Available Online: 20 April 2026 DOI: 10.54254/2977-3903/2026.32828

To what extent can we use statistical models to accurately predict the outcome of basketball games?

Zhijie Yang

Cogdel Cranleigh High School, Wuhan, China

2390981@qq.com

Abstract. Nowadays, basketball prediction is very popular, especially in the NBA, and many people enjoy the process of predicting game outcomes. This study reviews the existing literature to compare the effectiveness of different statistical approaches, including traditional linear, machine learning, and Markov models, to accurately predict the basketball game outcome and handle the potential issues of prediction. And this study summarises that data quality, core players' injury, home advantages and other contextual variables influence the basketball prediction result. Moreover, this research discusses the broad application of basketball prediction in team performance optimisation, gambling and market strategy. In addition, based on the combined research limitations and area development trend, this paper presents that the future has the potential to combine multiple models, integrate emerging data sources and improve model interpretability. These can support the landing application of basketball prediction. In conclusion, this research contributes to the information of basketball analytics and provides useful insights for the fans, coaches, team managers and sports prediction researchers.

Keywords: basketball outcome prediction, statistical models, machine learning, NBA, player performance metrics

1. Introduction

Basketball is a sport with widespread popularity and influence worldwide. Every year, people all over the world are tuning in to watch professional basketball games for the love of the game. The most globally famous professional basketball league is the National Basketball Association (NBA). It has fans, the media, and corporate partners all over the world and in different regions. Over 70% of the 2.1 billion followers on social media of the NBA are not from the United States. This proves that the NBA has grown into one of the most influential sports in the whole world over the years [1]. For so many people, watching and discussing a basketball game is not only a form of entertainment but also an everyday routine and social culture.

With the development of basketball leagues such as the NBA, the match outcome prediction has become increasingly important. Many people want to know which team will win the game before it's played. Some do it for fun, while others think it will help them pick a favourite team or guide their actions and exercise. In recent years, statistical models have been widely applied to predict the result of a basketball game. This is based on real train data: shooting efficiency, turnover rate, offensive rebounding, free throw rate, and other variables that can be measured [2]. By studying such figures, people aim to make better predictions in events such as the Olympic results. But a basketball game is tough to predict. The team data offers significant observational evidence pointing to the strength of the team and performance. There are colours, but there's a lot that can go wrong to change the ultimate game. For instance, an injury to a star player can influence a team's probability of winning. And it can impact the atmosphere, emotions, and even the referee at a football match. These things make basketball match analysis interesting and hard to predict. Statistical models are a useful heuristic, but they do not and cannot completely control or explain any actual case.

The prediction in sports has some practical significance. First, it can improve interest in watching basketball games. The audience may be immersed in the basketball games with their own feelings. When their favourite team wins the games, they can feel a sense of happiness and surprise due to the amazing prediction. Second, it can create a topic and discussion, such as which team is the champion in the NBA final games, and who is the MVP in the regular season. It can improve concentration and the hot of NBA games all around the world. Third, it is related to sports betting, and it can create huge economic profits. Some people can win some money by gambling. So, these kinds of activities can provide some interest to people who are addicted to sports betting. Fourth, it is convenient for coaches to make the plans and tactics. According to the prediction of people's performance, coaches can arrange the time in the games and promote teamwork. Therefore, the prediction of outcomes of basketball games plays an important role.

There are three main objectives in my research. Firstly, use statistical data to predict the outcome of basketball matches. Then, analyse how players' performance affects winning basketball games. Finally, explore methods to improve and evaluate the accuracy of predictive models.

The structure of this dissertation is as follows: firstly, the literature review will demonstrate the background on Basketball Prediction and important Metrics and Statistical Dimensions in Basketball Modelling. Then it introduces the main predictive modelling methods for the meantime. It tries to empirically test and evaluate the performance of some simple models based on actual NBA game data in the Development and Discussion Section. That is because this section explores the influence of various elements on how well models perform. In the final part of our paper, the main findings resulting from this research were outlined and reflect the significance of such research. Suggestions for improvement were also made to guide future researchers along. We hope this study will help us gain a deeper understanding of the game of basketball through statistical models and provide helpful information to both fans and professionals alike.

