for player types. With ten distinct player types, representing five starting players and five bench players across the five positions on the court, the modification in Algorithm 3 (lines

Algorithm 3: Dynamic Programming Approach with Skyline Dominates

```
   Input: List of current selection (cSelect), current budget
          (cBudget), current time (cTime)
   Result: Best player selection based on skyline dominance
1  function findBestCombination(cSelect, cBudget, cTime)
2      if (tuple(cSelect), cBudget, cTime) is in memo then
3          return memo[(tuple(cSelect), cBudget, cTime)]
4      end
5      for player in allplayers do
6          if player["Time"] ≤ cTime and player["Salary"] ≤ cBudget
            then
7              if player["Type"] ∈ cSelect["Type"] then
8                  oldPlayer ← cSelect["Type"].find(player["Type"]);
9                  newSelect ← cSelect.swap(oldPlayer, player);
10                 player["Salary"] ← player["Salary"] -
                    oldPlayer["Salary"];
11                 player["Time"] ← player["Time"] - oldPlayer["Time"];
12             end
13             else
14                 newSelect ← cSelect + [player];
15             end
16             memo[(tuple(newSelect), cBudget, cTime)] ←
                findBestCombination( newSelect, cBudget -
                player["Salary"], cTime - player["Time"])
17             if skylineDominates(newSelect, bestSelection) then
18                 bestSelection ← newSelect
19             end
20         end
21     end
22     return memo[(tuple(newSelect), cBudget, cTime)]
23 end
```

7 to 12) focused on refining player selection. This adjustment entailed a systematic approach as if a player of the same type was already selected, the algorithm replaced them with the new player. Adjustments to both salary and time were necessary as the players were swapped. Conversely, if a certain player type remained unselected, the algorithm

included the new player in the selection. This strategic refinement ensured a well-balanced and optimized player roster tailored to the specific roles and positions required for effective gameplay.

3.3.2 Optimal Solution and Linear Programming

Linear programming [21] is a mathematical method to optimize complex systems by allocating limited resources. Employed in diverse fields, including economics, engineering, and operations research, linear programming involves maximizing or minimizing a linear objective function subject to linear constraints.

$$\begin{aligned} &\text{Maximize } c^T \\ &\text{subject to } Ax \leq b \end{aligned} \tag{3.6}$$

In equation 3.6, c and x represent the objective function and decision variables, respectively. A is the constraints matrix, and b is the vector of constraint coefficients. This mathematical framework is a powerful tool for optimizing resource allocation and decision-making processes, aiding researchers and practitioners in efficiently and effectively addressing complex real-world problems.

In operations research and optimization, the simplex method is a fundamental algorithm for solving linear programming problems. First introduced by George Dantzig in the mid-20th century, the simplex method [22] offers a systematic approach to navigating the feasible region defined by linear constraints to optimize the objective function. Through a series of iterative steps, the simplex method moves from one vertex of the feasible region to another, gradually improving the objective value until it reaches an optimal solution. Despite its reliance on computational resources, the simplex method remains a cornerstone in linear programming due to its reliability and effectiveness in tackling complex optimization problems.The utilization of this linear programming approach in NBA player selection strategies is noteworthy. NBA teams are challenged to assemble a roster within specific constraints, such as salary cap limitations, player positions, team chemistry considerations, and court coverage. By framing this problem as a linear programming model with these constraints, teams can pursue various strategies to achieve their objectives. For example, one approach may center on maximizing team performance and winning potential by strategically selecting players with complementary skills and attributes. Conversely, another strategy may prioritize financial efficiency by minimizing the total team salary while ensuring competitive strength on the court.

CHAPTER 4

WINNING RATE PREDICTION AND MODEL EVALUATION

4.1 Datasets

4.1.1 Player Statistics

The traditional player datasets were systematically acquired from the official National Basketball Association (NBA) website [23] through an extensive web scraping mechanism. This process guaranteed the collection of precise and reliable player-centric information, which served as the foundational dataset for this analysis. Furthermore, crucial additional datasets, such as Player Efficiency Rating (PER) [24], Real Plus-Minus (RPM) [25], and detailed salary information [26], were meticulously sourced from the ESPN website. Integrating these supplementary datasets enriched the overall dataset, enabling a comprehensive evaluation of player performance and financial dynamics within the NBA.

Fig. 4.1 visually represents the dataset containing detailed statistics of NBA players for the 2021-2022 season. This dataset included various performance metrics such as points scored, rebounds, assists, salary, and other relevant player attributes, serving as a comprehensive resource for basketball analytics and research.

	PlayerName	TEAM	PlayerType	Position	GP	Min	PTS	FGM	FGA	3PM	3PA	FTM	FTA	OREB	DREB	REB	AST	TOV	STL	BLK	PF	PER	RPM	Salary
0	Deandre Ayton	PHX	Starter	C	58	29.5	17.2	7.6	12.0	0.1	0.3	1.8	2.4	2.6	7.7	10.2	1.4	1.6	0.7	0.7	2.4	21.99	2.94	12632950
1	Furkan Korkmaz	PHI	Backup	SF	67	21.1	7.6	2.7	7.0	1.1	4.0	1.0	1.3	0.3	2.3	2.6	1.9	0.7	0.5	0.1	0.9	10.22	-2.47	4629630
2	Dalano Banton	TOR	Backup	PG	64	10.9	3.2	1.3	3.2	0.2	0.8	0.4	0.7	0.6	1.3	1.9	1.5	0.8	0.4	0.2	1.1	10.06	-5.60	925258
3	Oshae Brissett	IND	Backup	SF	67	23.3	9.1	3.1	7.6	1.2	3.5	1.7	2.4	1.6	3.7	5.3	1.1	0.8	0.7	0.4	1.7	12.92	-3.28	1701593
4	James Harden	PHI	Starter	PG	65	37.2	22.0	6.3	15.3	2.3	6.9	7.2	8.2	0.8	6.8	7.7	10.3	4.4	1.3	0.6	2.4	20.92	4.62	44310840
5	RJ Barrett	NYK	Starter	SF	70	34.5	20.0	7.0	17.1	2.0	5.8	4.1	5.8	0.9	4.9	5.8	3.0	2.2	0.6	0.2	2.0	13.72	1.10	8623920
6	Gordon Hayward	CHA	Starter	SF	49	31.9	15.9	5.8	12.6	1.8	4.5	2.6	3.0	0.8	3.8	4.6	3.6	1.7	1.0	0.4	1.7	15.11	-0.40	29925000
7	Julius Randle	NYK	Starter	PF	72	35.3	20.1	7.1	17.3	1.7	5.4	4.2	5.6	1.7	8.2	9.9	5.1	3.4	0.7	0.5	2.8	15.80	1.64	21780000
8	Dwight Howard	LAL	Backup	C	60	16.2	6.2	2.2	3.7	0.1	0.3	1.6	2.4	2.0	4.0	5.9	0.6	0.8	0.6	0.6	1.9	18.25	-1.42	1669178
9	Isaac Okoro	CLE	Starter	SG	67	29.6	8.8	3.1	6.4	0.8	2.3	1.8	2.3	1.1	1.9	3.0	1.8	0.9	0.8	0.3	2.4	10.39	-0.32	6720720
10	Theo Maledon	OKC	Backup	C	51	17.8	7.1	2.3	6.2	0.9	2.9	1.5	2.0	0.4	2.2	2.6	2.2	1.3	0.6	0.2	1.3	10.60	-3.29	2000000
11	Tyrese Maxey	PHI	Starter	PG	75	35.3	17.5	6.4	13.3	1.8	4.1	2.8	3.3	0.3	2.9	3.2	4.3	1.2	0.7	0.4	2.1	16.27	3.43	2602920
12	LaMelo Ball	CHA	Starter	PG	75	32.3	20.1	7.2	16.7	2.9	7.5	2.8	3.2	1.4	5.2	6.7	7.6	3.3	1.6	0.4	3.2	19.76	4.43	8231760
13	Kevin Huerter	ATL	Starter	SG	74	29.6	12.1	4.7	10.3	2.2	5.6	0.6	0.7	0.4	3.0	3.4	2.7	1.2	0.7	0.4	2.5	11.91	0.91	4253357
14	Cody Martin	CHA	Backup	SF	71	26.3	7.7	2.9	6.0	0.9	2.2	1.1	1.5	1.2	2.9	4.0	2.5	0.9	1.2	0.5	1.6	12.73	-3.60	1782621

Fig. 4.1. Example of player statistics for season 2021-2022.

4.1.2 Dataset Preprocessing

Datasets were preprocessed to utilize team performance. The preprocessing step for creating team statistics from individual player performance is illustrated. This process involved aggregating the performance metrics of both starter and bench players to derive comprehensive insights into team dynamics and contributions.

Each team's overall performance was quantified through various metrics such as Total Player Efficiency Rating (PER), Real Plus-Minus (RPM), defensive stats, points scored (PTS), assists (AST), and more. By combining the individual contributions of players, this preprocessing step provides a holistic view of team performance, enabling deeper analysis and comparison across teams. For instance, in Fig. 4.2, the Total PER (starter) for a team like the Atlanta Hawks was computed by summing the Player Efficiency Rating (PER) of all the starters within the team. Similarly, the Total PER (backup) reflected the combined PER of the bench players. This aggregation process was repeated for other performance metrics like RPM, defensive statistics, points scored, and assists, ensuring a comprehensive assessment of each team's overall performance. Through such preprocessing, analysts can identify key trends, strengths, and areas for improvement within teams, facilitating informed decision-making and strategic planning in the context of NBA team management and analytics.

	TEAM	TotalS_PER	TotalS_RPM	TotalS_DEF	TotalB_PER	TotalB_RPM	TotalB_DEF	WinPTC
0	Atlanta Hawks	88.23	14.53	567.3	75.71	-6.99	552.1	0.524
1	Boston Celtics	93.33	27.85	519.2	61.37	-17.03	534.0	0.622
2	Brooklyn Nets	85.84	9.58	556.3	62.04	-17.58	560.3	0.537
3	Charlotte Hornets	85.18	15.48	557.5	70.10	-11.59	561.4	0.524
4	Chicago Bulls	86.76	7.21	564.3	62.43	-19.01	545.9	0.561
5	Cleveland Cavaliers	83.42	13.80	539.6	63.30	-9.76	531.7	0.537
6	Dallas Mavericks	96.71	19.40	550.2	55.38	-10.66	525.8	0.634
7	Denver Nuggets	89.00	24.60	552.3	60.55	-14.29	554.2	0.585
8	Detroit Pistons	66.72	2.42	562.8	71.06	-9.92	560.2	0.280
9	Golden State Warriors	83.40	18.91	526.8	79.49	-0.90	522.6	0.646
10	Houston Rockets	71.83	4.67	575.3	62.09	-14.16	576.7	0.244
11	Indiana Pacers	82.16	10.86	559.5	64.62	-13.62	570.8	0.305
12	LA Clippers	77.36	11.78	549.3	74.43	-7.34	536.1	0.512
13	Los Angeles Lakers	87.59	7.94	558.0	66.23	-13.24	561.0	0.402
14	Memphis Grizzlies	92.39	18.60	540.4	79.72	-5.58	534.7	0.683

Fig. 4.2. Processed team statistics for starter and bench players

4.1.3 Player Attributes Correlation Analysis

This study explored the relationships among crucial player metrics, including Real Plus-Minus (RPM), Player Efficiency Rating (PER), Defensive Rating (DEF), and the Win Rate of each team. The analysis employed a pair plot, a visual tool that allows simultaneously exploring the pairwise relationships between these attributes.

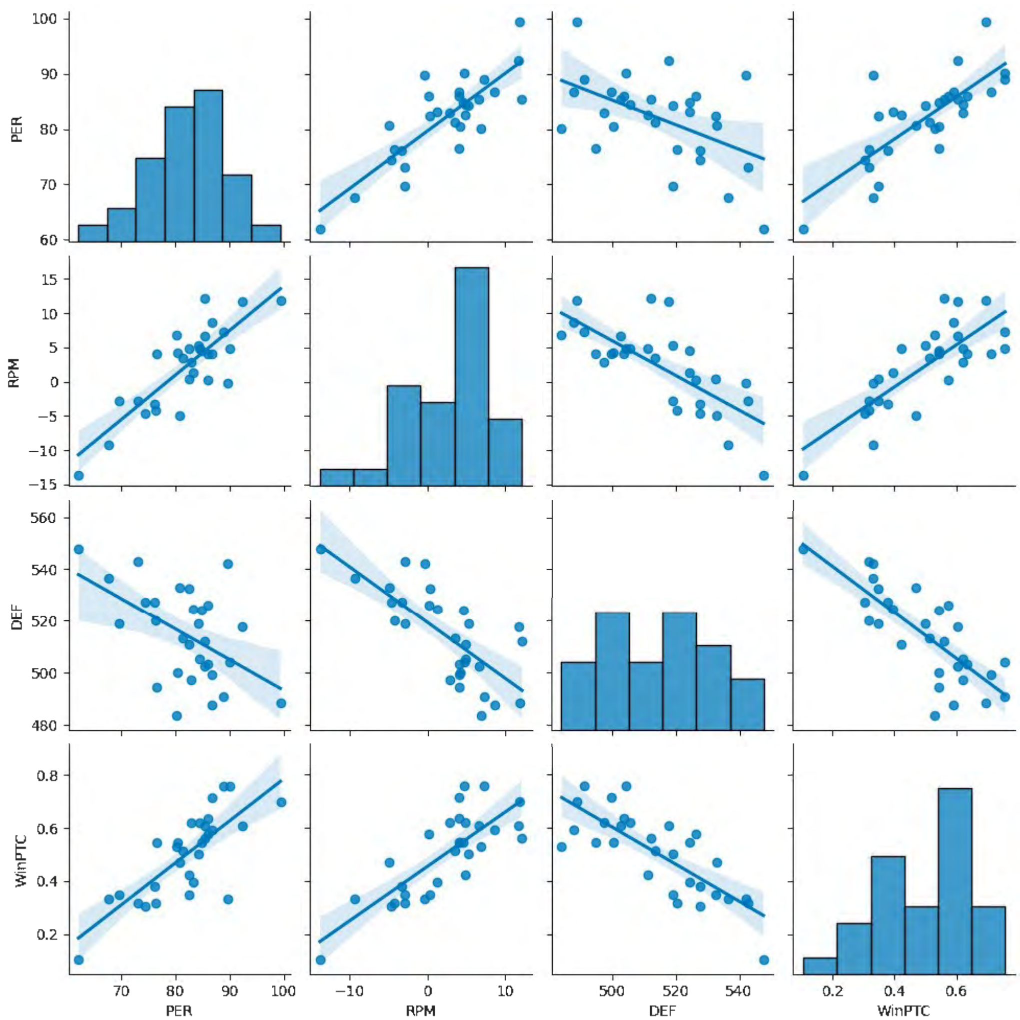


Fig. 4.3. Attributes Correlation Analysis

Through this method of Fig. 4.3, intricate patterns and potential correlations between the performance indicators of individual players and overall team success can be discerned. This analysis played a pivotal role in unraveling the interplay between player attributes and team outcomes, contributing valuable insights to the broader understanding of the factors influencing team performance in the dynamic landscape of professional basketball.

4.2 Winning Rate Prediction

The National Basketball Association (NBA) is the North American professional basketball league comprising 30 teams. Three machine learning models: Lasso, Ridge, and logistic regressions, were used to predict the winning rate. Ten seasons of data were collected from 2012-2013 to 2021-2022. Seasons 2015-2016, 2017-2018, 2019- 2020, and 2021-2022 were used for prediction, and the rest were used for training. First, PER, DEF, and RPM were used to evaluate seven weights (w0, w1 ... w6), and the following 15 parameters were used to predict 31 weights (w0, w1 ... w30). With these machine learning models, weights were trained, and used these weights in the linear programming model to select the optimized team.

4.2.1 Based on 3 attributes

The winning rates are predicted based on three attributes- player efficiency rating (PER), real plus-minus (RPM), and defensive rating (DEF). When examining the data from the 2015-16 and 2017-18 seasons presented in Tables 4.1 and 4.2, respectively, it becomes apparent that certain teams consistently outperformed or underperformed their predicted win rates across both seasons. For instance, perennial contenders like the Golden State Warriors demonstrated remarkable consistency in exceeding their predicted win rates, showcasing the strength of their roster and strategic prowess. Conversely, teams like the Brooklyn Nets faced challenges in meeting their predicted win rates, indicating potential areas for improvement in their performance or discrepancies in the predictive models utilized. Notably, the Atlanta Hawks displayed a notable variation in their actual win rates, ranging from 0.293 in the 2017-18 season to 0.585 in the 2015-16 season, reflecting the team's fluctuating performance over time.