2. Research review

2.1. Background on basketball prediction

2.1.1. Importance of basketball prediction

Before we can answer the question of how accurate statistical prediction of basketball results is, we must first examine the importance of basketball predictions. Its importance can be mainly reflected in three aspects: sports betting, strategy, and player development. Firstly, some betting companies can set the odds according to the prediction of basketball games. And participants can compare these odds, which rely on their own prediction, in order to discover the under-evaluated basketball teams [3].

Firstly, some betting companies can set the odds according to the prediction of basketball games. Participants can compare these odds, which rely on their own prediction, in order to discover the under-evaluated basketball teams. They can best judge whether or not to obtain the maximum profits. Secondly, it can help participants improve their betting strategies. They can adjust the parameters of the model (such as the joining injury of players, home advantage) by predicting continuous data. In this way, they can possess more stable logic in this field rather than relying on fortune. In addition, the betting market can be regarded as a predictive tool. It can receive the updated information rapidly. This is because the market can digest any new information quickly, and the reaction rate of it is more rapid than conventional model analysis. And it can modify the error of prediction in time.

Moreover, there are some applications of betting strategies in sports betting. In the basketball field, basketball prediction surrounds data of players and the whole team. Due to the quick pace of basketball games, the data includes different kinds of sides. So, the information density is higher, and the prediction model can help the betting market to reduce the time of error in information. Finally, prediction promotes player development. It helps predict key performance indicators such as rebounds, defence, and efficiency, guiding players to focus on their weaknesses and improve training outcomes [4]. Players may show their best performances to win, and they try their best to help basketball teams win the games and show their beautiful data in games. So, teams may have competitive ability and attract more fans. The main reason to win the games is their help. Sometimes, some high-tech devices like Sport VU can provide more accurate data on players' actions with more details. It can show the real value of platers. For example, although players cannot get many scores, he prevents opponents' attacks many times and wins the games. So, these roles can also lead to victory in games. Therefore, the prediction is really important.

2.1.2. Why predicting basketball outcomes is challenging

Although it is of great value to predict the results of basketball matches, various related factors make it difficult to predict. Firstly, some uncontrollable factors increase the uncertainty of basketball prediction. For example, in the United States, teams must travel long distances to play. Many studies have shown that teams moving east tend to perform more strongly than those moving west [5]. Similarly, travel-related fatigue is associated with fluctuations in player performance, which makes it difficult for prediction models to fully capture these effects. Player injuries and player transactions further exacerbate this uncertainty, which will suddenly change the strength of the team this season.

Research on the first scoring event also shows the role of Randomness: basketball is a kind of team cooperation, so everyone on the basketball court has the opportunity to get the first score. As a result, it is difficult to predict how early events will change momentum. A study tried to predict this by simulating casinos, but even with deeper data than NBA data, the model is difficult to replicate gambling probability, once again showing how complex these uncertainties are [6]. In addition, environmental factors such as home court advantage, gender, and team cooperation ability play an important role in Spanish basketball games. A study shows that these factors vary according to gender and team ability, which makes it more difficult to establish statistical models [7].

To sum up, these research results show that the prediction of basketball game results involves many dynamic factors, such as travel fatigue, shifting player availability, and early-game randomness. These variables affect the accuracy of prediction, so it is very challenging to be completely accurate.

2.2. Important metrics and statistical dimensions in basketball modelling

Since the accuracy of any statistical prediction depends on the selection of the most relevant inputs, this section explores key player-specific and team-specific metrics used in basketball modelling and explains why these variables are critical to assessing the accuracy of our prediction of game results.

2.2.1. Player-specific metrics

Basketball players are really important for winning games. Their contributions to teams are more important than their own personal performances. Therefore, understanding the player-specific metrics helps to clarify how much statistical models can capture the individual impact that influences game outcomes.

Research has shown that analysing players' statistical data in combination with team winning rates provides better insights into their overall impact. The NBA season is divided into regular seasons and postseasons. Although the regular season makes up most of the games, success is not measured in professional sports solely by an athlete's regular-season performance. The true measure of greatness is how a player performs in the postseason [8], as the playoffs are characterised by stronger opponents, higher tactical intensity, greater pressure, and more decisive personal contributions. So, the contribution from scores, catching rebounds, and giving assists of individual players is a more evident function in the postseasons. This idea is consistent with Oliver's argument [2], which states that the basic player metrics, such as Points Per Game (PPG) and efficiency ratings, are the basis of basketball analytics and provide higher modelling starting points [2].