TABLE 4.1
WINNING RATE PREDICTIONS FOR SEASON 2015-16 WITH 3 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.585	0.63101471	0.63103616	0.637326
Boston Celtics	0.585	0.63172495	0.63175123	0.639091
Brooklyn Nets	0.256	0.30158884	0.30154968	0.297381
Charlotte Hornets	0.585	0.46547993	0.46547871	0.461735
Chicago Bulls	0.512	0.41116329	0.41113854	0.403396
Cleveland Cavaliers	0.695	0.61389514	0.6138969	0.619886
Dallas Mavericks	0.512	0.47458006	0.47457391	0.472752
Denver Nuggets	0.402	0.46828794	0.46827821	0.465493
Detroit Pistons	0.537	0.47839209	0.47837889	0.476422
Golden State Warriors	0.89	0.82563478	0.82568152	0.804024
Houston Rockets	0.5	0.56940356	0.56940432	0.574857
Indiana Pacers	0.549	0.53376362	0.53376029	0.53528
LA Clippers	0.646	0.68051902	0.68053173	0.684988
Los Angeles Lakers	0.207	0.14772363	0.14767373	0.177837
Memphis Grizzlies	0.512	0.50654967	0.50655978	0.506634
Miami Heat	0.585	0.59515596	0.59514645	0.598992
Milwaukee Bucks	0.402	0.44449939	0.44447464	0.439238
Minnesota Timberwolves	0.354	0.41809684	0.41808427	0.412209
New Orleans Pelicans	0.366	0.40858026	0.4085759	0.401728
New York Knicks	0.39	0.40494369	0.40492452	0.397381
Oklahoma City Thunder	0.671	0.65841127	0.65841321	0.664733
Orlando Magic	0.427	0.45007217	0.45005603	0.444967
Philadelphia 76ers	0.122	0.33806585	0.33803361	0.330858
Phoenix Suns	0.28	0.39407887	0.39406365	0.387315
Portland Trail Blazers	0.537	0.47322491	0.47322847	0.470971
Sacramento Kings	0.402	0.46578852	0.46578895	0.462328
San Antonio Spurs	0.817	0.80964842	0.80969935	0.791242
Toronto Raptors	0.683	0.54620833	0.54620944	0.547599
Utah Jazz	0.488	0.53682464	0.5368219	0.538647
Washington Wizards	0.5	0.52698794	0.52698838	0.528996

TABLE 4.2
WINNING RATE PREDICTIONS FOR SEASON 2017-18 WITH 3 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.293	0.37244682	0.37243973	0.364574
Boston Celtics	0.671	0.60423839	0.60425224	0.612154
Brooklyn Nets	0.341	0.37683402	0.37683097	0.369242
Charlotte Hornets	0.439	0.44144474	0.44143983	0.437396
Chicago Bulls	0.329	0.25415746	0.2541219	0.256803
Cleveland Cavaliers	0.61	0.45775432	0.45776825	0.453897
Dallas Mavericks	0.293	0.38742482	0.38743925	0.378606
Denver Nuggets	0.561	0.51341701	0.5134384	0.514299
Detroit Pistons	0.476	0.50851175	0.50853254	0.508887
Golden State Warriors	0.707	0.79281322	0.79288146	0.779703
Houston Rockets	0.793	0.77897797	0.77902691	0.769417
Indiana Pacers	0.585	0.50007584	0.50006892	0.5001
LA Clippers	0.512	0.50871056	0.50874317	0.508695
Los Angeles Lakers	0.427	0.47979306	0.47982126	0.478164
Memphis Grizzlies	0.268	0.3477913	0.34776725	0.341163
Miami Heat	0.537	0.49504843	0.49505658	0.493312
Milwaukee Bucks	0.537	0.49586125	0.49585016	0.496347
Minnesota Timberwolves	0.573	0.56057528	0.56059411	0.565199
New Orleans Pelicans	0.585	0.57621695	0.57620991	0.582057
New York Knicks	0.354	0.36898931	0.36895967	0.359795
Oklahoma City Thunder	0.585	0.60995392	0.60997273	0.618743
Orlando Magic	0.305	0.39580627	0.39580014	0.387315
Philadelphia 76ers	0.634	0.65071626	0.65073937	0.659
Phoenix Suns	0.256	0.25014933	0.25011272	0.252969
Portland Trail Blazers	0.598	0.4623455	0.462327	0.458254
Sacramento Kings	0.329	0.24195155	0.24192316	0.244737
San Antonio Spurs	0.573	0.60165316	0.60167106	0.607932
Toronto Raptors	0.72	0.67142573	0.67148666	0.676437
Utah Jazz	0.585	0.60083393	0.60084894	0.607158
Washington Wizards	0.524	0.49792169	0.49792198	0.498145

TABLE 4.3
WINNING RATE PREDICTIONS FOR SEASON 2019-20 WITH 3 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.299	0.29008863	0.29004733	0.285783
Boston Celtics	0.667	0.6678575	0.66789083	0.674685
Brooklyn Nets	0.486	0.41549939	0.4155067	0.410928
Charlotte Hornets	0.354	0.29614814	0.29612345	0.292867
Chicago Bulls	0.338	0.41419658	0.41419362	0.407712
Cleveland Cavaliers	0.292	0.20853106	0.20849667	0.219713
Dallas Mavericks	0.573	0.61492246	0.61496617	0.623437
Denver Nuggets	0.63	0.58533312	0.58537443	0.593858
Detroit Pistons	0.303	0.34664028	0.34665697	0.339896
Golden State Warriors	0.231	0.34456795	0.3445579	0.338936
Houston Rockets	0.611	0.57037637	0.57039835	0.576491
Indiana Pacers	0.616	0.50756865	0.50757509	0.507314
LA Clippers	0.681	0.70308715	0.70314619	0.707644
Los Angeles Lakers	0.732	0.70469005	0.70473962	0.709649
Memphis Grizzlies	0.466	0.43365038	0.43365854	0.427747
Miami Heat	0.603	0.52103652	0.52106433	0.524507
Milwaukee Bucks	0.767	0.85593681	0.85601297	0.825502
Minnesota Timberwolves	0.297	0.39559031	0.39558636	0.388676
New Orleans Pelicans	0.417	0.4783321	0.47834454	0.478681
New York Knicks	0.318	0.22492093	0.22487438	0.231208
Oklahoma City Thunder	0.611	0.56402025	0.56405372	0.568984
Orlando Magic	0.452	0.52266869	0.52268878	0.5238
Philadelphia 76ers	0.589	0.58971878	0.58973399	0.597064
Phoenix Suns	0.466	0.4959387	0.49595509	0.495071
Portland Trail Blazers	0.473	0.45938932	0.45937729	0.456395
Sacramento Kings	0.431	0.41474363	0.41473525	0.407948
San Antonio Spurs	0.451	0.42534142	0.42536521	0.418257
Toronto Raptors	0.736	0.69641947	0.69650382	0.702926
Utah Jazz	0.611	0.5177384	0.51777068	0.520022
Washington Wizards	0.347	0.28366486	0.28366811	0.281074

TABLE 4.4
WINNING RATE PREDICTIONS FOR SEASON 2021-22 WITH 3 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.524	0.57603198	0.57608486	0.58091
Boston Celtics	0.622	0.69248943	0.69247502	0.694655
Brooklyn Nets	0.537	0.41882986	0.41879325	0.413076
Charlotte Hornets	0.524	0.51553649	0.51555458	0.516546
Chicago Bulls	0.561	0.41122203	0.41117602	0.404508
Cleveland Cavaliers	0.537	0.54869341	0.54868716	0.551233
Dallas Mavericks	0.634	0.67562612	0.67563594	0.678587
Denver Nuggets	0.585	0.60381206	0.60382443	0.608464
Detroit Pistons	0.28	0.28690213	0.2868847	0.285878
Golden State Warriors	0.646	0.69658172	0.69664234	0.699691
Houston Rockets	0.244	0.26883714	0.26880893	0.269918
Indiana Pacers	0.305	0.41871046	0.41870089	0.413462
LA Clippers	0.512	0.50742361	0.50744112	0.507639
Los Angeles Lakers	0.402	0.45099817	0.45098734	0.447543
Memphis Grizzlies	0.683	0.69651064	0.6965678	0.699626
Miami Heat	0.646	0.62388781	0.62391904	0.629689
Milwaukee Bucks	0.622	0.59956211	0.59953722	0.604455
Minnesota Timberwolves	0.561	0.56675195	0.56677313	0.570843
New Orleans Pelicans	0.439	0.47888893	0.47890526	0.477584
New York Knicks	0.451	0.45914566	0.45912414	0.45456
Oklahoma City Thunder	0.293	0.33089005	0.33086344	0.325598
Orlando Magic	0.268	0.32057056	0.32052079	0.315749
Philadelphia 76ers	0.6221	0.5943406	0.59434698	0.599141
Phoenix Suns	0.78	0.68457878	0.68459207	0.687399
Portland Trail Blazers	0.329	0.42340843	0.42340971	0.418588
Sacramento Kings	0.366	0.35030684	0.35029689	0.344133
San Antonio Spurs	0.415	0.42740843	0.42736714	0.421666
Toronto Raptors	0.585	0.52208135	0.52205465	0.522189
Utah Jazz	0.598	0.69527421	0.69532088	0.697729
Washington Wizards	0.427	0.42450585	0.42451152	0.41988

The winning rate predictions for the 2019-20 and 2021-22 NBA seasons, as shown in Tables 4.3 and 4.4, present a more nuanced picture of team performance. While some teams closely aligned with their predicted win rates, others experienced notable deviations, suggesting the influence of unpredictable factors such as player injuries or team chemistry. The consistency of certain franchises, like the Milwaukee Bucks and Philadelphia 76ers, in meeting or surpassing their predicted win rates underscored the effectiveness of their roster management and coaching strategies. However, discrepancies between predicted and actual win rates for other teams highlighted the inherent complexity of NBA dynamics and the challenges of accurately forecasting team success.

4.2.2 Based on 15 attributes

Winning rates were also predicted using 15 parameters. In Tables 4.5 and 4.6, the winning rate predictions for the NBA seasons 2015-16 and 2017-18, respectively, are presented alongside the actual win rates. These tables showcase the performance of three different prediction models, Lasso, Ridge, and Logistic regression, in forecasting team success based on various attributes. For instance, in the 2015-16 season, the models' predictions generally aligned closely with the actual win rates across different teams. Notably, for teams like the Golden State Warriors and San Antonio Spurs, which had high actual win rates, the prediction models also yielded relatively accurate estimations, albeit with some variation.

TABLE 4.5
WINNING RATE PREDICTIONS FOR SEASON 2015-16 WITH 15 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.585	0.62771538	0.62642684	0.635996
Boston Celtics	0.585	0.57262704	0.56980373	0.576524
Brooklyn Nets	0.256	0.3251687	0.32548255	0.328
Charlotte Hornets	0.585	0.48180876	0.48014584	0.475319
Chicago Bulls	0.512	0.42343615	0.42401592	0.416893
Cleveland Cavaliers	0.695	0.62437021	0.62432706	0.630531
Dallas Mavericks	0.512	0.49585978	0.49504288	0.492561
Denver Nuggets	0.402	0.34280933	0.33823448	0.331097
Detroit Pistons	0.537	0.48645046	0.48428935	0.479798
Golden State Warriors	0.89	0.88023936	0.8836514	0.84194
Houston Rockets	0.5	0.51954316	0.5197452	0.517186
Indiana Pacers	0.549	0.53152659	0.53032023	0.528809
LA Clippers	0.646	0.67966271	0.67757674	0.674608
Los Angeles Lakers	0.207	0.15381881	0.15258283	0.178793
Memphis Grizzlies	0.512	0.42825556	0.42688222	0.419733
Miami Heat	0.585	0.57890759	0.5783239	0.579894
Milwaukee Bucks	0.402	0.40427044	0.40490646	0.397366
Minnesota Timberwolves	0.354	0.42666031	0.42603785	0.421225
New Orleans Pelicans	0.366	0.38728631	0.38676929	0.376716
New York Knicks	0.39	0.43267918	0.43274176	0.427373
Oklahoma City Thunder	0.671	0.67920084	0.67999604	0.684514
Orlando Magic	0.427	0.47210919	0.47108281	0.465439
Philadelphia 76ers	0.122	0.2585331	0.2579322	0.259586
Phoenix Suns	0.28	0.30183654	0.30113375	0.300635
Portland Trail Blazers	0.537	0.48679331	0.48719933	0.481237
Sacramento Kings	0.402	0.4359119	0.43712074	0.428992
San Antonio Spurs	0.817	0.82696462	0.82667995	0.805933
Toronto Raptors	0.683	0.56718123	0.56750251	0.566434
Utah Jazz	0.488	0.57218781	0.57226097	0.574085
Washington Wizards	0.5	0.52425516	0.5258017	0.527794

TABLE 4.6
WINNING RATE PREDICTIONS FOR SEASON 2017-18 WITH 15 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.293	0.37055835	0.36915888	0.365697
Boston Celtics	0.671	0.61040173	0.61037048	0.616668
Brooklyn Nets	0.341	0.39819811	0.39802648	0.392652
Charlotte Hornets	0.439	0.45329137	0.4517396	0.442942
Chicago Bulls	0.329	0.28668008	0.28904788	0.29194
Cleveland Cavaliers	0.61	0.51480308	0.51527647	0.518057
Dallas Mavericks	0.293	0.43233557	0.43156455	0.419377
Denver Nuggets	0.561	0.54417986	0.54761825	0.559592
Detroit Pistons	0.476	0.48279398	0.48346965	0.479683
Golden State Warriors	0.707	0.73094582	0.72854335	0.737258
Houston Rockets	0.793	0.77573816	0.77636393	0.758526
Indiana Pacers	0.585	0.52417141	0.52626993	0.531098
LA Clippers	0.512	0.52938694	0.52726292	0.528459
Los Angeles Lakers	0.427	0.43599294	0.43516669	0.425315
Memphis Grizzlies	0.268	0.2763992	0.27250631	0.269952
Miami Heat	0.537	0.46931107	0.4691103	0.468127
Milwaukee Bucks	0.537	0.50095345	0.49912086	0.502204
Minnesota Timberwolves	0.573	0.62042009	0.62104637	0.627575
New Orleans Pelicans	0.585	0.56361785	0.56429894	0.579595
New York Knicks	0.354	0.35881177	0.35816746	0.352621
Oklahoma City Thunder	0.671	0.64280198	0.64351583	0.653779
Orlando Magic	0.305	0.37574753	0.374976	0.373448
Philadelphia 76ers	0.634	0.64153495	0.64056899	0.654074
Phoenix Suns	0.256	0.27336136	0.27390335	0.272595
Portland Trail Blazers	0.598	0.50443114	0.50263802	0.495973
Sacramento Kings	0.329	0.20380068	0.20445232	0.218908
San Antonio Spurs	0.573	0.60011054	0.59955163	0.608051
Toronto Raptors	0.72	0.684797	0.68254516	0.684624
Utah Jazz	0.585	0.64018247	0.63834796	0.650029
Washington Wizards	0.524	0.50503867	0.50579674	0.505691

TABLE 4.7
WINNING RATE PREDICTIONS FOR SEASON 2019-20 WITH 15 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.299	0.30346954	0.30300678	0.296235
Boston Celtics	0.667	0.6861262	0.68348696	0.686468
Brooklyn Nets	0.486	0.42778855	0.42561217	0.419296
Charlotte Hornets	0.354	0.3109069	0.31126763	0.304987
Chicago Bulls	0.338	0.44470432	0.44616485	0.444174
Cleveland Cavaliers	0.292	0.23326217	0.23493171	0.248805
Dallas Mavericks	0.573	0.65426834	0.65222946	0.657145
Denver Nuggets	0.63	0.60724814	0.60976333	0.622863
Detroit Pistons	0.303	0.44496189	0.44787907	0.447182
Golden State Warriors	0.231	0.26336727	0.25969614	0.268251
Houston Rockets	0.611	0.58797999	0.5876712	0.588782
Indiana Pacers	0.616	0.56845802	0.56953153	0.580504
LA Clippers	0.681	0.70703983	0.70782932	0.710102
Los Angeles Lakers	0.732	0.68419341	0.68493142	0.693888
Memphis Grizzlies	0.466	0.48390256	0.48336297	0.490218
Miami Heat	0.603	0.59245413	0.59310539	0.601882
Milwaukee Bucks	0.767	0.86026066	0.8549884	0.824762
Minnesota Timberwolves	0.297	0.40188873	0.39833006	0.390816
New Orleans Pelicans	0.417	0.47257954	0.47476279	0.479594
New York Knicks	0.318	0.25652428	0.25898587	0.258232
Oklahoma City Thunder	0.611	0.55460603	0.55175889	0.553884
Orlando Magic	0.452	0.48877766	0.48447353	0.487836
Philadelphia 76ers	0.589	0.62050757	0.62098288	0.631142
Phoenix Suns	0.466	0.57629521	0.57757288	0.587054
Portland Trail Blazers	0.473	0.44937886	0.44897725	0.441887
Sacramento Kings	0.431	0.42378524	0.42461206	0.421956
San Antonio Spurs	0.451	0.46368636	0.46303203	0.463419
Toronto Raptors	0.736	0.66745908	0.66290817	0.66441
Utah Jazz	0.611	0.58146302	0.58223393	0.586939
Washington Wizards	0.347	0.3576417	0.36090837	0.348542

TABLE 4.8
WINNING RATE PREDICTIONS FOR SEASON 2021-22 WITH 15 ATTRIBUTES

Team Name	Actual Win	Lasso	Ridge	Logistic
Atlanta Hawks	0.524	0.59378792	0.59348148	0.602127
Boston Celtics	0.622	0.66653437	0.66457038	0.670409
Brooklyn Nets	0.537	0.44253314	0.44062795	0.440641
Charlotte Hornets	0.524	0.51717647	0.51763357	0.518822
Chicago Bulls	0.561	0.46097818	0.46235881	0.452827
Cleveland Cavaliers	0.537	0.62199035	0.62295403	0.621481
Dallas Mavericks	0.634	0.65065568	0.64998935	0.651548
Denver Nuggets	0.585	0.54977313	0.54533898	0.548955
Detroit Pistons	0.28	0.24266008	0.24039506	0.245589
Golden State Warriors	0.646	0.66792587	0.66558055	0.67433
Houston Rockets	0.244	0.22706228	0.22476425	0.233449
Indiana Pacers	0.305	0.3951774	0.39473192	0.385991
LA Clippers	0.512	0.47359592	0.47118669	0.46852
Los Angeles Lakers	0.402	0.40124047	0.39819673	0.388224
Memphis Grizzlies	0.683	0.66768625	0.66958454	0.677364
Miami Heat	0.646	0.68865354	0.69025195	0.692369
Milwaukee Bucks	0.622	0.6062648	0.60660696	0.609291
Minnesota Timberwolves	0.561	0.51831096	0.51622528	0.513602
New Orleans Pelicans	0.439	0.46063377	0.46035223	0.458409
New York Knicks	0.451	0.49462008	0.49497012	0.491598
Oklahoma City Thunder	0.293	0.27393162	0.27057649	0.267207
Orlando Magic	0.268	0.23191567	0.22747376	0.236911
Philadelphia 76ers	0.6221	0.65875436	0.65975611	0.662942
Phoenix Suns	0.78	0.73573624	0.73634988	0.730308
Portland Trail Blazers	0.329	0.42369155	0.42368555	0.411185
Sacramento Kings	0.366	0.33246084	0.33085964	0.321434
San Antonio Spurs	0.415	0.42665392	0.42704387	0.422832
Toronto Raptors	0.585	0.59145216	0.59319984	0.59194
Utah Jazz	0.598	0.70257323	0.70047906	0.697761
Washington Wizards	0.427	0.44376191	0.44360076	0.438166

Tables 4.7 and 4.8 display predictions for the 2017-18 and 2021-22 NBA seasons, respectively. Here, despite certain discrepancies between the predicted and actual win rates, the models still demonstrated a reasonable level of predictive accuracy. For instance, while some teams like the Boston Celtics and Miami Heat exhibited closer alignment between predicted and actual win rates, others, such as the Phoenix Suns and Los Angeles Lakers, showed more variability. These tables underscore the complexity of forecasting team performance in the NBA, where various factors contribute to outcomes beyond what can be captured by statistical models alone.

4.2.3 Trained Regression Weights

With the coordination of the machine learning model, the weights were generated. These weights will later be used in the linear programming model to select the optimized players for team formation. The Table 4.9 presents the weight coefficients for Lasso, Ridge, and Logistic regressions applied to a dataset with 15 attributes, encompassing both starter and backup players in the NBA. The regression models aimed to predict team winning rates, with each weight (w) corresponding to a specific attribute. Analyzing the weight coefficients from the regression models revealed insightful patterns in the context of NBA team winning rate predictions. The constant terms w_0 were consistently positive, indicating a baseline positive influence on the predicted winning rates. Among the positive weights, high values were observed for attributes like w_8 (0.148430) and w_{23} (0.198431), suggesting that strong performance in these areas significantly contributed to higher predicted winning rates. Conversely, negative weights, such as w_7 (-0.094385) and w_{22} (-0.093507), highlighted attributes where subpar performance may lead to decreased winning probabilities. The regularization effects of Lasso and Ridge were evident in the sparsity of certain weights, emphasizing the model's focus on key contributing factors. Overall, these weight coefficients provided a nuanced understanding of the impact of different players and team attributes on the outcomes predicted by the regression models.