Taken together, Berri and Oliver both emphasised that understanding player-specific metrics is crucial for building accurate prediction models, because individual performance (especially in high-risk environments) determines the success of the team.

2.2.2. Team-specific and situational metrics

Building on the focus on individual metrics, team-specific and situational metrics also capture the performance of the team as the coordination unit, which is crucial to evaluate the degree to which the statistical model can accurately predict the game outcomes.

Team-specific and situational metrics can capture general performance from a team, including its tactics, season goals, and strategies during the match. These metrics are really important to prediction, as statistical models not only need to consider individual abilities, but also how teams organise offence, defence, and decision-making under different conditions. All these factors can directly affect the possibility of winning.

For example, the strategies of a team in different game situations can affect the efficiency of players' coordination and execution of the game. An excellent strategy can help players communicate with each other well and have wonderful performances in games. In addition, situational metrics like the performance in critical cases are considered. It can practice the ability to execute as a team and the individual performances of basketball celebrities.

In the FIBA World Cup, the study examines the relationship between match Key Performance Indicators (KPIs) and outcome in elite men's basketball, and identifies the most suitable analysis to model this relationship during the men's basketball tournament [9]. These KPIs include team shooting efficiency, total rebounds, assists, turnovers, and defensive actions, all of which describe the contribution of team performance. According to these figures, they can adjust the starting lineup and let everyone achieve the biggest functions in the games.

In addition, the Elo rating system is always used in NBA playoffs to track a team's relative strength throughout the entire season. Elo is a dynamic indicator of team strength that increases when the team wins and decreases when the team fails, resulting in greater gains from failed victories or huge winning profits [10]. It can calculate the real-time combat power scores. and make the curves about the decrease in physical fitness of core players. So, coaches can arrange the time that they play and prepare the tactics to face these problems. Importantly, Elo can use a sigmoid function, namely the logistic, based on rating differences to convert it into a winning probability estimate. This makes the Elo really related to prediction, as it provides a simple, continuously updated estimate of the possibility of one team defeating another.

Overall, the evidence suggests that both individual and team metrics affect the outcomes of the basketball game. The accuracy of the statistical model in prediction will be improved when considering both these categories of factor.

2.2.3. Limitations in current data usage

Although player and team metrics constitute the basis and core input of most prediction models, there are still important gaps in how to use data in current research.

Current research usually uses recordable and numerical data, but there are some events, such as player trades and injuries are characterised by unpredictability and immediate impact. For instance, a sudden injury to a core player may lead to a sharp short-term decline in a team's strength, while a blockbuster trade could quickly boost a team's competitiveness. Such events are difficult to predict in advance using traditional "historical data trend-based" models. Furthermore, their impacts tend to be phased—an injury may last for several weeks, and it may take 1-2 months for a team to integrate a newly traded lineup. The injuries are serious in any sports games. Therefore, it is necessary to clarify how injuries affect NBA teams, healthcare professionals, medical staff, and coaches, and it is an issue that deserves attention [11].

These limitations show that even a well-designed statistical model can only be accurate when its basic variables reflect the real situation of basketball. This therefore suggests that there may be a limit to how far they can answer the research question.

2.3. Types of statistical models used in basketball prediction

The choice of model determines how well these metrics can be used to predict results, so the next section reviews five main statistical models commonly applied in basketball outcome prediction—linear models, logistic regression, metric-based analysis, machine-learning methods, and state-transition models.

2.3.1. Linear models

Linear models assume a straight-line relationship between variables, making them one of the simplest ways to understand how different factors are related to basketball performance. They feature simple calculations and interpretable results, so researchers often use them first to identify which performance indicators may be important.

Linear models can improve match prediction because they can explain results. Leicht et al. studied Olympic basketball games and found that shooting efficiency and several other key statistics had a high correlation with success [12]. This demonstrates that the linear relationship between team statistics and match outcome can serve as an early indicator for more accurate prediction. A case study at California State University, Northridge, assessed the performance of simple and multiple linear regressions in predicting the number of regular-season wins. The simple model was capable of explaining only a portion of the variation in results, and the multiple model, which included point differential, shooting percentages, and turnovers, proved to be more accurate [13]. This comparison illustrates that multiple linear models are more effective when several important variables are entered into the equation.