TABLE 4.9
WEIGHT COEFFICIENTS OF LASSO, RIDGE, AND LOGISTIC REGRESSIONS FOR 15 ATTRIBUTES

#	Lasso	Ridge	Logistic	#	Lasso	Ridge	Logistic
w_0	2.726073	2.748405	9.207153				
w_1	-0.005777	-0.006420	-0.034572	w_{16}	-0.001388	-0.002091	-0.007241
w_2	0.040083	0.042495	0.196955	w_{17}	0.023944	0.025997	0.105707
w_3	0.005412	0.005757	0.029405	w_{18}	0.003370	0.003456	0.015688
w_4	-0.005744	-0.006554	-0.021241	w_{19}	-0.002405	-0.002537	-0.014515
w_5	-0.001001	-0.001019	-0.003823	w_{20}	-0.003203	-0.003237	-0.013572
w_6	-0.004150	-0.004284	-0.013567	w_{21}	-0.003902	-0.004018	-0.012871
w_7	-0.021583	-0.021940	-0.094385	w_{22}	-0.020256	-0.020611	-0.093507
w_8	0.035559	0.036356	0.148430	w_{23}	0.044190	0.044911	0.198431
w_9	0.018646	0.021131	0.084800	w_{24}	-0.005743	-0.006139	-0.036679
w_{10}	-0.006576	-0.009603	-0.033711	w_{25}	0.018202	0.018700	0.085889
w_{11}	0.016646	0.017146	0.071058	w_{26}	0.020446	0.021307	0.097968
w_{12}	-0.007647	-0.007599	-0.035362	w_{27}	-0.001592	-0.001723	-0.007977
w_{13}	-0.000750	-0.000798	-0.000274	w_{28}	0.002470	0.003029	0.018084
w_{14}	-0.019323	-0.019973	-0.080043	w_{29}	-0.027344	-0.027587	-0.109293
w_{15}	0.003413	0.003174	0.014545	w_{30}	0.003753	0.003477	0.016985

Table 4.10 displays the weight coefficients for Lasso, Ridge, and Logistic regressions applied to a dataset with three attributes encompassing starter and backup players in the NBA. The constant terms w_0 were consistently positive, with values of 0.641259 for Lasso, 0.673022 for Ridge, and 0.698581 for Logistic regression, suggesting a baseline positive influence on the predicted outcomes. Positive weights, exemplified by w_2 (0.039990), suggested that excelling in a specific area substantially enhanced predicted outcomes. In contrast, negative weights like w_4 (-0.003344) underscored attributes where below-average performance. This utilization of positive and negative weights was a crucial aspect in

balancing the importance of attributes during the training of machine learning models, ensuring a nuanced understanding of how different attributes influence predictions.

TABLE 4.10
WEIGHT COEFFICIENTS OF LASSO, RIDGE, AND LOGISTIC REGRESSIONS FOR 3 ATTRIBUTES

#	Lasso	Ridge	Logistic
w_0	0.641259	0.673022	0.698581
w_1	-0.000301	-0.000385	-0.001407
w_2	0.008714	0.009124	0.039990
w_3	0.000656	0.000591	0.002225
w_4	-0.000797	-0.000763	-0.003344
w_5	0.012749	0.013050	0.057696
w_6	0.004781	0.004747	0.019982

4.2.4 Model Evaluation and Result Analysis

Several metrics were crucial in assessing the model's predictive performance in regression model evaluation. Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are commonly employed measures that quantify the differences between predicted and observed values. MSE calculates the average squared differences, emphasizing more significant errors, while RMSE provides a more interpretable scale by taking the square root of MSE. Conversely, MAE computes the average absolute differences, offering a straightforward interpretation of average prediction accuracy. In addition to these metrics, R-Square, also known as the coefficient of determination, stands out as a vital indicator of the model's goodness of fit.

$$R^2 = 1 - \frac{\text{Sum of squared errors}}{\text{Total sum of squares}} \tag{4.1}$$

R-Square measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A high R-Square value, close to one, signifies that the model captured a significant portion of the variability in the response variable, indicating a better fit. R-Square comprehensively assesses the model's explanatory power, making it a preferred metric for evaluating regression models. Its intuitive interpretation and ability to capture the proportion of variability explained make R-Square a valuable tool in determining the effectiveness of a regression model in capturing the underlying patterns in the data.

TABLE 4.11
R2 VALUES FOR 3 ATTRIBUTES

Model	Lasso	Ridge	Logistic
2015-16	0.808740625	0.808761405	0.811131373
2017-18	0.813513481	0.813488793	0.823577859
2019-20	0.832752595	0.832740077	0.848202821
2021-22	0.818612063	0.818556351	0.816241400

Table 4.11 presents R^2 values for machine learning algorithms, namely Lasso, Ridge, and Logistic regression, applied to predict NBA team performance based on a dataset with three attributes across four seasons: 2015-16, 2017-18, 2019-20, and 2021-22. The R^2 values quantified the proportion of variance in team performance explained by each model, serving as a metric for predictive accuracy. Across the four seasons, all three algorithms consistently demonstrated strong performance, with R^2 values ranging from 0.808 to 0.848. Remarkably, Logistic regression consistently exhibited the highest R^2 values across all seasons, suggesting its superior predictive power for the given dataset and attributes.

However, the differences between the algorithms were relatively small, emphasizing their comparable effectiveness in capturing the underlying patterns in NBA team performance. While Logistic regression performs marginally better based on the R^2 values, other factors such as model interpretability and computational efficiency should also be considered when determining the most suitable algorithm for a given application.

TABLE 4.12
R2 VALUES FOR 15 ATTRIBUTES

Model	Lasso	Ridge	Logistic
2015-16	0.877329186	0.876143454	0.87043161
2017-18	0.851194148	0.852948728	0.85934959
2019-20	0.847181817	0.84782169	0.852746332
2021-22	0.862518322	0.86059582	0.860842011

The presented Table 4.12 illustrates the R^2 values for three machine learning algorithms— Lasso, Ridge, and Logistic regression applied to predict NBA team performance over the following selected four seasons: 2015-16, 2017-18, 2019-20, and 2021-22, based on a dataset comprising 15 attributes. The R^2 values indicate the model's ability to explain the variance in team performance. Across all seasons, the three algorithms consistently demonstrated high predictive accuracy, with R^2 values ranging from 0.847 to 0.877. Like the three attributes, Logistic regression consistently outperformed Lasso and Ridge regression regarding R^2 values across all seasons for 15 attributes. The use of 15 attributes appeared to provide additional information, contributing to the improved predictive power of these machine learning algorithms compared to a dataset with only three attributes.

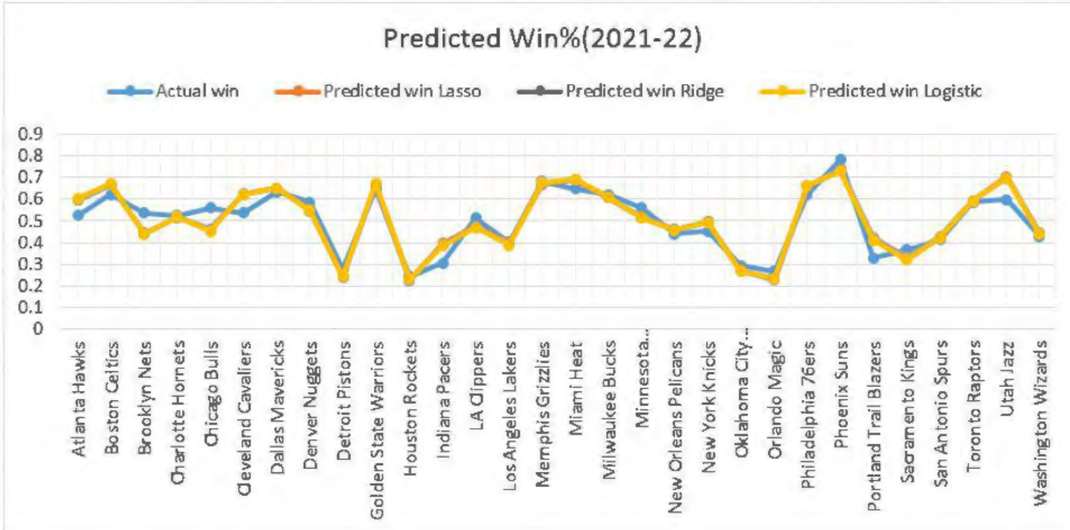


Fig. 4.4. Predicted winning rate versus actual winning rate of season 2021-22 (15 attributes).

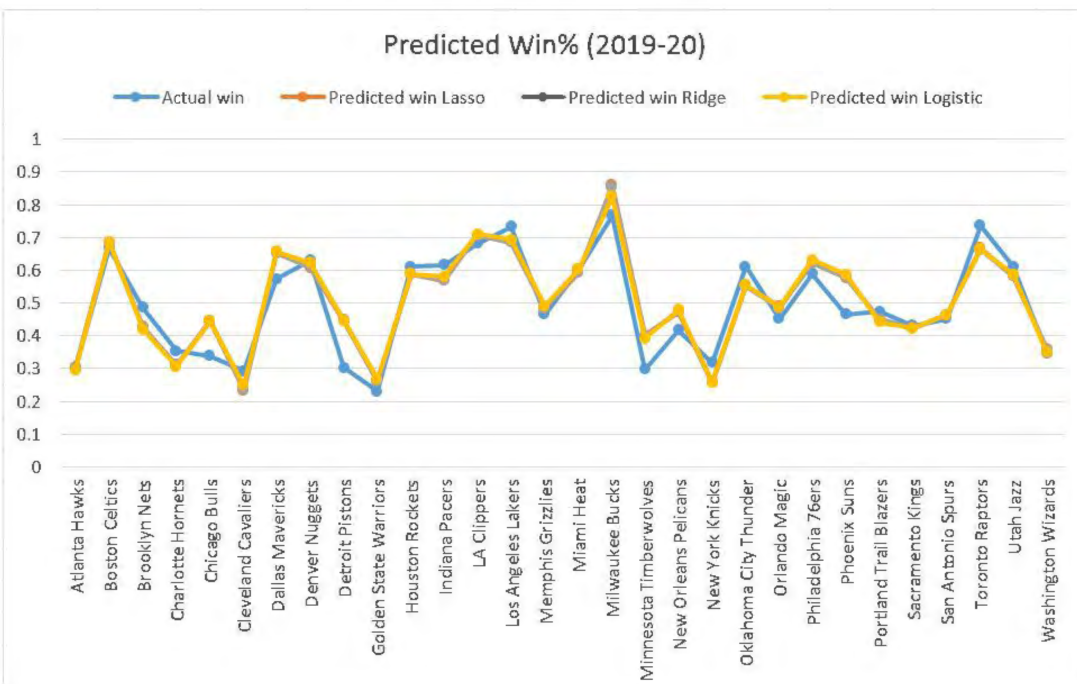


Fig. 4.5. Predicted winning rate versus actual winning rate of season 2019-20 (15 attributes).

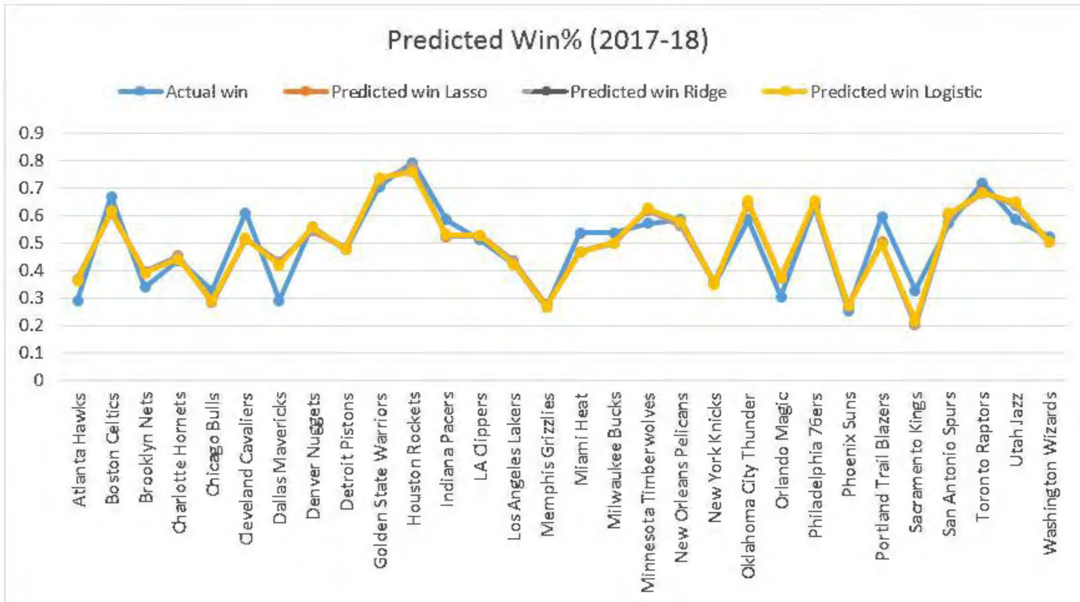


Fig. 4.6. Predicted winning rate versus actual winning rate of season 2017-18 (15 attributes).

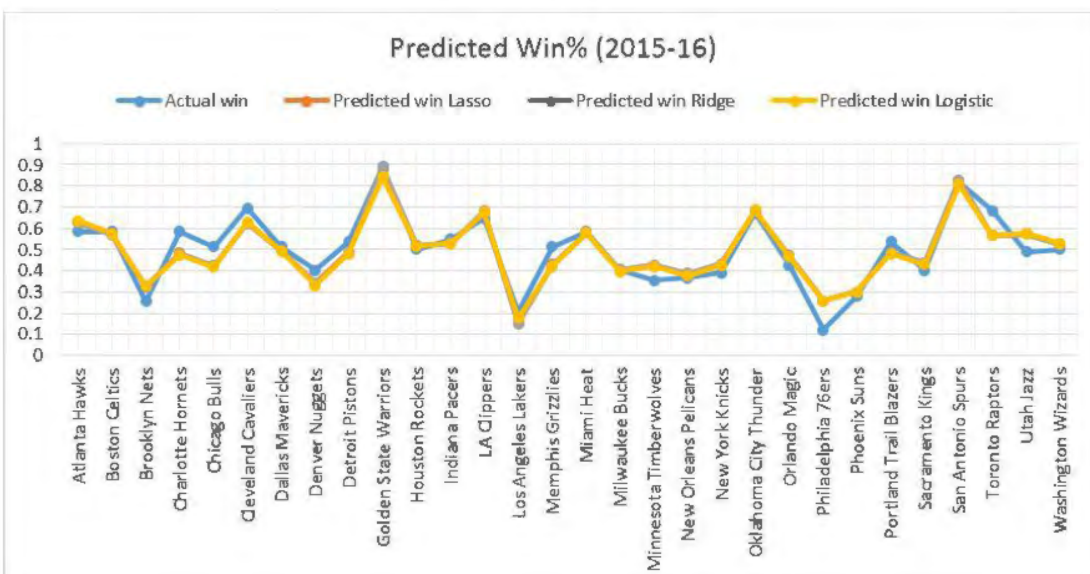


Fig. 4.7. Predicted winning rate versus actual winning rate of season 2015-16 (15 attributes).

The Figs. present a comparative analysis of predicted winning rates versus actual winning rates for four NBA seasons with 15 attributes. Specifically, Fig. 4.4 showcases the predictive performance for the 2021-22 season, while Fig. 4.5 highlights the same for the 2019-20 season. Similarly, Fig. 4.6 provides insights into the 2017-18 season, and Fig. 4.7 does so for the 2015-16 season. These visualizations allow stakeholders to assess the accuracy of predictive models in capturing the complexities of NBA team dynamics. By comparing the predicted values with actual outcomes, the reliability of the models in forecasting team performance can be evaluated. Additionally, including Lasso, Ridge, and Logistic regression plots alongside actual winning rates allows for a comprehensive visual examination of each model's predictive capabilities based on the 15 attributes considered in the analysis.

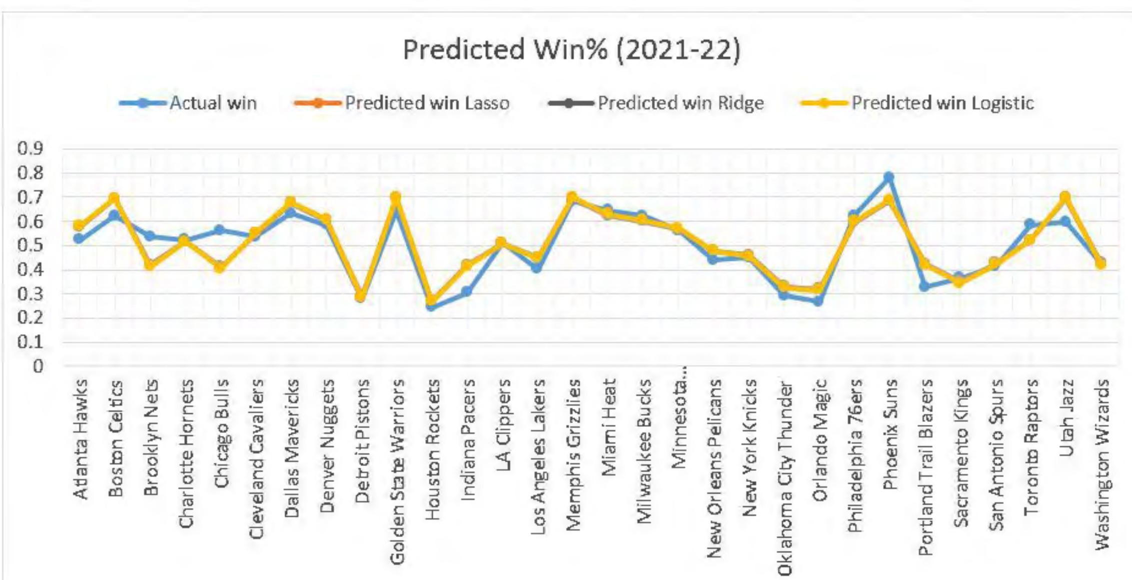


Fig. 4.8. Predicted winning rate versus actual winning rate of season 2021-22 (3 attributes).

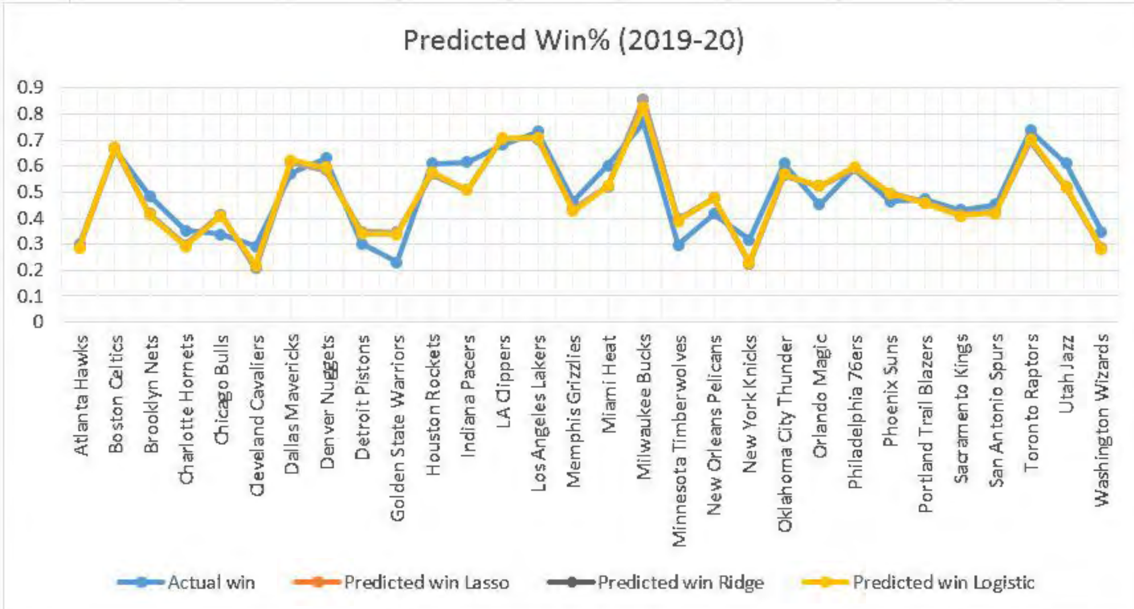


Fig. 4.9. Predicted winning rate versus actual winning rate of season 2019-20 (3 attributes).

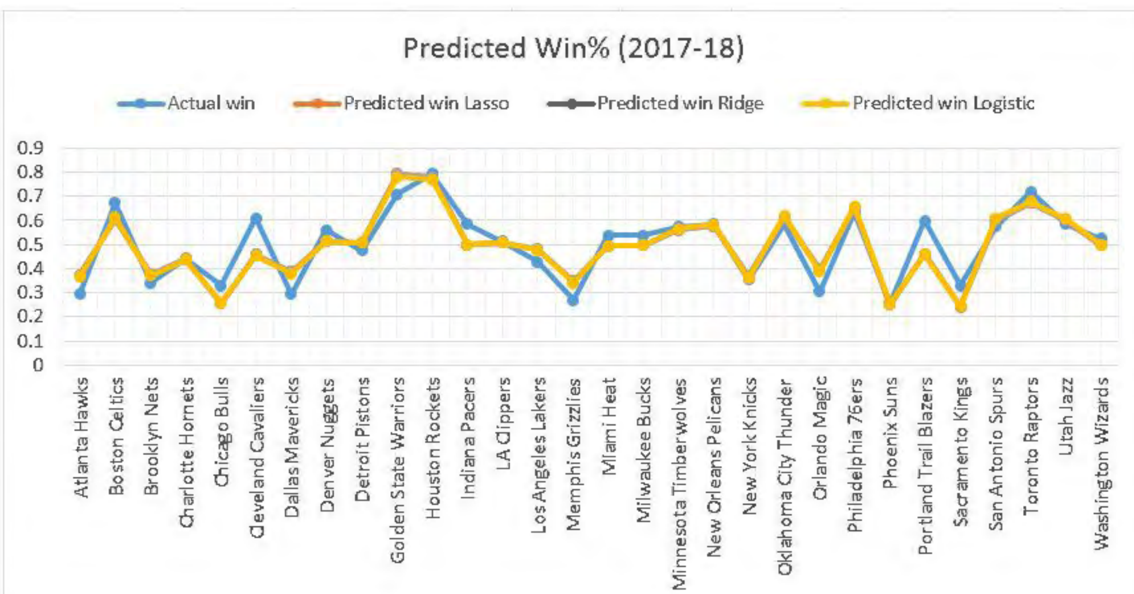


Fig. 4.10. Predicted winning rate versus actual winning rate of season 2017-18 (3 attributes).

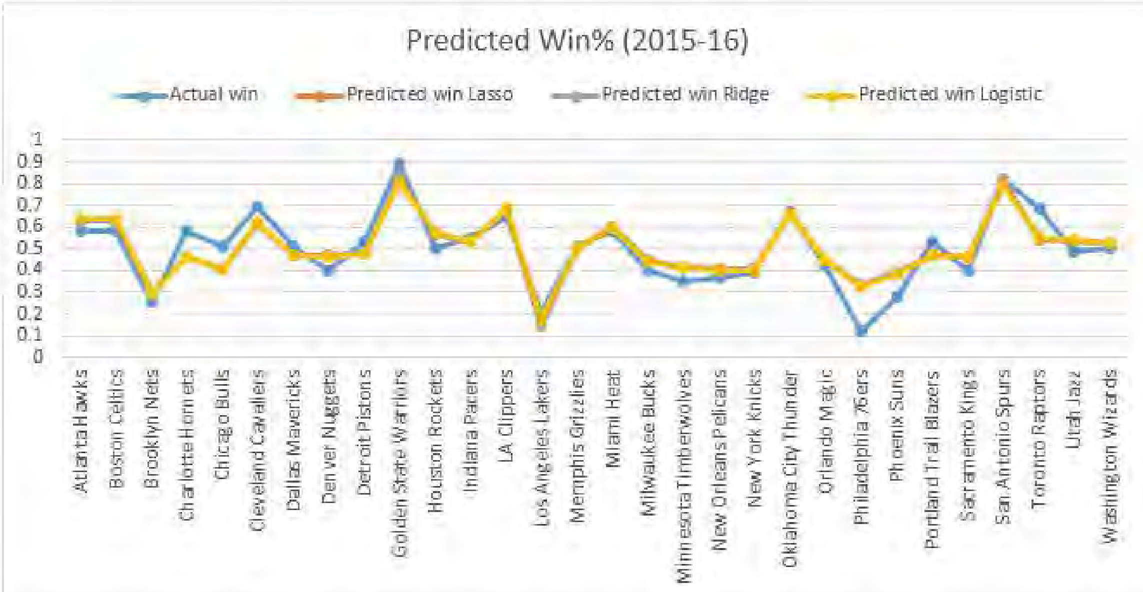


Fig. 4.11. Predicted winning rate versus actual winning rate of season 2015-16 (3 attributes).

The Figs. compared predicted (Lasso, Ridge, and Logistic) and actual winning rates for four NBA seasons, each analyzed with only three attributes. Specifically, Fig. 4.8 illustrates the predictive performance for the 2021-22 season, while Fig. 4.9 showcases the same for the 2019-20 season. Fig. 4.10 compares the 2017- 18 season, and Fig. 4.11 does so for the 2015-16 season. These visualizations offer stakeholders insights into the effectiveness of predictive models when utilizing a reduced set of attributes.

CHAPTER 5

TEAM OPTIMIZATION STRATEGIES AND MODEL EVALUATION

5.1 Team Optimization with Skyline Algorithm

Table 5.1 showcases a player selection based on three key attributes—Minutes (Min), Defensive Rating (DEF), and Regularized Plus-Minus (RPM) utilizing the dynamic programming approach with Skyline Dominates algorithm. The algorithm systematically evaluated different combinations of players, considering constraints such as budget and time, and selected a subset of players that formed the skyline concerning the specified attributes. In this context, the skyline represents a set of non-dominated solutions, providing a diverse selection of players with superior performance in the considered metrics.

TABLE 5.1
PLAYER SELECTION BASED ON 3 ATTRIBUTES BY USING THE SKYLINE ALGORITHM

Name	S/B	Position	Salary ($)	Team	Min	DEF	RPM	PER
Nikola Jokic	S	C	31579390	DEN	33.5	108.9	32.94	11.78
Jayson Tatum	S	SF	28103500	BOS	35.9	103.4	21.87	8.96
Luka Doncic	S	PG	10174391	DAL	35.4	110.3	25.13	6.50
Desmond Bane	S	SG	2033160	MEM	29.8	107.8	17.62	5.43
Bobby Portis	S	PF	4347600	MIL	28.2	108.5	17.79	2.34
Brandon Clarke	B	PF	4347600	MEM	19.5	105.7	23.75	-0.68
JaVale McGee	B	C	4347600	PHX	15.8	104.6	22.65	-0.83
Gary Payton II	B	PG	1669178	GSW	17.6	102.3	17.86	0.86
Otto Porter Jr.	B	SF	1669178	GSW	22.2	103.3	15.94	0.36
Jordan Poole	B	SG	2161440	GSW	30.0	105.5	16.20	2.16

The selected players in the Table are highlighted for their noteworthy contributions across the chosen attributes. For instance, Nikola Jokic, positioned as a center (C) for the

Denver Nuggets (DEN), demonstrated an exceptional RPM of 32.94, underlining his impact on the court. Similarly, Jayson Tatum, playing as a small forward (SF) for the Boston Celtics (BOS), exhibited a well-rounded performance with a high DEF of 103.4 and a competitive RPM of 21.87. These selections balanced offensive and defensive capabilities, maximizing the team's overall performance within the defined constraints. As the weights were already generated with the Lasso, Ridge, and Logistic machine learning algorithms, those weights were used to predict the winning rate. The team winning rate was 93% with logistic weights, 93.58% with lasso weights, 94.78% for ridge weights, and the total salary was $8,9464,717 USD.

TABLE 5.2
PLAYER SELECTION BASED ON 15 ATTRIBUTES BY USING THE SKYLINE ALGORITHM

Name	S/B	P	Salary($)	Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEFRTG	RPM
Joel Embiid	S	C	31579390	PHI	3.7	1.4	9.8	19.6	9.6	11.8	2.1	9.6	4.2	3.1	1.1	1.5	2.7	107.8	9.83
Luka Doncic	S	PG	10174391	DAL	8.8	3.1	9.9	21.6	5.6	7.5	0.9	8.3	8.7	4.5	1.2	0.6	2.2	110.3	6.50
Jayson Tatum	S	SF	28103500	BOS	8.6	3.0	9.3	20.6	5.3	6.2	1.1	6.9	4.4	2.9	1.0	0.6	2.3	103.4	8.96
Desmond Bane	S	SG	2033160	MEM	6.9	3.0	6.7	14.5	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Bobby Portis	S	PF	4347600	MIL	4.7	1.8	5.8	12.1	1.2	1.6	2.5	6.6	1.2	1.3	0.7	0.7	2.4	108.5	2.34
Jordan Poole	B	SG	2161440	GSW	7.6	2.8	6.2	13.9	3.2	3.5	0.4	3.0	4.0	2.5	0.8	0.3	2.7	105.5	2.16
JaVale McGee	B	C	5000000	PHX	0.1	0.0	3.9	6.2	1.4	2.0	2.2	4.5	0.6	1.3	0.3	1.1	2.4	104.6	-0.83
Obi Toppin	B	PF	5105160	NYK	2.3	0.7	3.5	6.6	1.3	1.7	1.0	2.8	1.1	0.8	0.3	0.5	1.4	104.0	-0.29
Rudy Gay	B	SF	5890000	UTA	3.7	1.3	2.9	6.9	1.1	1.4	1.0	3.4	1.0	0.9	0.5	0.3	1.7	104.9	-2.56
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86

Table 5.2 presents the results of player selection based on the Skyline dynamic programming algorithm, considering a comprehensive set of 15 attributes to maximize team success. The attributes included key performance metrics such as three-point attempts (3PA), field goals made (FGM), free throws made (FTM), offensive rebounds (OREB), defensive rebounds (DREB), assists (AST), and various defensive indicators like steals (STL), blocks (BLK), and defensive rating (DEFRTG). The Lasso algorithm effectively

optimized player choices, striking a balance between offensive and defensive contributions.

The selected roster reflected a strategic combination of star players and supporting roles, each contributing significantly to the team's overall performance. Notably, Joel Embiid, positioned as a center (C) for the Philadelphia 76ers (PHI), stood out with an impressive RPM of 9.83, showcasing his dominance on both ends of the court. The team's winning rate is 0.92 with logistics, 0.94 with the lasso, and 0.95 with the ridge weights.

Furthermore, the total team salary of $96,063,819 and the cumulative playing time of 262.5 minutes aligned with practical considerations for team management. This comprehensive approach to player selection, combining offensive and defensive skills while considering financial and time constraints, contributed valuable insights to sports analytics.

5.2 Team Optimization with Traditional Statistics

5.2.1 Strategy to Maximize Team Winning

The strategy employed to maximize team winning percentage in this research was formulated as an integer linear programming (ILP) problem, addressing the inherent complexity of player selection under budget constraints. The objective function, as outlined in equations 5.1 and 5.2, was derived from the team winning rate formula. The goal was to select players efficiently to achieve the highest team winning percentage while operating within predefined financial limitations.

$$\begin{aligned} maximize \quad & w_0 + \sum_{i=1}^{5n} \sum_{j=1}^{15} (S_i \times P_{ij} \times w_j) + \\ & \sum_{i=5n+1}^{10n} \sum_{j=1}^{15} (S_i \times P_{ij} \times w_{j+14}) \end{aligned} \tag{5.1}$$

$$s.t. \quad \sum_{i=1}^{10n}(Salary_i \times S_i) \leq SalaryCap \tag{5.2}$$

$$\sum_{i=1}^{n} S_i = 1 \tag{5.3}$$

$$\sum_{i=n+1}^{2n} S_i = 1 \tag{5.4}$$

$$\sum_{i=2n+1}^{3n} S_i = 1 \tag{5.5}$$

$$\sum_{i=3n+1}^{4n} S_i = 1 \tag{5.6}$$

$$\sum_{i=4n+1}^{5n} S_i = 1 \tag{5.7}$$

$$\sum_{i=5n+1}^{6n} S_i = 1 \tag{5.8}$$

$$\sum_{i=6n+1}^{7n} S_i = 1 \tag{5.9}$$

$$\sum_{i=7n+1}^{8n} S_i = 1 \tag{5.10}$$

$$\sum_{i=8n+1}^{9n} S_i = 1 \tag{5.11}$$

$$\sum_{i=9n+1}^{10n} S_i = 1 \tag{5.12}$$

$$S_i \in \{0, 1\} \tag{5.13}$$

Here, w is the notation of weights generated by machine learning algorithms. The player selection is divided into 5 starters (i = 1 to 5n) and 5 bench (i = 5n + 1 to 10n) players. S is the salary of i^{th} player, and P is the j^{th} attribute of i^{th} player. The integer linear programming nature of the problem implies that it is theoretically NP-hard, underscoring the computational challenges associated with optimizing player selections for maximum team success. To tackle this problem, the research leveraged the Python PuLP library for coding and computation. This strategic approach integrated mathematical optimization techniques into the player selection process, demonstrating a systematic and algorithmic methodology to address the complexities of assembling a winning team in professional basketball.

TABLE 5.3

PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	P	Salary ($)	Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Trae Young	S	PG	8326471	ATL	8.0	3.1	9.4	20.3	6.6	7.3	0.7	3.1	9.7	4.0	0.9	0.1	1.7	114.9	6.95
Robert Williams III	S	C	3661976	BOS	0.0	0.0	4.4	6.0	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Jimmy Butler	S	SG	36016200	MIA	2.0	0.5	7.0	14.5	6.9	8.0	1.8	4.1	5.5	2.1	1.6	0.5	1.5	108.4	4.19
Giannis Antetokounmpo	S	PF	39344970	MIL	3.6	1.1	10.3	18.6	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Mikal Bridges	S	SF	5557725	PHX	3.8	1.4	5.6	10.5	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Caleb Martin	B	SF	1782621	MIA	2.6	1.1	3.5	6.8	1.2	1.6	1.2	2.7	1.1	0.9	1.0	0.5	1.7	106.2	-1.34
Montrezl Harrell	B	C	9720900	CHA	0.2	0.1	5.0	7.8	3.0	4.2	2.1	4.0	2.0	1.0	0.4	0.6	1.9	108.5	-0.77
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Amir Coffey	B	SG	153488	LAC	3.7	1.4	3.1	6.8	1.5	1.7	0.4	2.5	1.8	0.7	0.6	0.2	1.3	111.0	-1.29
Brandon Clarke	B	PF	2726880	MEM	0.3	0.1	4.5	7.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68

Table 5.3 presents the outcome of player selection utilizing a threshold salary of $123,655,000. The cumulative salary of the selected players was reported as $108,960,409. The objective function value, which signifies the winning rate, was documented as 1.129 when employing the Lasso regression method and 1.14 with the Ridge regression technique. It is worth noting that the winning rate exceeding one was a consequence of the inherent characteristics of Lasso and Ridge regression models. Although normalization could be applied to these values, the overarching objective remained to achieve high predictive accuracy for the winning rate using these regression methodologies.

TABLE 5.4

PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (RIDGE)

Name	S/B	P	Salary ($)	Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Robert Williams III	S	C	3661976	BOS	0.0	4.4	6.0	0.0	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Jimmy Butler	S	SG	36016200	MIA	2.0	7.0	14.5	0.5	6.9	8.0	1.8	4.1	5.5	2.1	1.6	0.5	1.5	108.4	4.19
Giannis Antetokounmpo	S	PF	39344970	MIL	3.6	10.3	18.6	1.1	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Tyrese Maxey	S	PG	2602920	PHI	4.1	6.4	13.3	1.8	2.8	3.3	0.3	2.9	4.3	1.2	0.7	0.4	2.1	109.7	3.43
Mikal Bridges	S	SF	5557725	PHX	3.8	5.6	10.5	1.4	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Montrezl Harrell	B	C	9720900	CHA	0.2	5.0	7.8	0.1	3.0	4.2	2.1	4.0	2.0	1.0	0.4	0.6	1.9	108.5	-0.77
Gary Payton II	B	PG	1669178	GSW	1.7	3.0	4.8	0.6	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Luke Kennard	B	SG	13347727	LAC	6.0	4.1	9.1	2.7	1.0	1.1	0.3	3.0	2.1	0.9	0.6	0.1	1.4	106.6	-2.86
Brandon Clarke	B	PF	2726880	MEM	0.3	4.5	7.0	0.1	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68
Cameron Johnson	B	SF	4437000	PHX	5.9	4.2	9.2	2.5	1.5	1.7	0.6	3.5	1.5	0.7	0.9	0.2	1.7	104.6	-0.59

Table 5.4 presents the player selection outcomes for the current season, where the threshold salary remained consistent at \$123,655,000. The aggregated salary for the chosen players amounted to \$119,085,476, while the objective function value, denoting the winning rate, registered at 1.14 with the application of the Ridge regression method.

TABLE 5.5
PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	P	Salary (\$)	Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Trae Young	S	PG	8326471	ATL	8.0	3.1	9.4	20.3	6.6	7.3	0.7	3.1	9.7	4.0	0.9	0.1	1.7	114.9	6.95
Robert Williams III	S	C	3661976	BOS	0.0	0.0	4.4	6.0	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Jimmy Butler	S	SG	36016200	MIA	2.0	0.5	7.0	14.5	6.9	8.0	1.8	4.1	5.5	2.1	1.6	0.5	1.5	108.4	4.19
Giannis Antetokounmpo	S	PF	39344970	MIL	3.6	1.1	10.3	18.6	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Mikal Bridges	S	SF	5557725	PHX	3.8	1.4	5.6	10.5	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Montrezl Harrell	B	C	9720900	CHA	0.2	0.1	5.0	7.8	3.0	4.2	2.1	4.0	2.0	1.0	0.4	0.6	1.9	108.5	-0.77
Otto Porter Jr.	B	SF	2389641	GSW	3.4	1.3	3.1	6.6	0.8	1.0	1.4	4.4	1.5	0.6	1.1	0.5	1.3	103.3	0.36
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Luke Kennard	B	SG	13347727	LAC	6.0	2.7	4.1	9.1	1.0	1.1	0.3	3.0	2.1	0.9	0.6	0.1	1.4	106.6	-2.86
Brandon Clarke	B	PF	2726880	MEM	0.3	0.1	4.5	7.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68

The player selection model presented in Table 5.5 employed Logistic regression to optimize team performance based on 15 key attributes. Under a salary cap constraint of \$123655000, the selected lineup exhibited a total salary of \$122,761,668, showcasing effective financial management. The logistic regression model achieved a team winning rate of 94.2%, emphasizing the balance between player attributes and salary considerations. These results underscored the utility of logistic regression in strategic player selection for basketball teams, providing a valuable framework for teams seeking an optimal blend of performance and financial efficiency.