Overall, Linear models provide a clear and interpretable way to see which factors are related to winning. They are useful as a starting point, but because they only capture linear relationships, so can't fully reflect the more complex patterns common in real basketball games.

2.3.2. Logistic regression models

Logistic regression is a type of generalised linear model, which is used in a division problem to predict discrete classifications, rather than predict continuous scores. This makes it very important to evaluate the possibility of a team winning a basketball game.

Logistic regression includes binary logistic regression, multiple logistic regression, and sequential logistic regression. Among them, the binary logistic regression model is the most commonly used in sports outcome prediction. Research evidence supports its usefulness in solving my problem about the accuracy of statistical prediction. For example, Kvam and Sokol used a combined logistic regression Markov chain model to predict the results of NCAA Championships, and achieved good results using only basic team statistics [14]. This shows that logistic regression can successfully describe the relationship between game indicators and game results. Other studies provide useful comparisons. Dobson pointed out that although Elo rating is simple and effective, running logistic regression on the same data set can match or better than Elo [15]. This comparison shows that logistic regression provides a more statistically based estimation method, even though both methods rely on a relatively simple relationship.

In general, logistic regression provides a clear explanation and strong baseline accuracy, although its linear assumptions limit its flexibility in a more complex basketball environment.

2.3.3. Metric-driven analysis

Unlike the early linear and logical models, metric-driven analysis starts from the metrics themselves, not from the assumed functional form. Metric-driven analysis solves a key limitation of linear models and logistic regression models: the outcome of a basketball game is affected by many nonlinear factors. The way does not assume a simple linear relationship, but is based on the core indicators of the competition, such as Player-Specific Metrics, Team-Specific, and Situational Metrics. It can more accurately capture the nonlinear relationship and multi-dimensional influencing factors of basketball.

Kubatko et al. have provided a starting point for analysing basketball statistics [16]. They pointed the possession concept is central to basketball analysis, and sorted out useful tools such as attack and defence score and real shooting percentage to help unify the basis of research. The work helps to improve the accuracy of predicting basketball games. With the growth of technology, analytics has started to change the game of basketball.

Subsequent research, like Kubatko, also emphasised the importance of core competition data, but began to use more advanced technology to obtain more detailed competition information. Terner and Franks reviewed modern methods for quantifying and characterising basketball gameplay and showed how measurement-based modelling can more accurately evaluate team strategies, game modes, and individual contributions than traditional models [17].

Together, these studies show that metric-driven analysis seems to provide a more flexible and realistic basis for basketball prediction than linear-based methods. On the whole, it increases prediction accuracy through capturing the non-linear and context-sensitive aspects of the game.

2.3.4. Machine learning models

The above statistical models and metric-driven models have some limitations; machine learning provides a more flexible method to capture complex and non-linear relationships in the basketball data. Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed [18]. It has been used in data analysis since the 1950s. Later work shows these models can process large and complex data more effectively than traditional statistical tools, becoming more and more popular in the 20th century, which is essential for the analysis of basketball performance.

The research on applying machine learning to basketball shows this advantage. For example, Kohli used 26 predictive variables and a binary response variable representing playoff qualification to characterise the team, and promoted the prediction accuracy through classification tree pruning and random forest modelling [19]. The results show that the defensive factors, such as the team's playoff qualification, the opponent's average assists, and the opponent's average two-point shots, play a key role in predicting playoff eligibility. This supports the view that machine learning can reveal influential features that simple linear models may ignore. Current studies have expanded this model by using support vector machines, random forests, and neural networks to combine historical data, in-game statistics, and real-time data. These models optimised betting strategies and identified value bets, ultimately improving profitability [20].

In a word, machine learning provides a flexible and powerful approach for basketball prediction, especially when the data is high-dimensional, and the relationship between the data is nonlinear.

2.3.5. Distribution-based and state-transition models

Finally, models based on distribution and state transition provide an effective method to describe the characteristics of probability and dynamics in basketball games. Among them, the Poisson model can simulate the probability of discrete events, which is in line with the gradual control structure of basketball. The method is widely used in basketball prediction, such as the NBA winning probability model of FiveThirtyEight, which uses the Poisson model to generate two teams' remaining scoring probability before the last 90 seconds of the game, and then transforms it into the probability of winning. This shows how the distribution-based model can provide timely and data-driven predictions, which are closely combined with the development in the game. Other evidence highlights different advantages.