TABLE 5.6
PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	Position	Salary ($)	Team	DEF	RPM	PER
Trae Young	S	PG	8326471	ATL	114.9	6.95	25.48
Jayson Tatum	S	SF	28103500	BOS	103.4	8.96	21.87
Giannis Antetokounmpo	S	PF	39344970	MIL	107.9	8.18	32.12
Nikola Jokic	S	C	31579390	DEN	108.9	11.78	32.94
Desmond Bane	S	SG	2033160	MEM	107.8	5.43	17.62
Andre Drummond	B	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	B	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	B	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	B	PG	1669178	GSW	102.3	0.86	17.86
Isaiah Roby	B	PF	1782621	OKC	113.4	0.09	18.35

Table 5.6 represents the team selection for Lasso based on 3 attributes: PER, RPM, and DEFRTG. The total salary for the team was $119,059,549, and the objective value was 1.29.

TABLE 5.7
PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (RIDGE)

Name	S/B	Position	Salary ($)	Team	DEF	RPM	PER
Trae Young	S	PG	8326471	ATL	114.9	6.95	25.48
Jayson Tatum	S	SF	28103500	BOS	103.4	8.96	21.87
Giannis Antetokounmpo	S	PF	39344970	MIL	107.9	8.18	32.12
Nikola Jokic	S	C	31579390	DEN	108.9	11.78	32.94
Desmond Bane	S	SG	2033160	MEM	107.8	5.43	17.62
Obi Toppin	B	PF	5105160	NYK	104.0	-0.29	18.4
Andre Drummond	B	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	B	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	B	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	B	PG	1669178	GSW	102.3	0.86	17.86

In Table 5.7, the total salary for ridge team selection was $122,382,088, and the objective value was 1.30.

TABLE 5.8
PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Trae Young	S	PG	8326471	ATL	114.9	6.95	25.48
Jayson Tatum	S	SF	28103500	BOS	103.4	8.96	21.87
Giannis Antetokounmpo	S	PF	39344970	MIL	107.9	8.18	32.12
Nikola Jokic	S	C	31579390	DEN	108.9	11.78	32.94
Desmond Bane	S	SG	2033160	MEM	107.8	5.43	17.62
Andre Drummond	B	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	B	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	B	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	B	PG	1669178	GSW	102.3	0.86	17.86
Isaiah Roby	B	PF	1782621	OKC	113.4	0.09	18.35

In leveraging Logistic regression for player selection based on Defensive Rating (DEF), Real Plus-Minus (RPM), and Player Efficiency Rating (PER), Table 5.8 revealed an adept optimization of team composition under a salary cap threshold of $123,655,000. With a total salary of $119,059,549, the model attained an impressive team-winning rate of 97.1% and a minimized objective function value of 3.49. Notably, including high-salaried players such as Jayson Tatum, Giannis Antetokounmpo, and Nikola Jokic, it underscored the delicate balance between financial constraints and the impactful contributions of top-tier athletes to overall team success. This highlighted the nuanced decision-making in crafting an optimal roster that maximized performance and budgetary considerations.

5.2.2 Strategy to Minimize Team Salary

The strategy to minimize team salary was designed to address teams' specific needs, prioritizing cost savings as their primary objective. This approach formulated an integer linear programming problem where the goal was to minimize the overall team expenditure, all while ensuring a predefined level of team performance. Like the initial strategy, this optimization model encompassed constraints from 5.3 to 5.13. However, a pivotal divergence lay in the objective function, which was replaced by the minimization of team salary, as articulated in Equation 5.14. Constraint (5.2) was substituted with a new inequality constraint (5.15) to uphold a predetermined bottom winning rate, set in this instance as the 81.7% winning rate of the 2016-17 national championship team, the Golden State Warriors.

$$minimize \quad \sum_{i=1}^{10n}(Salary_i \times S_i) \tag{5.14}$$

$$\begin{aligned} s.t. \quad & w_0 + \sum_{i=1}^{5n}\sum_{j=1}^{15}(S_i \times P_{ij} \times w_{ij}) + \\ & \sum_{i=5n+1}^{10n}\sum_{j=1}^{15}(S_i \times P_{ij} \times w_{j+15}) \geq WinThreshold \end{aligned} \tag{5.15}$$

For logistic regression, the left side of constraint (5.15) signifies y instead of Win%, prompting an adjustment in the Win Threshold formula.

$$\ln \frac{WinThreshold}{1 - WinThreshold}$$

Incorporating pre-defined starter/bench status for players in the new team ensured position and status. This strategic approach offers teams a systematic method to curtail expenses while carefully balancing the trade-off between financial constraints and maintaining a competitive performance level, catering to organizations primarily focusing on financial efficiency in the competitive sports landscape.

TABLE 5.9
PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MINIMIZE THE SALARY (LASSO)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Robert Williams III	S	C	3661976	BOS	0.0	0.0	4.4	6.0	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Tyrese Maxey	S	PG	2602920	PHI	4.1	1.8	6.4	13.3	2.8	3.3	0.3	2.9	4.3	1.2	0.7	0.4	2.1	109.7	3.43
Desmond Bane	S	SG	2033160	MEM	6.9	3.0	6.7	14.5	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Jarred Vanderbilt	S	PF	4050000	MIN	0.2	0.0	2.9	4.9	1.2	1.8	2.9	5.5	1.3	1.0	1.3	0.6	2.4	110.5	0.89
Mikal Bridges	S	SF	5557725	PHX	3.8	1.4	5.6	10.5	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Max Strus	B	SF	1669178	MIA	6.5	2.7	3.7	8.3	0.6	0.8	0.4	2.6	1.4	0.8	0.4	0.2	1.7	108.1	-0.58
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Amir Coffey	B	SG	153488	LAC	3.7	1.4	3.1	6.8	1.5	1.7	0.4	2.5	1.8	0.7	0.6	0.2	1.3	111.0	-1.29
Brandon Clarke	B	PF	2726880	MEM	0.3	0.1	4.5	7.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68
Hassan Whiteside	B	C	1669178	UTA	0.0	0.0	3.3	5.1	1.6	2.6	2.6	5.0	0.4	0.8	0.3	1.6	2.8	109.0	-1.08

In the Lasso regression model Table 5.9, the total salary amounted to $25,793,683, achieving a team winning rate of 95.84%. Notable selections such as Robert Williams III, Desmond Bane, and Mikal Bridges highlighted the effectiveness of the Lasso regression in cost-effective player choices.

Similarly, the Ridge regression model Table 5.10, also yielded a total salary of $25,793,683, maintaining a high team winning rate of 95.84%. Key players like Tyrese Maxey, Desmond Bane, and Mikal Bridges continued to demonstrate the efficacy of the Ridge model in achieving a balance between team performance and financial constraints.

TABLE 5.10
PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MINIMIZE THE SALARY (RIDGE)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Robert Williams III	S	C	3661976	BOS	0.0	0.0	4.4	6.0	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Tyrese Maxey	S	PG	2602920	PHI	4.1	1.8	6.4	13.3	2.8	3.3	0.3	2.9	4.3	1.2	0.7	0.4	2.1	109.7	3.43
Desmond Bane	S	SG	2033160	MEM	6.9	3.0	6.7	14.5	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Jarred Vanderbilt	S	PF	4050000	MIN	0.2	0.0	2.9	4.9	1.2	1.8	2.9	5.5	1.3	1.0	1.3	0.6	2.4	110.5	0.89
Mikal Bridges	S	SF	5557725	PHX	3.8	1.4	5.6	10.5	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Max Strus	B	SF	1669178	MIA	6.5	2.7	3.7	8.3	0.6	0.8	0.4	2.6	1.4	0.8	0.4	0.2	1.7	108.1	-0.58
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Amir Coffey	B	SG	153488	LAC	3.7	1.4	3.1	6.8	1.5	1.7	0.4	2.5	1.8	0.7	0.6	0.2	1.3	111.0	-1.29
Brandon Clarke	B	PF	2726880	MEM	0.3	0.1	4.5	7.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68
Hassan Whiteside	B	C	1669178	UTA	0.0	0.0	3.3	5.1	1.6	2.6	2.6	5.0	0.4	0.8	0.3	1.6	2.8	109.0	-1.08

TABLE 5.11
PLAYER SELECTION BASED ON 15 ATTRIBUTES TO MINIMIZE THE SALARY (LOGISTIC)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
John Collins	S	PF	2686560	ATL	3.6	1.4	8.6	14.8	2.9	3.7	2.8	7.3	1.5	1.8	0.8	1.6	3.4	112.1	1.6
Jarrett Allen	S	C	2376840	BKN	0.1	0.0	4.3	6.6	2.4	3.9	3.1	6.5	1.6	1.1	0.6	1.3	2.3	109.6	1.64
Duncan Robinson	S	SF	1416852	MIA	8.3	3.7	4.4	9.4	0.9	1.0	0.1	3.0	1.4	1.0	0.5	0.3	2.6	106.2	3.83
James Harden	S	SG	38199000	HOU	12.4	4.4	9.9	22.3	10.2	11.8	1.0	5.5	7.5	4.5	1.8	0.9	3.3	108.2	7.49
Damian Lillard	S	PG	29802321	POR	10.2	4.1	9.5	20.4	7.0	7.8	0.5	3.8	8.0	2.9	1.1	0.3	1.7	114.3	3.14
Christian Wood	B	PF	1645357	DET	2.3	0.9	4.6	8.2	2.9	3.9	1.7	4.6	1.0	1.4	0.5	0.9	1.6	108.8	3.53
George Hill	B	PG	9133907	MIL	3.0	1.4	3.3	6.4	1.4	1.7	0.8	2.2	3.1	1.0	0.8	0.1	1.4	102.1	1.37
Mitchell Robinson	B	C	1599712	NYK	0.0	0.0	4.1	5.6	1.4	2.4	3.0	4.0	0.6	0.6	0.9	2.0	3.2	109.1	0.13
Rondae Hollis-Jefferson	B	SF	2500000	TOR	0.4	0.1	2.6	5.5	1.8	2.4	1.8	2.9	1.8	1.0	0.8	0.4	1.9	104.3	-0.15
Seth Curry	B	SG	7461380	DAL	5.0	2.3	4.4	9.0	1.3	1.5	0.4	1.8	1.9	1.0	0.6	0.1	1.8	110.6	0.23

The logistic regression model Table 5.11 minimized team salary to $115,858,013 while achieving a team winning rate of 94.08%. Star players such as James Harden, Damian Lillard, and Christian Wood emerged as noteworthy selections in this Table, highlighting the logistic regression's capability to identify high-impact players. In summary, each regression technique offered a unique perspective on the trade-off between cost efficiency and acquiring star players.

TABLE 5.12

PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MINIMIZE SALARY (LASSO)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Jarrett Allen	S	C	2376840	BKN	109.6	1.64	20.69
Duncan Robinson	S	SF	1416852	MIA	106.2	3.83	13.12
Giannis Antetokounmpo	S	PF	25842697	MIL	97.4	10.3	31.94
Luka Doncic	S	PG	7683360	DAL	111.4	4.28	27.65
Damion Lee	S	SG	842327	GSW	109.7	1.1	12.75
Cody Martin	B	SF	1173310	CHA	107.7	-0.74	10.54
Shaquille Harrison	B	PG	1620564	CHI	99.7	2.83	17.81
Christian Wood	B	PF	1645357	DET	108.8	3.53	23.22
Mitchell Robinson	B	C	1599712	NYK	109.1	0.13	23.51
Terence Davis	B	SG	898310	TOR	103.9	1.21	13.88

Table 5.12 illustrates the player selection process focused on minimizing team salary while considering three essential attributes. Notably, Giannis Antetokounmpo emerged as the standout star player within the lineup, positioned as the power forward. Giannis's presence significantly elevated the team's competitive edge despite his substantial salary of $25,842,697, as his impact on the court transcended monetary considerations. With an impressive Defensive Efficiency (DEF) rating of 97.4, Giannis showcased his defensive prowess, fortifying the team's defensive capabilities. The team's total salary amounted to $45,099,329, and despite the emphasis on minimizing expenses, the team maintained a formidable winning rate of 0.952. This underscores the efficacy of the Lasso regression model in optimizing player selection while ensuring competitiveness within budgetary constraints.

TABLE 5.13

PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MINIMIZE SALARY (RIDGE)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Jarrett Allen	S	C	2376840	BKN	109.6	1.64	20.69
Duncan Robinson	S	SF	1416852	MIA	106.2	3.83	13.12
Giannis Antetokounmpo	S	PF	25842697	MIL	97.4	10.3	31.94
Luka Doncic	S	PG	7683360	DAL	111.4	4.28	27.65
Damion Lee	S	SG	842327	GSW	109.7	1.1	12.75
Christian Wood	B	PF	1645357	DET	108.8	3.53	23.22
Furkan Korkmaz	B	SF	1620564	PHI	106.6	0.16	12.08
Terence Davis	B	SG	898310	TOR	103.9	1.21	13.88
Chris Clemons	B	PG	563347	HOU	107.7	1.33	15.05
Nerlens Noel	B	C	1620564	OKC	105.3	0.65	20.45

Table 5.13 presents the player selection process employing the Ridge regression method to minimize team salary while considering essential player attributes. Among the notable players listed, Luka Doncic stands out for his exceptional defensive prowess. With an impressive Defensive Efficiency (DEF) rating of 111.4, Doncic showcased his defensive skills, further solidifying his importance to the team’s success. Despite a total team salary of $44,510,218 and a winning rate of approximately 0.953, the Ridge regression model effectively balanced financial constraints with competitive performance.

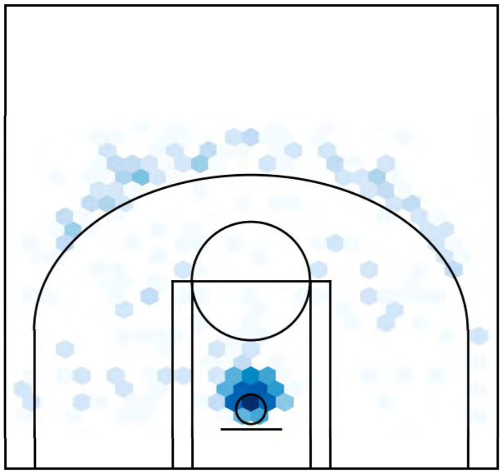
TABLE 5.14
PLAYER SELECTION BASED ON 3 ATTRIBUTES TO MINIMIZE SALARY (LOGISTIC)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Jayson Tatum	S	SF	7830000	BOS	103.5	3.87	20.45
Bam Adebayo	S	C	3454080	MIA	107.7	2.67	20.36
Giannis Antetokounmpo	S	PF	25842697	MIL	97.4	10.3	31.94
Luka Doncic	S	PG	7683360	DAL	111.4	4.28	27.65
James Harden	S	SG	38199000	HOU	108.2	7.49	29.11
Enes Freedom	B	C	4767000	BOS	103.9	1.37	22.48
Shaquille Harrison	B	PG	1620564	CHI	99.7	2.83	17.81
Christian Wood	B	PF	1645357	DET	108.8	3.53	23.22
Donte DiVincenzo	B	SG	2905800	MIL	99.6	2.08	14.09
Mikal Bridges	B	SF	4161000	PHX	108.1	0.93	12.97

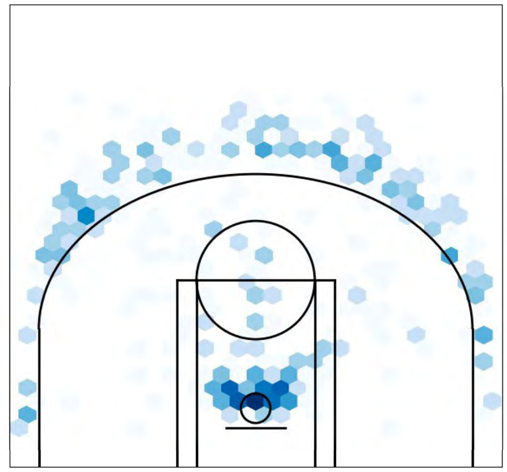
Based on Table 5.14, the player selection process focused on minimizing team salary while considering three key attributes. The total team salary resulting from this selection strategy amounted to $98,108,858. Despite emphasizing the reduction of salary expenses, the team maintained a competitive edge, achieving a winning rate of approximately 0.954.

5.3 Team Optimization with Player Shot Charts and Court Coverage

The NBA, known for its dazzling display of skill and precision, relies heavily on players finding their sweet spots on the court. With cutting-edge technology, the league gathers detailed datasets of NBA players' shooting patterns [27], meticulously recording the (X, Y) coordinates that pinpoint the location of every shot attempted. Leveraging these comprehensive insights, analysts discern the distinctive sweet spots for each player, identifying the zones where their shooting accuracy and effectiveness are maximized. By delving into these intricate shooting patterns, teams and coaches can strategize more effectively, enabling players to capitalize on their strengths and elevate their performance on the hardwood.



(a) Shot charts of LeBron James



(b) Shot charts of Stephen Curry

Fig. 5.1. Shooting patterns for season 2021-2022.

Fig. 5.1 presents shot charts for two prominent NBA players during the 2021- 2022 season, LeBron James and Stephen Curry. In Fig. 5.1a, the shot chart of LeBron James illustrates his shooting patterns, showcasing the areas of the court where he attempted shots throughout the season. Conversely, Fig. 5.1b displays the shot chart of Stephen Curry, providing insights into his shooting tendencies and preferred scoring zones on the court.

The NBA court's dimensions, originally 50 feet in width and 94 feet in height, have been standardized through normalization, scaling the width to a range of -250 to +250 and the height accordingly. The Y-axis begins at -47.5, with the centerline positioned at 422.5, resulting in a sum of 470, which mirrors the calculated midpoint of the court at (94 feet * 10) / 2). This meticulous normalization ensures the basket hoop resides precisely at the origin point (0,0) on the court, facilitating accurate analysis and interpretation of basketball shot data. Fig. 5.2 is the shot chart plotted for all players in session 2021-22. These will be useful for the Base Vector experiment in Section 5.3.2.

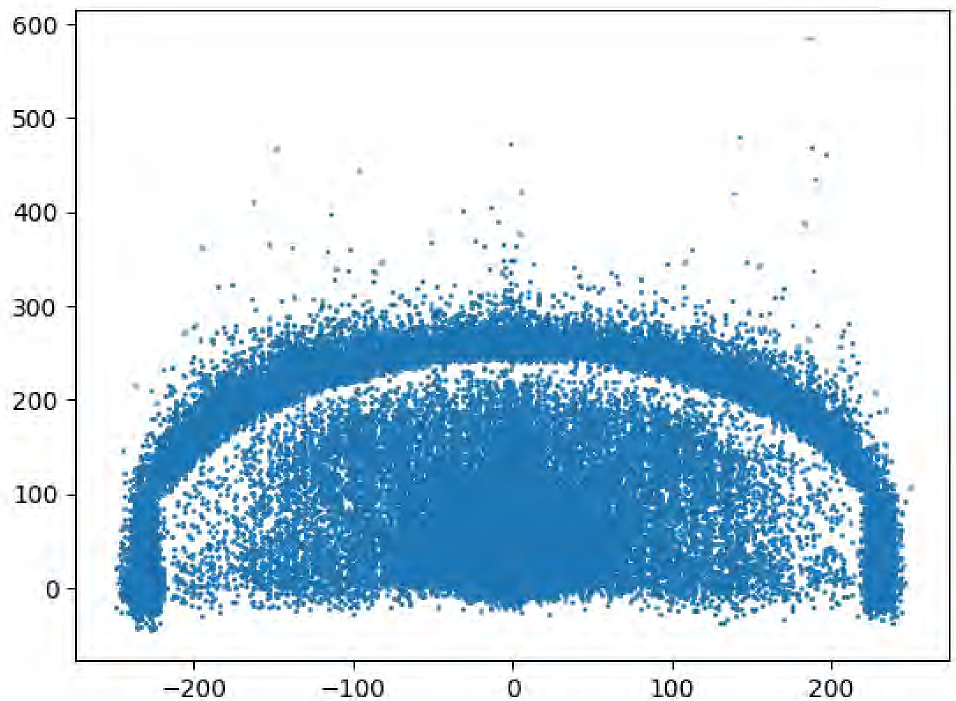


Fig. 5.2. Shot charts for all players in the season 2021-22.

5.3.1 Background

Gaussian process (GP) is leveraged as a powerful probabilistic model for capturing intricate patterns in spatial data. The GP is defined as a stochastic process with normally distributed sample paths, making it particularly suitable for modeling functions $f: \mathbf{X} \rightarrow R$. Here, $f_n \equiv f(x_n)$ corresponds to the evaluation of the function at a specific point $x_n \in X$. The spatial covariance between two points in $\mathbf{X}$ encodes prior beliefs about the function f, allowing to capture the properties such as differentiability, smoothness, and periodicity. The main focus on utilizing the smoothness property to embed an inductive bias into the model, reflecting the assumption that shooting habits exhibit smooth variations across the basketball court.

5.3.1.1 Gaussian Smoothing

Gaussian smoothing is incorporated as a vital component of the methodology. The Gaussian smoothing operator, achieved through 2-D convolution, facilitates the "blurring" of images by removing fine details and noise. This operator employs a Gaussian-shaped kernel, whose standard deviation governs the degree of smoothing. Notably, the Gaussian smoothing operator produces a weighted average of each pixel's neighborhood, emphasizing the central pixels more. This nuanced approach to averaging results in gentler smoothing, preserving edges more effectively than a uniformly weighted mean filter. The choice of the Gaussian filter is justified by its frequency response, as it acts as a lowpass frequency filter, selectively removing high spatial frequency components and enhancing the ability to discern meaningful spatial patterns.

5.3.1.2 Non-Negative Matrix Factorization

The research incorporated Non-Negative Matrix Factorization (NMF) as a dimensionality reduction technique. NMF assumes that a matrix Λ can be approximated by the product of two low-rank matrices W and B, where Λ consists of data points, B comprises basis vectors, and W contains non-negative weight vectors. The reconstruction of each vector was expressed as a linear combination of the basis vectors weighted by the associated non-negative weights. The optimization procedure for determining the optimal matrices W^* and B^* minimized a measure of reconstruction error while enforcing non-negativity constraints on all elements. This integration of the Gaussian process, Gaussian smoothing, and Non-Negative Matrix Factorization formed a comprehensive approach for capturing and understanding spatial patterns in this study.

5.3.2 Team Optimization with Base Vector Experimentation

The initial step in the analysis involved discretizing the basketball court into bins and counting the number of shots made by each player within each bin. Subsequently, the bins were vectorized, resulting in row vectors representing the shot distribution for individual players. To smooth the histograms and transition from a discrete to a continuous representation, the ndimage.filters.gaussianfilter() function was employed, ensuring a refined presentation of shot data.

Normalization of the smoothed histograms was imperative to standardize players, considering varying shot attempts. Using kernel smoothing and normalization, empirical distribution functions were obtained, enabling a consistent comparison across players. Notably, the histograms were smoothed to mitigate noise and highlight shooting patterns, which is essential for the subsequent analyses.

Nonnegative matrix factorization (NMF) was executed on the data following the preprocessing steps. NMF, with its unique non-negativity constraint, facilitated factorization of the matrix Λ into bases comprising non-negative values. The constructed matrix Λ featured columns representing the vectorized smoothed shot densities for each player, resulting in a $N \times V$ matrix, where N is the number of bins, and V is the number of players. With a total of 15,750 bins, Λ captured the nuanced shot count for each player within specific court regions.

Fig. 5.3 presents approximated error for different number of bases. Following the elbow rule, ten was chosen as the base number to minimize the error with efficiency. The factorized representation included non-negative r-base vectors in matrix W, signifying distinct shooting styles, while matrix B contains coefficients representing the importance of each base for

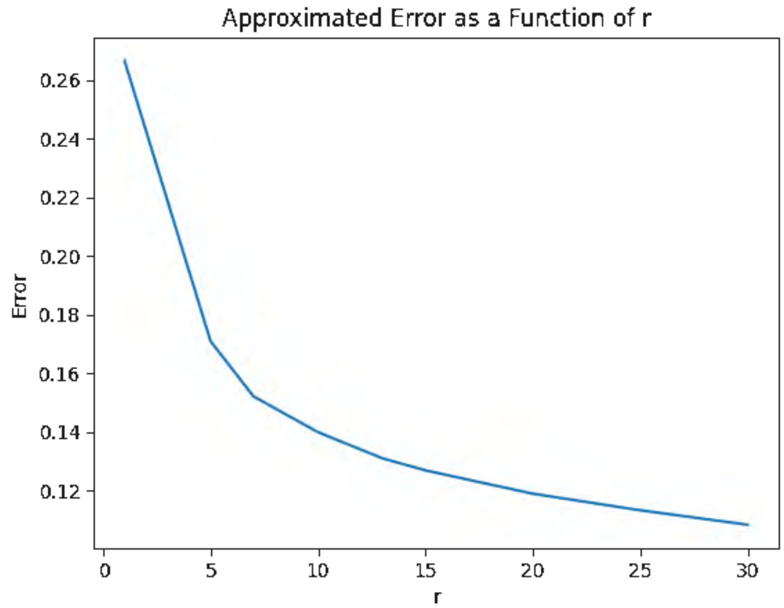


Fig. 5.3. Approximated error for different number of bases.

individual players. These coefficients essentially weighted the influence of different shooting styles, contributing to a comprehensive understanding of each player's shooting tendencies. Determining the optimal number of bases was a critical aspect of the analysis, influencing the granularity and interpretability of the derived shooting styles.

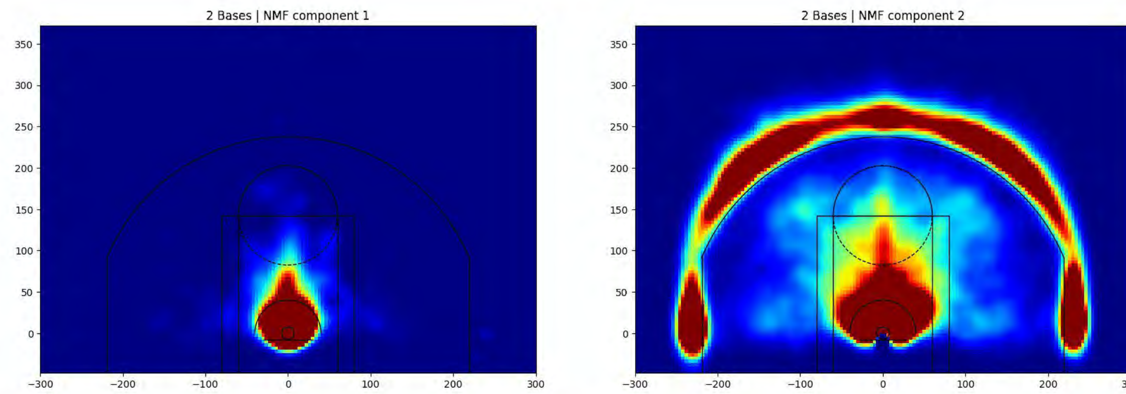
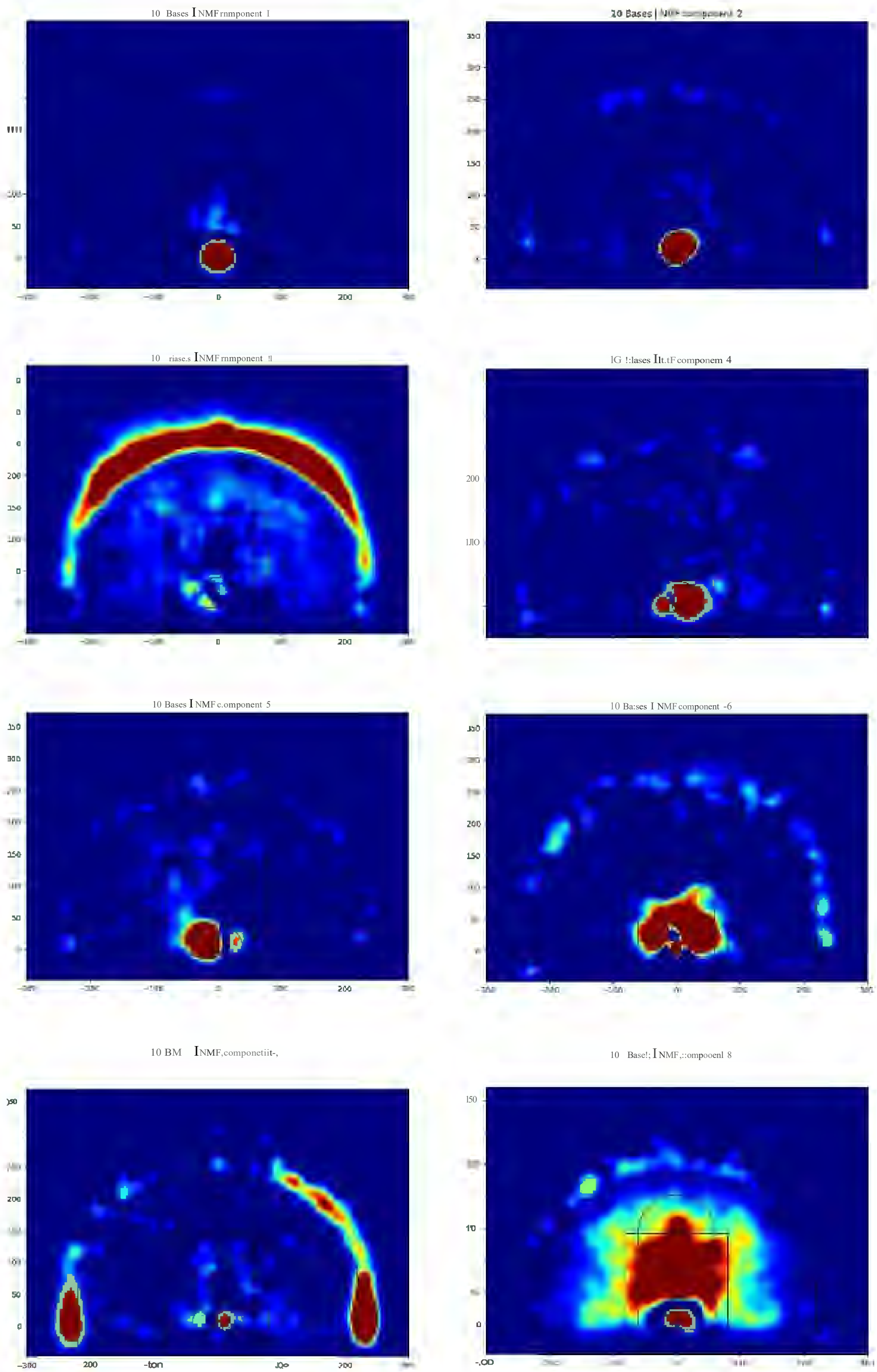


Fig. 5.4. Bases 2 — NMF component.

The crucial task in optimizing the Non-Negative Matrix Factorization (NMF) components was determining an appropriate value for the number of components (r). The analysis revealed that employing insufficient components, such as 2 or 5, resulted in a lack of diversity in the identified shooting modes, implying oversimplification. Conversely, a larger r beyond ten failed to provide substantial new information, leading to duplicated bases and the risk of overfitting. To make an informed decision about r, the approximate error plotted against different r values, identifying a distinct “elbow” at r = 10. This “elbow” is a pivotal point, indicating the optimal r value where further additions cease to enhance the model’s representation significantly. Striking a balance between model complexity and informativeness, this observation guided the selection of an optimal r for a meaningful and concise representation of shooting modes in professional basketball players.



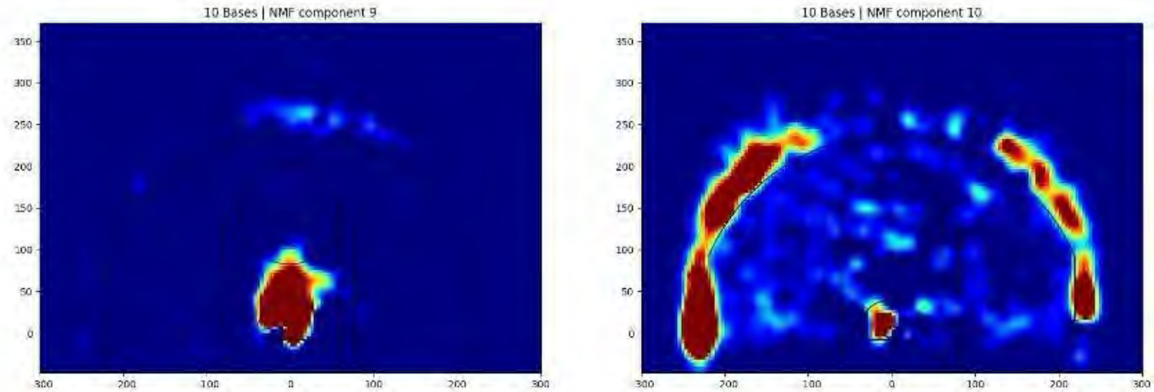


Fig. 5.5. Bases 10 — NMF component.

Incorporating insights from the Non-Negative Matrix Factorization (NMF) analysis, where ten coefficients characterized players' shooting patterns, the player selection model extended to ensure balanced court coverage for the chosen team. Constraints based on the average values of these coefficients promoted equitable distribution of shooting styles across court zones. The team selection process required the sum of each coefficient for the chosen ten players to exceed ten times the average value. This approach guaranteed comprehensive court coverage, emphasizing a balanced representation of shooting patterns. Optimization objectives, whether maximizing winning percentage or minimizing team salary, remain consistent with previous models, while additional constraints were introduced to enhance overall court coverage in player selection. The 1st base coefficient of player i as $Z_i(1)$, the new constraints can be defined as:

$$\sum_{i=1}^{10}(Z_i(1) \times S_i) \geq 10 \times Z_{\text{avg}}(1) \quad (5.16)$$

$$\sum_{i=1}^{10}(Z_i(2) \times S_i) \geq 10 \times Z_{\text{avg}}(2) \quad (5.17)$$

$$\sum_{i=1}^{10}(Z_i(3) \times S_i) \geq 10 \times Z_{\text{avg}}(3) \quad (5.18)$$

$$\sum_{i=1}^{10}(Z_i(4) \times S_i) \geq 10 \times Z_{\text{avg}}(4) \quad (5.19)$$

$$\sum_{i=1}^{10}(Z_i(5) \times S_i) \geq 10 \times Z_{\text{avg}}(5) \quad (5.20)$$

$$\sum_{i=1}^{10}(Z_i(6) \times S_i) \geq 10 \times Z_{\text{avg}}(6) \quad (5.21)$$

$$\sum_{i=1}^{10}(Z_i(7) \times S_i) \geq 10 \times Z_{\text{avg}}(7) \quad (5.22)$$

$$\sum_{i=1}^{10}(Z_i(8) \times S_i) \geq 10 \times Z_{\text{avg}}(8) \quad (5.23)$$

$$\sum_{i=1}^{10}(Z_i(9) \times S_i) \geq 10 \times Z_{\text{avg}}(9) \quad (5.24)$$

$$\sum_{i=1}^{10}(Z_i(10) \times S_i) \geq 10 \times Z_{\text{avg}}(10) \quad (5.25)$$

Table 5.15 presents the ten coefficients corresponding to different shooting styles for star players in the 2021-2022 season, including Kevin Durant, Giannis Antetokounmpo, Nikola Jokic, Stephen Curry, and LeBron James.

TABLE 5.15
COEFFICIENTS BASED ON 10 BASES FOR SOME STAR PLAYER

Player Name	1	2	3	4	5	6	7	8	9	10
Kevin Durant	0.013377	0.006032	0.016061	0.008704	0.003303	0.000000	0.001745	0.027946	0.000000	0.000000
Giannis Antetokounmpo	0.023633	0.078488	0.006980	0.013411	0.015336	0.000809	0.000000	0.007303	0.019799	0.000000
Nikola Jokic	0.020168	0.011755	0.006171	0.012729	0.013798	0.023851	0.000000	0.018851	0.036111	0.000000
Stephen Curry	0.003048	0.009718	0.028253	0.011034	0.015894	0.013281	0.000000	0.004295	0.001823	0.005664
LeBron James	0.019035	0.053619	0.016962	0.019009	0.017918	0.002332	0.000000	0.000000	0.019075	0.001477

5.3.2.1 Team Optimization and Players Selection

Table 5.16 encapsulates the results derived from the player selection model incorporating Lasso regularization. Notable players such as Jayson Tatum and Nikola Jokic contributed to a team composition with a total salary of $94,556,026 and an impressive winning rate of approximately 94.86%. Each player's attributes, defensive metrics, and shooting proficiency across distinct court zones were meticulously considered, resulting in a balanced and competitive team.

TABLE 5.16
PLAYER SELECTION BASED ON 3 ATTRIBUTES AND HOT SPOTS TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	Position	Salary($)	Time	Current Team	DEF	RPM	PER	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7	Zone8	Zone9	Zone10
Jayson Tatum	S	SF	28103500	35.9	BOS	103.4	8.96	21.87	0.058	0.301	0.163	0.126	0.110	0.059	0.024	0.033	0.105	0.020
Bobby Portis	S	PF	4347600	28.2	MIL	108.5	2.34	17.79	0.200	0.249	0.086	0.042	0.0	0.058	0.136	0.129	0.065	0.036
Nikola Jokic	S	C	31579390	33.5	DEN	108.9	11.78	32.94	0.141	0.082	0.043	0.089	0.096	0.166	0.0	0.131	0.252	0.0
Monte Morris	S	PG	8449074	29.9	DEN	110.4	4.55	14.75	0.0	0.041	0.124	0.281	0.187	0.095	0.054	0.097	0.0	0.122
Desmond Bane	S	SG	2033160	29.8	MEM	107.8	5.43	17.62	0.117	0.178	0.258	0.136	0.028	0.056	0.034	0.107	0.018	0.068
Lou Williams	B	SG	5000000	14.3	ATL	109.5	-3.72	12.17	0.136	0.138	0.038	0.016	0.0	0.0	0.079	0.295	0.074	0.225
Troy Brown Jr.	B	SF	5170564	16.0	CHI	111.7	-5.09	10.16	0.287	0.072	0.041	0.161	0.009	0.087	0.197	0.004	0.039	0.102
Goga Bitadze	B	C	3098400	14.6	IND	116.4	-0.1	18.06	0.241	0.234	0.060	0.121	0.194	0.038	0.004	0.0	0.108	0.0
Obi Toppin	B	PF	5105160	17.1	NYK	104.0	-0.29	18.4	0.445	0.171	0.001	0.106	0.153	0.0	0.033	0.0	0.011	0.081
Gary Payton II	B	PG	1669178	17.6	GSW	102.3	0.86	17.86	0.140	0.156	0.0	0.046	0.303	0.103	0.099	0.0	0.116	0.038

Table 5.17 showcases outcomes obtained by implementing Ridge regularization in the player selection model. Featured players like Darius Garland and Giannis Antetokounmpo contributed to a team with a total salary of $99,775,759 and an enhanced winning rate of around 95.52%. The model emphasized defensive capabilities, RPM, PER, and shooting accuracy in various zones, ensuring a comprehensive evaluation of player performance and strategic contributions.