The Exponential Model uses the exponential random variable to model the waiting time between successive wins of seeds in a round [21]. In contrast, a Markov chain can draw the transfer process between competition states. For example, some scholars have modelled the attack round in basketball games as a team-specific non-stationary Markov Decision Process (MDP), and the state transition probability depends on the 24-second attack clock [22].

In these studies, the methods based on distribution and state transition have different concerns. The Poisson model emphasises event frequency, the Exponential model emphasises waiting time, and the Markov chain emphasises game flow, but they are all helpful tools for modelling basketball probability structure. In general, these models provide valuable and professional insights to help capture different aspects of uncertainty in basketball prediction.

2.4. Limitations of existing research and future directions

Although many studies have provided useful insights, there are still some gaps in the current research, which limit how accurately basketball outcomes can be predicted

2.4.1. Structural gaps in current studies

First, most of basketball prediction research focuses on quantitative statistical analysis, such as scoring, rebounds, and assists, lacking attention to qualitative factors and real-time competition information. For example, factors such as on-site tactical adjustments, game pace are often not taken into account in the model. Moreover, there is too little integration between different model methods. Linear regression, logistic regression, machine learning, or state transition models are independent and have not yet formed a unified framework. Therefore, future research needs to make up for these blind spots by adding qualitative and real-time factors, combining the advantages of different models.

2.4.2. Optimisation directions for statistical methods

In order to improve the accuracy of basketball game prediction, future research can explore combining continuous and discrete models, applying Markov chains to real-time data analysis at the same time. Basketball games are a dynamic process, so we may consider relating players' state transitions to their overall distribution during the game. Also, the outcomes of the prediction need to be interpretable and easy to understand for teams and coaches. These are some optimisation directions for statistical methods on basketball game prediction.

After discussing the limitations of existing research and future directions, the research review ends here. The following discussion will study how these insights can be applied to real basketball data and practical modelling methods.

3. Discussion/ development

3.1. Overview

My project is used to predict the outcomes of basketball games accurately. Basketball games are highly uncertain and influenced by various objective and subjective factors, such as the injury of core players and home advantage. Therefore, basketball prediction is both challenging and practically valuable.

There are many statistical models in basketball prediction. Linear regression is a statistical method used to model the linear relationship between independent variables and a dependent variable. The purpose of it is to find the best-fit line in order to describe the relationship between variables. In addition, I have also analysed logistic regression, machine learning, state transition models, and Markov decision processes. These models can help us record all the information about the data of basketball games or some excellent figures of individual players. Some line charts and bar graphs can demonstrate the performances of basketball teams that are compared directly with other teams. And it can show the core competitive ability of a team. Therefore, these models can help us predict the winner of basketball games better.

3.2. Evaluation of modelling approaches in literature

To judge how far statistical models can accurately predict basketball outcomes, it helps to begin with the traditional methods outlined in the literature review. These include simple linear regression, multiple linear regression and logistic regression, which together constitute the basic kit of early basketball prediction research.

3.2.1. Traditional statistical models

In these traditional models, linear regression is usually the first method used because it checks how one variable changes in relation to another, which makes it a straightforward way to test whether key performance statistics can help explain win patterns.

The linear regression model is one of the important models to predict the outcomes of basketball games. The simple regression is used to build a relational model between a response variable and a predictor variable.

For instance, one study predicted NBA regular-season wins based on shooting percentage (FG%) at first, and found that scoring efficiency has some explanatory power for team wins(R^2= 0.44). Then, a multiple linear regression model including offensive and defensive indicators was established, such as two-point field goal percentage, three-point field goal percentage, and so on (see Figure 1). The R^2 of the multiple model has been increased to 0.905, highly improving model accuracy [13].