TABLE 5.17
PLAYER SELECTION BASED ON 3 ATTRIBUTES AND HOT SPOTS TO MAXIMIZE THE TEAM WINNING (RIDGE)

Name	S/B	Position	Salary ($)	Time	Current Team	DEF	RPM	PER	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7	Zone8	Zone9	Zone10
Darius Garland	S	PG	7040880	35.7	CLE	106.9	4.88	19.04	0.2566	0.0756	0.1389	0.1250	0.0408	0.0200	0.0265	0.1941	0.0751	0.0485
Giannis Antetokounmpo	S	PF	39344970	32.9	MIL	107.9	8.18	32.12	0.1426	0.4735	0.0421	0.0809	0.0925	0.0049	0.0	0.0441	0.1194	0.0
Nikola Jokic	S	C	31579390	33.5	DEN	108.9	11.78	32.94	0.1406	0.0820	0.0430	0.0887	0.0962	0.1663	0.0	0.1314	0.2518	0.0
Jae'Sean Tate	S	SF	1517981	26.4	HOU	114.2	2.08	14.36	0.0868	0.2097	0.0355	0.0426	0.4417	0.0	0.0381	0.0134	0.1251	0.0071
Desmond Bane	S	SG	2033160	29.8	MEM	107.8	5.43	17.62	0.1174	0.1776	0.2578	0.1356	0.0281	0.0559	0.0344	0.1068	0.0181	0.0684
Lou Williams	B	SG	5000000	14.3	ATL	109.5	-3.72	12.17	0.1360	0.1376	0.0380	0.0162	0.0	0.0	0.0790	0.2945	0.0735	0.2252
Payton Pritchard	B	PG	2137440	14.1	BOS	107.2	-2.42	14.95	0.0256	0.0775	0.2846	0.1241	0.1145	0.0556	0.1233	0.0158	0.0	0.1789
Dewayne Dedmon	B	C	1669178	15.9	MIA	106.0	-2.5	15.9	0.2289	0.1710	0.0380	0.2040	0.0972	0.1367	0.0110	0.0114	0.1017	0.0
Obi Toppin	B	PF	5105160	17.1	NYK	104.0	-0.29	18.4	0.4445	0.1714	0.0008	0.1056	0.1525	0.0	0.0334	0.0	0.0109	0.0809
Maurice Harkless	B	SF	4347600	18.4	SAC	110.8	-2.92	7.69	0.1851	0.1620	0.0	0.1370	0.0686	0.1099	0.3118	0.0	0.0256	0.0

TABLE 5.18
PLAYER SELECTION BASED ON 3 ATTRIBUTES AND HOT SPOTS TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	Position	Salary ($)	Time	Current Team	DEF	RPM	PER	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7	Zone8	Zone9	Zone10
Giannis Antetokounmpo	S	PF	39344970	32.9	MIL	107.9	8.18	32.12	0.1426	0.4735	0.0421	0.0809	0.0925	0.0049	0.0	0.0441	0.1194	0.0
Nikola Jokic	S	C	31579390	33.5	DEN	108.9	11.78	32.94	0.1406	0.0820	0.0430	0.0887	0.0962	0.1663	0.0	0.1314	0.2518	0.0
Monte Morris	S	PG	8449074	29.9	DEN	110.4	4.55	14.75	0.0	0.0414	0.1236	0.2806	0.1872	0.0948	0.0541	0.0965	0.0	0.1218
Desmond Bane	S	SG	2033160	29.8	MEM	107.8	5.43	17.62	0.1174	0.1776	0.2578	0.1356	0.0281	0.0559	0.0344	0.1068	0.0181	0.0684
Keldon Johnson	S	SF	2145720	31.9	SAS	111.1	1.71	15.26	0.2067	0.0523	0.0979	0.0750	0.1098	0.0943	0.0620	0.1075	0.1184	0.0762
Lou Williams	B	SG	5000000	14.3	ATL	109.5	-3.72	12.17	0.1360	0.1376	0.0380	0.0162	0.0	0.0	0.0790	0.2945	0.0735	0.2252
Goga Bitadze	B	C	3098400	14.6	IND	116.4	-0.1	18.06	0.2411	0.2338	0.0595	0.1205	0.1945	0.0383	0.0041	0.0	0.1081	0.0
Obi Toppin	B	PF	5105160	17.1	NYK	104.0	-0.29	18.4	0.4445	0.1714	0.0008	0.1056	0.1525	0.0	0.0334	0.0	0.0109	0.0809
Gary Payton II	B	PG	1669178	17.6	GSW	102.3	0.86	17.86	0.1396	0.1556	0.0	0.0461	0.3034	0.1027	0.0989	0.0	0.1159	0.0378
Maurice Harkless	B	SF	4347600	18.4	SAC	110.8	-2.92	7.69	0.1851	0.1620	0.0	0.1370	0.0686	0.1099	0.3118	0.0	0.0256	0.0

Table 5.18 revealed insights from the player selection model utilizing Logistic regression. Star players such as Giannis Antetokounmpo and Nikola Jokic shaped a team with a total salary of $102,772,652 and a commendable winning rate of approximately 89.13%. The model incorporated a predefined threshold salary constraint of $112,414,000, demonstrating its adaptability to financial considerations while focusing on player attributes and performance metrics. Collectively, these Tables underscored the effectiveness of the player selection model in constructing competitive basketball teams through a nuanced consideration of player skills, defensive prowess, and strategic contributions across different court zones.

TABLE 5.19
PLAYER SELECTION BASED ON 15 ATTRIBUTES AND HOT SPOTS TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	Position	Salary ($)	Time	Current Team	3PA	FGM	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Jimmy Butler	S	SG	36016200	33.9	MIA	2.0	7.0	6.9	8.0	1.8	4.1	5.5	2.1	1.6	0.5	1.5	108.4	4.19
Giannis Antetokounmpo	S	PF	39344970	32.9	MIL	3.6	10.3	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Mitchell Robinson	S	C	1802057	25.7	NYK	0.0	3.6	1.2	2.5	4.1	4.5	0.5	0.8	0.8	1.8	2.7	111.2	4.25
Monte Morris	S	PG	8449074	29.9	DEN	4.2	5.0	1.0	1.1	0.4	2.7	4.4	1.0	0.7	0.2	1.2	110.4	4.55
Mikal Bridges	S	SF	5557725	34.8	PHX	3.8	5.6	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Lou Williams	B	SG	5000000	14.3	ATL	1.8	2.2	1.2	1.4	0.3	1.3	1.9	0.8	0.5	0.1	0.9	109.5	-3.72
Payton Pritchard	B	PG	2137440	14.1	BOS	3.5	2.3	0.3	0.3	0.5	1.4	2.0	0.6	0.4	0.1	0.9	107.2	-2.42
Otto Porter Jr.	B	SF	2389641	22.2	GSW	3.4	3.1	0.8	1.0	1.4	4.4	1.5	0.6	1.1	0.5	1.3	103.3	0.36
JaVale McGee	B	C	5000000	15.8	PHX	0.1	3.9	1.4	2.0	2.2	4.5	0.6	1.3	0.3	1.1	2.4	104.6	-0.83
Eric Paschall	B	PF	1782621	12.7	UTA	1.9	2.0	1.1	1.5	0.5	1.3	0.6	0.5	0.2	0.1	0.9	111.3	-2.75

Table 5.19 presents the outcomes of a player selection model based on Lasso regularization, focusing on 15 attributes and strategic hotspots to maximize team success. Star players such as Jimmy Butler and Giannis Antetokounmpo contributed to a team with a total salary of $107,479,728 and a commendable winning rate of approximately 92.22%. The model intricately considered various player attributes, including three-point attempts, field goals, free throws, and defensive efficiency, ensuring a comprehensive evaluation of player contributions and strategic effectiveness across different court zones.

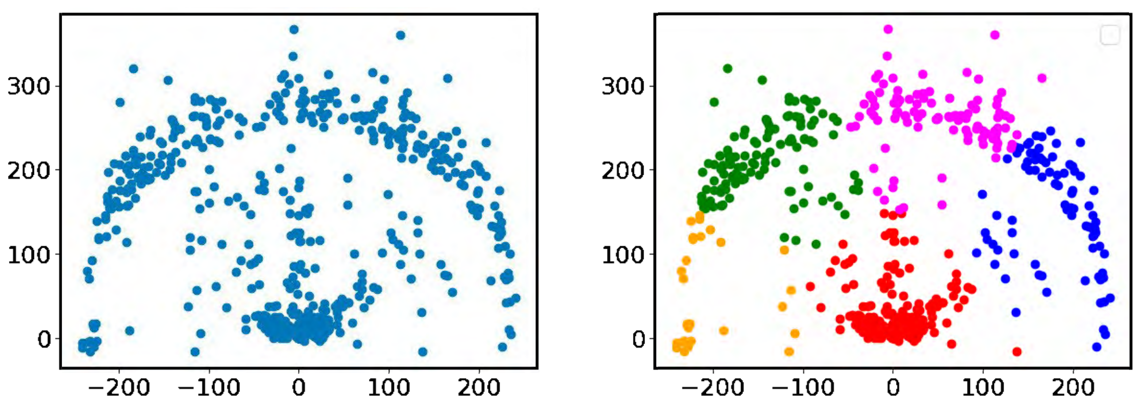
Table 5.20 presents player selection using Logistic regression to optimize team success based on 15 attributes and hotspots. Noteworthy players like Bruce Brown and Giannis Antetokounmpo shaped a team with a total salary of $96,951,123 and a winning rate of approximately 86.14%. The model emphasized player attributes such as three-point attempts, field goals made, free throws made, and defensive efficiency, providing valuable insights into the strategic composition of a competitive team.

TABLE 5.20
PLAYER SELECTION BASED ON 15 ATTRIBUTES AND HOT SPOTS TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	Position	Salary ($)	Time	Current Team	3PA	FGM	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Bruce Brown	S	PG	4736102	24.6	BKN	1.3	3.6	1.3	1.8	1.4	3.4	2.1	0.8	1.1	0.7	2.4	112.0	-0.85
Giannis Antetokounmpo	S	PF	39344970	32.9	MIL	3.6	10.3	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Nikola Jokic	S	C	31579390	33.5	DEN	3.9	10.3	5.1	6.3	2.8	11.0	7.9	3.8	1.5	0.9	2.6	108.9	11.78
Desmond Bane	S	SG	2033160	29.8	MEM	6.9	6.7	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Mikal Bridges	S	SF	5557725	34.8	PHX	3.8	5.6	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Quentin Grimes	B	SG	2168760	17.1	NYK	4.1	2.1	0.3	0.4	0.5	1.5	1.0	0.6	0.7	0.2	1.6	105.0	-2.36
Khem Birch	B	C	6350000	18.0	TOR	0.3	1.8	1.0	1.3	2.2	2.1	1.0	0.5	0.5	0.5	1.9	112.4	-3.75
Gary Payton II	B	PG	1669178	17.6	GSW	1.7	3.0	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Keita Bates-Diop	B	SF	1729217	16.2	SAS	0.9	2.3	0.8	1.0	1.1	2.9	0.7	0.8	0.5	0.2	1.0	106.2	-3.57
Eric Paschall	B	PF	1782621	12.7	UTA	1.9	2.0	1.1	1.5	0.5	1.3	0.6	0.5	0.2	0.1	0.9	111.3	-2.75

Collectively, these tables underscored the versatility of the player selection model in accommodating different regularization techniques while focusing on crucial player attributes and strategic contributions across various court zones.

5.3.3 Team Optimization with Individual Sweet Spots



(a) Shot charts of Stephen Curry (b) Clustering Shot charts of Stephen Curry

Fig. 5.6. Clustering shot charts data.

Fig. 5.6a provides an overview of Curry's shot distribution across the court, illustrating his shooting tendencies and scoring patterns. In contrast, Fig. 5.6b organizes Curry's shot data into five clusters based on similarities in shot location and frequency, providing insights into his preferred scoring areas and offensive strategies.

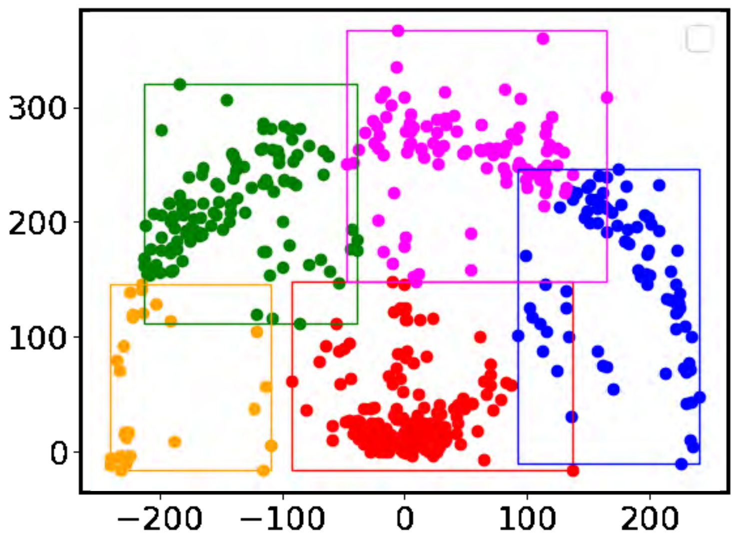


Fig. 5.7. Clustered area of Stephen Curry.

The clustered area of Stephen Curry's shot charts were divided into five distinct clusters represented by rectangles with different colors, as depicted in Fig. 5.7. Each cluster corresponded to a specific region on the basketball court and was marked by a unique color, facilitating the visualization of Curry's court coverage.

The clustering analysis revealed players' shot charts, featuring notable NBA figures like Stephen Curry from the Golden State Warriors (GSW) and LeBron James from the Los Angeles Lakers (LAL), as detailed in Table 5.21. The clusters were determined based on the minimum and maximum values of the X and Y coordinates (MinX, MinY, MaxX, MaxY). Each player had shot charts divided into five clusters, labeled from one to five.

For example, Stephen Curry's shot charts were divided into clusters with respective coordinate ranges (-92, -15, 137.0, 148) for Cluster 1, (-47, 148, 164.0, 366) for Cluster 2, (-212, 112, -39, 320) for Cluster 3, (92, -10, 240, 245) for Cluster 4, and (-240, -16, -109, 146) for Cluster 5 to maximize the court coverage.

TABLE 5.21
CLUSTERING PLAYERS SHOT CHARTS

Players	Team	MinX	MinY	MaxX	MaxY	Cluster	Avg Field Goals
Stephen Curry	GSW	-92	-15	137.0	148	1	6.5667
Stephen Curry	GSW	-47	148	164.0	366	2	6.0196
Stephen Curry	GSW	-212	112	-39	320	3	6.0178
Stephen Curry	GSW	92	-10	240	245	4	5.3333
Stephen Curry	GSW	-240	-16	-109	146	5	3.347
LeBron James	LAL	-70	-29	127	120	1	14.1071
LeBron James	LAL	-221	140	-35	337	2	5.8510
LeBron James	LAL	-29	159	154	297	3	4.1428
LeBron James	LAL	72	-4	235	211	4	3.8000
LeBron James	LAL	-229	-11	-59	131	5	3.7692

5.3.3.1 Modeling and Constraints

Optimizing player selection for efficient court coverage becomes pivotal for team success. In this context, the grid coverage model was introduced as an integral component of a broader strategy. The objective was to maximize the cumulative field goals of selected players on their respective sweet spots, denoted by the binary decision variable A_{ij}, which signifies the assignment of player i to sweet spot j.

$$\text{maximize} \qquad \sum_{i=1}^{P} \sum_{j=1}^{S} A_{ij} \times M_{ij} \tag{5.26}$$

$$\text{s.t.} \qquad \sum_{i=1}^{P} \sum_{j=1}^{S} A_{ij} \leq N * K \tag{5.27}$$

$$\sum_{j=1}^{S} A_{ij} \leq K \tag{5.28}$$

$$\sum_{i=1}^{P} \sum_{j=1}^{S} C_{ijxy} * A_{ij} \leq G_{xy} * N \tag{5.29}$$

$$G_{xy} \leq \sum_{i=1}^{P} \sum_{j=1}^{S} C_{ijxy} * A_{ij} \tag{5.30}$$

$$\sum_{x=1}^{X_{max}} \sum_{y=1}^{Y_{max}} G_{xy} \geq X_{max} * Y_{max} * coverage\_ratio \tag{5.31}$$

$$A_{ij} \in \{0, 1\} \tag{5.32}$$

$$G_{xy} \in \{0, 1\} \tag{5.33}$$

Where S represents the clustered sweet spots of each player (e.g., S = 5), P represents the total number of players, N represents the number of players to be selected (e.g., N = 10), K represents the number of the sweet spots of a player is allowed to use for covering the court (e.g., K = 3), A_{ij} represents the assignment of player i on his sweet spot j, M_{ij} represents the average field goals made by player i on his sweep spot j, G_{xy} represents the court coverage at grid indexed by x-axis x and y-axis; C_{ijxy} represents the court coverage by player i's sweet spot j at grid indexed by x and y, X_{max} represents the max grid value of the court on its x-axis (e.g., 50), Y_{max} represents the max grid value of the court on its y-axis (e.g̅., 70), coverage ratio is the least court grid coverage rate (e.g., 60%).

In the formulation, Equation 5.26 was the objective function and formulated the total game field goals. In Equation 5.27 only N players were selected. In Equation 5.28 each player was only assigned to three of his sweet spots. Equations 5.29 and 5.30 provided that a court grid point was deemed "occupied" if at least one of the selected five players had his assigned sweet spot covering this grid point. It should be noted that this Equation flattened the circular sweet spots into rectangular form to simplify the mathematical optimization process. Equation 5.31 required court coverage be greater than or equal to the defined threshold.