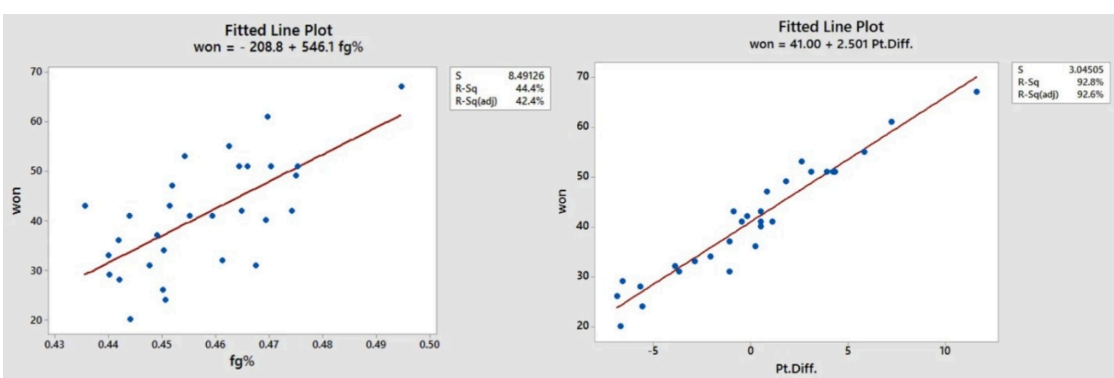

Figure 1. Model diagrams and results of simple linear regression and multiple linear regression [13]

Next, taking the Olympic Games as an example, Leicht et al. analysed the relationship between the team performance metrics and the game outcomes, emphasising the explanatory effect of linear and nonlinear analysis on the victory rate [12]. In Olympic games, coaches and players do not have many chances to interact. So, the index of the team's performances can provide important instructions to coaches. This can help teams improve the winning rate in some top competitions. Linear and nonlinear analyses examined the relationship between match outcome and team performance indicator characteristics, namely, binary logistic regression and a Conditional Interference (CI) classification tree [12].

Therefore, linear regression can provide direct and interpretable information about how individual and team statistics are related to basketball game outcomes. From the improvement of simple regression to multiple regression, adding a wider range of variables can enhance the model's ability to reflect real basketball performance. It should be noted, however, that high R^2 values mainly indicate a good fit to existing data and do not always guarantee strong predictive performance across different seasons or contexts. Despite this limitation, linear regression remains a useful and transparent starting point, as its main value lies in interpretability and baseline comparison rather than maximum prediction accuracy. It can therefore be concluded that linear regression provides a useful and transparent starting point for judging how far statistical models can predict results, even if more advanced methods are needed to improve the accuracy in the future.

3.2.2. Machine learning and advanced methods

Traditional models offer some clear and helpful insights, but their limitations have created a need for more sophisticated prediction techniques. Strong data-driven capabilities of machine learning and advanced techniques are changing how decisions are made in a variety of fields, including basketball.

Based on this transformation, recent research has increasingly applied machine learning techniques to predict the outcome of basketball games, demonstrating their potential to capture complex patterns. For instance, a Decision Tree Model was used to forecast the results of NCAA men's basketball tournament games. Its results are highly visual and easy to understand—even non-technical personnel can follow the judgment logic. As shown in Figure 2, some statistical models have achieved excellent performance across three core evaluation metrics: accuracy, F1 score, and Ranked Probability Score (RPS). Specifically, Jain et al. reported that their Hierarchical Fuzzy Support Vector Machine (HFSVM) model achieved 89.3% performance, which is much higher than the traditional logistic regression model (about 74%). Other models, such as random forests and Graph Convolutional Network (GCN) ensemble models, have also shown improvements, typically achieving results of 70% to 83% in the above metrics. RPS is an indicator where a lower score indicates a more calibrated and accurate probability ranking. In terms of reliability, Chen et al. achieved a powerful RPS of 8.180, indicating that its calibrated predictions outperform comparable models [23].