Only A_{ij} and G_{ij} were variables assigned by the LP model. All others in the formulation were constants which were preassigned based on the court dimension, court covering requirement, and the clustering results (e.g., $C_{ijxy} \in \{0, 1\}$)

5.3.3.2 Team Optimization and Players Selection

Examining the Tables, a comprehensive exploration of three regression approaches—Lasso, Ridge, and Logistic were employed for meticulously selecting basketball players to optimize team performance. Table 5.22, governed by the Lasso regression model, unveiled a strategic player selection strategy that prioritized cost-effectiveness. Featuring players like Mitchell Robinson and Tyrese Maxey, the team achieved an impressive winning rate of 91.75% while adhering to a conservative total salary of $23,933,764, well below the specified threshold. Notably, the team's efficiency, measured by the minimum objective function value of 240 minutes, aligned cohesively with the commitment to maintain a robust court coverage of 70%.

TABLE 5.22
PLAYER SELECTION BASED ON 15 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Mitchell Robinson	S	C	1802057	NYK	0.0	0.0	3.6	2.5	1.2	2.5	4.1	4.5	0.5	0.8	0.8	1.8	2.7	111.2	4.25
Tyrese Maxey	S	PG	2602920	PHI	4.1	1.8	6.4	13.3	2.8	3.3	0.3	2.9	4.3	1.2	0.7	0.4	2.1	109.7	3.43
Desmond Bane	S	SG	2033160	MEM	6.9	3.0	6.7	14.5	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Jarred Vanderbilt	S	PF	4050000	MIN	0.2	0.0	2.9	4.9	1.2	1.8	2.9	5.5	1.3	1.0	1.3	0.6	2.4	110.5	0.89
Mikal Bridges	S	SF	5557725	PHX	3.8	1.4	5.6	10.5	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Otto Porter Jr.	B	SF	1669178	GSW	3.4	1.3	3.1	6.6	0.8	1.0	1.4	4.4	1.5	0.6	1.1	0.5	1.3	103.3	0.36
Gary Payton II	B	PG	1669178	GSW	1.7	0.6	3.0	4.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Amir Coffey	B	SG	153488	LAC	3.7	1.4	3.1	6.8	1.5	1.7	0.4	2.5	1.8	0.7	0.6	0.2	1.3	111.0	-1.29
Isaiah Hartenstein	B	C	1669178	LAC	0.4	0.0	3.4	5.2	1.3	1.9	1.7	3.2	2.4	1.2	0.7	1.1	2.5	105.0	-0.87
Brandon Clarke	B	PF	2726880	MEM	0.3	0.1	4.5	7.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68

TABLE 5.23
PLAYER SELECTION BASED ON 15 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (RIDGE)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Giannis Antetokounmpo	S	PF	39344970	MIL	3.6	8.3	10.3	11.4	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Mitchell Robinson	S	C	1802057	NYK	0.0	0.0	3.6	2.5	1.2	2.5	4.1	4.5	0.5	0.8	0.8	1.8	2.7	111.2	4.25
Tyrese Maxey	S	PG	2602920	PHI	4.1	2.8	6.4	3.3	2.8	3.3	0.3	2.9	4.3	1.2	0.7	0.4	2.1	109.7	3.43
Desmond Bane	S	SG	2033160	MEM	6.9	1.8	6.7	2.0	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Keldon Johnson	S	SF	2145720	SAS	5.3	2.4	6.3	3.1	2.4	3.1	1.1	5.0	2.1	1.2	0.8	0.2	2.0	111.1	1.71
Max Strus	B	SF	1669178	MIA	6.5	0.6	3.7	0.8	0.6	0.8	0.4	2.6	1.4	0.8	0.4	0.2	1.7	108.1	-0.58
Gary Payton II	B	PG	1669178	GSW	1.7	0.5	3.0	0.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Amir Coffey	B	SG	153488	LAC	3.7	1.5	3.1	1.7	1.5	1.7	0.4	2.5	1.8	0.7	0.6	0.2	1.3	111.0	-1.29
Isaiah Hartenstein	B	C	1669178	LAC	0.4	1.3	3.4	1.9	1.3	1.9	1.7	3.2	2.4	1.2	0.7	1.1	2.5	105.0	-0.87
Brandon Clarke	B	PF	2726880	MEM	0.3	1.3	4.5	2.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68

Turning attention to Table 5.23, where Ridge regression takes the lead, a distinct set of players, including Giannis Antetokounmpo and Stephen Curry, comes to the forefront. Despite a higher financial investment of $55,816,729, the team excelled with an outstanding winning rate of 96.97%. The consistent minimum objective function value (240.11) underscored the unwavering commitment to team efficiency, while the court coverage remained a formidable 70%.

TABLE 5.24
PLAYER SELECTION BASED ON 15 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	Position	Salary ($)	Current Team	3PA	3PM	FGM	FGA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	DEF	RPM
Robert Williams III	S	C	3661976	BOS	0.0	1.1	4.4	1.5	1.1	1.5	3.9	5.7	2.0	1.0	0.9	2.2	2.2	103.4	4.82
Giannis Antetokounmpo	S	PF	39344970	MIL	3.6	8.3	10.3	11.4	8.3	11.4	2.0	9.6	5.8	3.3	1.1	1.4	3.2	107.9	8.18
Stephen Curry	S	PG	45780966	GSW	11.7	4.3	8.4	4.7	4.3	4.7	0.5	4.7	6.3	3.2	1.3	0.4	2.0	103.4	9.48
Desmond Bane	S	SG	2033160	MEM	6.9	1.8	6.7	2.0	1.8	2.0	0.6	3.8	2.7	1.5	1.2	0.4	2.6	107.8	5.43
Mikal Bridges	S	SF	5557725	PHX	3.8	1.6	5.6	1.9	1.6	1.9	0.9	3.3	2.3	0.8	1.2	0.4	1.8	106.9	3.71
Montrezl Harrell	B	C	9720900	CHA	0.2	3.0	5.0	4.2	3.0	4.2	2.1	4.0	2.0	1.0	0.4	0.6	1.9	108.5	-0.77
Gary Payton II	B	PG	1669178	GSW	1.7	0.5	3.0	0.8	0.5	0.8	1.0	2.5	0.9	0.6	1.4	0.3	1.8	102.3	0.86
Austin Reaves	B	SG	925258	LAL	2.7	1.6	2.4	1.9	1.6	1.9	0.7	2.4	1.8	0.7	0.5	0.3	1.4	109.0	-2.25
Brandon Clarke	B	PF	2726880	MEM	0.3	1.3	4.5	2.0	1.3	2.0	2.1	3.2	1.3	0.5	0.6	1.1	1.9	105.7	-0.68
Cameron Johnson	B	SF	4437000	PHX	5.9	1.5	4.2	1.7	1.5	1.7	0.6	3.5	1.5	0.7	0.9	0.2	1.7	104.6	-0.59

In Table 5.24, Logistic regression introduced a unique perspective to player selection showcasing talents like Robert Williams III and Giannis Antetokounmpo. Despite the highest total salary expenditure among the three methodologies ($115,858,013), the team achieved a notable winning rate of 94.08%. The persistent minimum objective function value of 240.11 reinforced the overarching theme of prioritizing team efficiency. The targeted court coverage of 70% validated Logistic regression's effectiveness in navigating the delicate balance between individual player attributes and court presence, culminating in forming a competitive and well-rounded basketball team.

TABLE 5.25
PLAYER SELECTION BASED ON 3 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (LASSO)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Trae Young	S	PG	8326471	ATL	114.9	6.95	25.48
Miles Bridges	S	SF	5421493	CHA	112.4	4.51	17.97
Nikola Jokic	S	C	31579390	DEN	108.9	11.78	32.94
Desmond Bane	S	SG	2033160	MEM	107.8	5.43	17.62
Tyrese Haliburton	S	PF	4023600	IND	115.9	2.7	18.25
Andre Drummond	B	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	B	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	B	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	B	PG	1669178	GSW	102.3	0.86	17.86
Isaiah Roby	B	PF	1782621	OKC	113.4	0.09	18.35

Table 5.25 displays the outcomes of player selection using the Lasso regression model, emphasizing three critical attributes (DEF, RPM, PER) and considering essential court coverage. This strategic approach yielded a formidable team with a total salary of $61,056,172. The team's impressive winning rate of 99.33%, exceeding the set win

threshold of 92%, underscored the Lasso regression's efficacy in assembling a high-performing roster. Additionally, the team operated comfortably within the specified salary threshold of $123,655,000 while maintaining a robust court coverage of 70 percent.

TABLE 5.26
PLAYER SELECTION BASED ON 3 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (RIDGE)

Name	Lineup Position	Position	Salary ($)	Current Team	DEF	RPM	PER
Miles Bridges	Starter	SF	5421493	CHA	112.4	4.51	17.97
Darius Garland	Starter	PG	7040880	CLE	106.9	4.88	19.04
Nikola Jokic	Starter	C	31579390	DEN	108.9	11.78	32.94
Desmond Bane	Starter	SG	2033160	MEM	107.8	5.43	17.62
Tyrese Haliburton	Starter	PF	4023600	IND	115.9	2.7	18.25
Andre Drummond	Backup	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	Backup	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	Backup	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	Backup	PG	1669178	GSW	102.3	0.86	17.86
Brandon Clarke	Backup	PF	2726880	MEM	105.7	-0.68	23.75

Moving to Table 5.26, the player selection process under the Ridge regression model is presented, showcasing chosen players for both starting and backup lineups. Noteworthy players like Miles Bridges and Nikola Jokic contributed to the team's remarkable objective function value of 102.46, surpassing the winning threshold of 90%. Despite a total salary of $60,714,840, the team balanced financial considerations and on-court performance, adhering to the prescribed salary threshold and achieving 70% court coverage.

TABLE 5.27

PLAYER SELECTION BASED ON 3 ATTRIBUTES AND COURT COVERAGE TO MAXIMIZE THE TEAM WINNING (LOGISTIC)

Name	S/B	Position	Salary ($)	Current Team	DEF	RPM	PER
Trae Young	S	PG	8326471	ATL	114.9	6.95	25.48
Jayson Tatum	S	SF	28103500	BOS	103.4	8.96	21.87
Nikola Jokic	S	C	31579390	DEN	108.9	11.78	32.94
Jaren Jackson Jr.	S	PF	9180560	MEM	106.0	6.14	17.09
Desmond Bane	S	SG	2033160	MEM	107.8	5.43	17.62
Andre Drummond	B	C	1669178	BKN	109.8	1.15	21.04
Jordan Poole	B	SG	2161440	GSW	105.5	2.16	16.2
Otto Porter Jr.	B	SF	2389641	GSW	103.3	0.36	15.94
Gary Payton II	B	PG	1669178	GSW	102.3	0.86	17.86
Isaiah Roby	B	PF	1782621	OKC	113.4	0.09	18.35

Table 5.27 highlights the player selection process under the Logistic regression model, emphasizing the same three attributes and court coverage. The resulting team, with a total salary of $88,895,139, achieved a commendable winning rate of 93.88%. The utilization of Logistic regression demonstrated its effectiveness in navigating the delicate balance between player attributes and court coverage, resulting in a competitive team that aligned with financial constraints and strategic objectives. These findings collectively contributed to a nuanced understanding of player selection strategies based on different regression models, providing valuable insights for optimizing team performance in basketball.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In conclusion, this research conducted an in-depth exploration into the dynamics of NBA player statistics and their profound implications for team performance. Formulating two distinct player signing strategies, focused on maximizing team winning rates or minimizing team salary commitments, represents a novel and insightful approach to talent acquisition in professional basketball. Employing linear, logistic, and ridge regression models, the study integrated 14 crucial individual statistics to predict team winning rates accurately. The strategic optimization problems were translated into integer linear programming challenges, marking a pioneering effort in player selection research under the constraints of salary cap.

The rigorous evaluation of these models and strategies, utilizing data spanning ten NBA regular seasons, underscored their effectiveness in predicting team success and the capability to identify adept players under specified constraints. The groundbreaking nature of this work lies in its pioneering exploration of player selection research within the challenging realm of salary cap limitations. The identified strategies offer valuable insights for team managers and stakeholders seeking to balance performance excellence and fiscal responsibility.

The intersection of cutting-edge machine learning models and traditional basketball statistics represents a pivotal frontier for research in NBA team optimization. Integrating advanced techniques like deep learning or ensemble models offers the potential to extract intricate patterns and dependencies within player data. This approach not only refines player selection but also provides a more nuanced understanding of how metrics like PER, RPM, and DEFRTG collectively influence team dynamics. The synergy of these modern methods with established linear programming, lasso, ridge, and logistic regression models enables researchers to create a comprehensive framework that captures both the quantitative and qualitative aspects of player performance.

Looking ahead, future avenues of research could delve into integrating more sophisticated continuous models tailored explicitly for spatial-temporal data. Such an approach aims to refine the representation of shot patterns, providing a more nuanced understanding of player performance on the court. Additionally, exploring alternative models with enhanced prediction accuracy and the capacity to accommodate broader constraints aligns with the practical business needs of NBA teams. Continuous refinement and expansion of these analytical frameworks promise to yield increasingly insightful perspectives on optimal NBA teaming strategies, ensuring this research's continued relevance and applicability in the dynamic landscape of professional basketball.

REFERENCES

[1] "NBA sets salary cap at $136 million for 2023-24 season." https://www.nba.com/news/nba-sets-salary-cap-at-136-million-for-2023-24-season. Accessed on November 5, 2023.

[2] E. Kiplagat, "Top 15 highest paid NBA players for the 2023/24 season." https://sportsbrief.com/nba/ 44389-top-15-highest-paid-nba-players-202324-season/. Accessed on November 5, 2023.

[3] D. Cabarkapa, M. A. Deane, A. C. Fry, G. T. Jones, D. V. Cabarkapa, N. M. Philipp, and D. Yu, "Game statistics that discriminate winning and losing at the nba level of basketball competition," PLOS ONE, vol. 17, pp. 1–12, 08 2022.

[4] X. Feng, Y. Wang, and T. Xiong, "NBA player salary analysis based on multi- variate regression analysis," Highlights in Science, Engineering and Technology, vol. 49, pp. 157–166, 2023.

[5] L. Grassetti, R. Bellio, L. Di Gaspero, G. Fonseca, and P. Vidoni, "An extended regularized adjusted plus-minus analysis for lineup management in basketball using play-by-play data," IMA Journal of Management Mathematics, vol. 32, pp. 385–409, 10 2020.

[6] D. Oliver, Basketball on paper: rules and tools for performance analysis. U of Nebraska Press, 2011.

[7] K. Goldsberry, Sprawlball: A visual tour of the new era of the NBA. Mariner Books, 2019.

[8] "Calculating player efficiency rating." https://www.basketball-reference.com/about/per.html. Accessed on November 5, 2023.

[9] J. Engelmann, "Possession-based player performance analysis in basketball (adjusted+/–and related concepts)," in Handbook of statistical methods and analyses in sports, pp. 231–244, Chapman and Hall/CRC, 2017.

[10] E. Wong-Toi, H.-C. Yang, W. Shen, and G. Hu, "A joint analysis for field goal attempts and percentages of professional basketball players: Bayesian nonparametric resource," Journal of Data Science, vol. 21, no. 1, pp. 68–86, 2022.

[11] G. Hu, H.-C. Yang, Y. Xue, and D. K. Dey, "Zero-inflated Poisson model with clustered regression coefficients: Application to heterogeneity learning of field goal attempts of professional basketball players," Canadian Journal of Statistics, vol. 51, no. 1, pp. 157–172, 2023.

[12] R. Nagarajan, Y. Zhao, and L. Li, "Effective NBA player signing strategies based on salary cap and statistics analysis," in 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), pp. 138–143, IEEE, 2018.

[13] S. A. Kovalchik, "Player tracking data in sports," Annual Review of Statistics and Its Application, vol. 10, pp. 677–697, 2023.

[14] C. Osken and C. Onay, "Predicting the winning team in basketball: A novel approach," Heliyon, vol. 8, no. 12, 2022.

[15] B. S. Baumer, G. J. Matthews, and Q. Nguyen, "Big ideas in sports analytics and statistical tools for their investigation," WIREs Computational Statistics, vol. 15, no. 6, p. e1612, 2023.

[16] B. J. Coleman, "Identifying the "players" in sports analytics research," Inter- faces, vol. 42, no. 2, pp. 109–118, 2012.

[17] T. Horvat, J. Job, R. Logozar, and C̃. Livada, "A data-driven machine learning algorithm for predicting the outcomes of NBA games," Symmetry, vol. 15, no. 4, 2023.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 58, no. 1,

pp. 267–288, 1996.

[19] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970.

[20] D. R. Cox, "The regression analysis of binary sequences," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 20, no. 2, pp. 215–232, 1958.

[21] G. B. Dantzig, "Linear programming," Operations research, vol. 50, no. 1, pp. 42–47, 2002.

[22] G. B. Dantzig, A. Orden, P. Wolfe, et al., "The generalized simplex method for minimizing a linear form under linear inequality restraints," Pacific Journal of Mathematics, vol. 5, no. 2, pp. 183–195, 1955.

[23] "Players traditional stats." https://www.nba.com/stats/players/ traditional. Accessed on November 9, 2023.

[24] "Hollinger nba player statistics - all players." https://www.espn.com/nba/ hollinger/statistics. Accessed on November 9, 2023.

[25] "NBA real plus-minus." https://www.espn.com/nba/statistics/rpm/_/ sort/DRPM.

Accessed on November 9, 2023.

[26] "NBA player salaries - national basketball association - ESPN" https://www.espn.com/nba/salaries. Accessed on November 9, 2023.

[27] "NBA shot charts." http://nbashotcharts.com. Accessed on November 10, 2023.

CURRICULUM VITA

Shaykh Siddique

Prairie View A&M University
Department of Computer Science
P.O. Box 519, Prairie View, 77446

Phone: (979) 645-0271
Email: shaykhsiddiqee@gmail.com

Education

Prairie View A&M University, USA
Master of Science in Computer Science · (2022 - Present)

East West University, Bangladesh
Bachelor of Science in Computer Science and Engineering · (2020)

Experience

DevOps Engineer (Summer Intern), Hewlett Packard
Enterprise,
Spring, TX (May, 2023 - August, 2023)

Technical Skills

Languages: C, C++, Python, Java

Deep Learning Frameworks: Pytorch, Tensorflow, Keras

Others: AWS, CloudFront, API Gateway, Okta (IAM), CD/CI Pipelines, Git, Docker, Unix Server, Linux, Shell Scripting, TCP/IP Protocols, SQL/MySQL

Publications

Siddique S., Li L., Wang Y. (2024). Teaming Strategy Optimization: An Analysis of NBA Statistics, Shot Charts, and Constraints. Under Preparation, 2024.

Siddique S., Shultana S., Frizell S. (2023). An Enhanced Strategy of Detecting Neurological Disorders from Magnetic Resonance Images Using Deep Learning. (AIKE 2023) (pp. 99-105). IEEE.

Siddique S., Hridoy AA.I., Khushbu S.A., Das A.K. (2022). CVD: An Improved Approach of Software Vulnerability Detection for Object Oriented Programming Languages Using Deep Learning. Lecture Notes in Networks and Systems, vol 559 (Springer Nature).

Siddique S., Ahmed T., Talukder M. R. A. and Uddin M. M. (2020). English to Bangla Machine Translation Using Recurrent Neural Network. International Journal of Future Computer and Communication, 9(2).

Siddique S., Yasmin M., Taher T. B., and Alam M. (2021). The Reliability and Acceptance of Biometric System in Bangladesh: Users Perspective. International Journal of Computer Trends and Technology, vol. 69, no. 6, pp. 1-6, 2021.