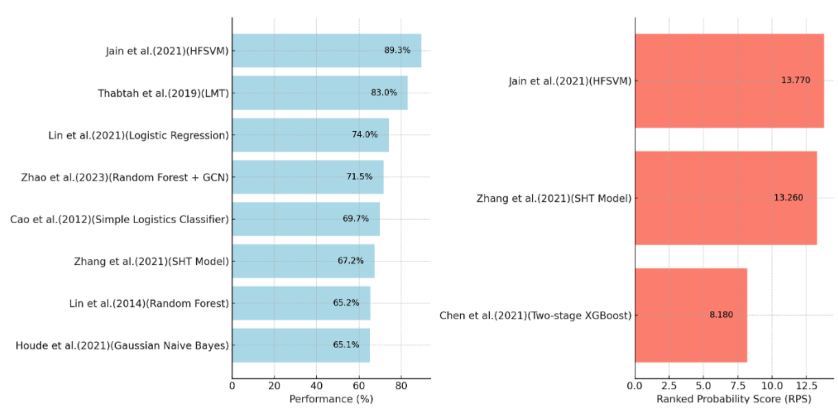

Figure 2. The best performances in basketball analytics based on accuracy, F1, and RPS [20]

The advantages of machine learning make it easier for us to predict the outcome of basketball games. To begin with, it is able to quickly and efficiently process high-dimensional, large-volume data. As a result, it can help us conduct real-time processing for situations such as real-time recommendation systems. Accordingly, such a function is suitable for basketball prediction, where large volumes of possession and tracking data must be processed on time. Second, machine learning models can be continuously improved. When they receive additional basketball game data, they learn from these games and make fewer mistakes. Consequently, the predictions improve over time; since the game of basketball evolves quickly, this can be advantageous during games. Third, machine learning can automate repetitive or complicated tasks, thus liberating workers from such activities. These technologies automatically process large volumes of data without intervention and separate them into meaningful categories. Basketball prediction can benefit from this, since it can quickly and consistently analyse and organise the shooting, defensive action, or contextual data by enabling the model to do so.

However, there are also disadvantages to machine learning. There is a strong reliance on high-quality data with machine learning. If the data is poorly generated, the model may not be effective. Building the models and maintaining them are also expensive. In the case of basketball prediction, the predictions will not be effective if the tracking or scoring data is not reliable. Most basketball teams or organisations do not have the budget or the technical personnel to build and maintain these models. In addition, some critics argue that machine learning models seem to outperform traditional methods mainly due to overfitting, data leakage, or accessing richer datasets that are not widely available, making it difficult for their reported accuracy to be transferred to real-world environments. This concern is reasonable. However, these issues have not rendered machine learning ineffective as a prediction method; On the contrary, they emphasise the need for appropriate model validation, regularisation, and context-specific applications.

On the whole, the strength of machine learning is its scalability and predictive power, yet its weakness is interpretability and accessibility. They are most effective when a large amount of reliable data and computing resources are available. For example, while complex models like deep neural networks can achieve higher accuracy, it is difficult for coaches and teams to understand the reasons behind a particular forecast. In contrast, simpler traditional model, though less accurate, is more interpretable, also is more adapted to decision support. So, the application of machine learning in basketball games depends on the balance of accuracy, interpretability, and resource availability, but not only on pursuing performance.

3.2.3. Markov models

Based on advanced statistical and machine learning methods, another method to predict the results of basketball games is the Markov model. The model is only based on the current state and focuses on the probability of moving from one state to another.

Markov models are the probability of a system's future state, which depends solely on its current state, independent of past states. The core features are no memory, discrete States, and stable transition probabilities. It applies to scenarios where both states and time are discrete, such as stock market ups and downs prediction and user behaviour path analysis. For example, play-by-play data from the 2007–08 and 2008–09 NBA seasons were used to apply a Markov chain Monte Carlo model to estimate the model's stationary distribution [24]. From this distribution, they get the number of points per transition, which, multiplied by the number of transitions, gives an estimate of the number of points scored. From Table 1, we can see details about the play-by-play data and how they were transformed into states.

Table 1. A table of an excerpt from the first quarter of a match between Seattle (A) and Phoenix (B) [24]

	Play-by-play data				Parsed model states			
	Team	Player	Event	Result	Possession	Method	Pts.on previous	State
1	PHX	Amare Stoudemire	Shot	2pt	SEA	Inbound pass	2	Ai2
2	SEA	Damien Wilkins	Shot	Missed				
3	PHX	Grant Hill	Def. rebound		PHX	Def. rebound	0	BdO
4	PHX	Steve Nash	Bad pass		SEA	Steal	0	AsO
5	SEA	Kevin Durant	Shot	2pt	PHX	Inbound pass	2	Bi2
6	PHX	Shawn Marion	Shot-3pt	Missed				
7	SEA	Earl Watson	Def. rebound		SEA	Def. rebound	0	AdO
8	SEA	Damien Wilkins	Shot	2pt	PHX	Inbound pass		

The core function of a parsed model is to convert unstructured/complex data into a structured and interpretable form, supporting subsequent analysis, decision-making, or system operation. So, we can easily understand the performances of individual players. In addition, Bayesian hierarchical models are employed in the modelling and parameterisation of the transition probabilities to borrow strength across players and through time [22].

3.2.4. Comparative insights from previous studies

Table 2. Summary of basketball prediction models [20]

Approaches	Work	Performance	Metrics	Features	Datasets
XGBoost	Chen et al. (2021)	91.82%	Accuracy	Defensive rebounds, 2 PFG%, FT%, Offensive rebounds, Assist, 3 PFG attempts	NBA 2018-2019 season (2460 game data points)
Logistic Regression	Cao (2012), Lin et al. (2014), Horvat et al. (2020), Houde (2021), Sukumaran et al. (2022)	69.67%, 68.75%, 60.82%, 65.1%, 93.20%	Accuracy	Comprehensive NBA statistics data mart, Points scored, FG attempts, Defensive rebounds, Assist, Turnovers, Record, Team assists, Steals, Personal fouls, FG attempts, Rebounds, Win percentage, FG%, 3PF%, FT%, Rebounds, Assist, Turnovers, Steals, Blocks, Plus-minus, Offensive rating, Defensive rating, True shooting percentage	NBA data from 5 regular seasons for training, 1 for scoring, NBA games from 1991-1998, NBA Enhanced Box Score and Standings (2012-2018) from Kaggle, Games from three NBA seasons (2018-2021)

Table 2. Continued

Random Forest	Lin et al. (2014), Zhang et al. (2021), Zhao et al. (2023), Cai et al. (2019), Houde (2021)	65.15%, 64.63%, 71.54%, 84%, 65.1%	Accuracy	Points scored, FG attempts, Defensive rebounds, Assist, Turnovers, Record, Team strength, Home court advantage, Player tiredness, Team assists, Steals, Personal fouls, FG attempts, Rebounds, Two-point shots, Three-point shots, Free throws, Attack, Defense, Assist	NBA games from 1991-1998, NBA official website (2016-2021), Professional Transactions Archive, Sportac/ESPN, NBA Enhanced Box Score and Standings (2012-2018) from Kaggle, NBA 2018-2019 season (2460 game data points)
SVM	Cao (2012), Jain and Kaur (2017), Li et al. (2021)	69.67%, 89.26%, 73.95%	Accuracy	Comprehensive NBA statistics data mart, 33 condition attributes including biometric data, college stats, draft order, Two-point shots, Three-point shots, Free throws, Defensive rebounds, Assist, Steals, Turnovers, Personal fouls	NBA data from 5 regular seasons for training, 1 for scoring, NBA data from 800 games in 2015-2016 regular season, NBA data from 2011-2012 to 2014-2015 seasons
Neural Networks	Oksen and Onay (2022)	74.33%, 76%	Accuracy	Field goals, 3-pointers, Free throws, Rebounds, Assist, Steals, Blocks, Turnovers, Personal fouls, Player efficiency rating, True shooting percentage, Effective field goal percentage	NBA 2007-2008 season data from ESPN, NBA regular season data from www.basketballreference.com and shot selection data from www.nbanner.com
Naive Bayes	Lin et al. (2014), Miljkovic et al. (2010), Jain and Kaur (2017)	65.15%, 67%, 89.26%	Accuracy	Points scored, FG attempts, Defensive rebounds, Assist, Turnovers, Record, Field goals, Three-pointers, Free throws, Rebounds, Assist, Turnovers, Steals, Blocks, Fouls, Points per game, standings attributes like total wins/losses, home/away records, current streaks, 33 condition attributes including biometric data, college stats, draft order	NBA games from 1991-1998, NBA data from the 2009-2010 season, NBA data from 800 games in 2015-2016 regular season

Table 2. Continued

Ensemble Methods	Cai et al. (2019), Sukumaran et al. (2022)	84%, 84%	Accuracy	Two-point shots, Three-point shots, Free throws, Attack, Defense, Assist, Multiple features evaluated across various studies	CBA dataset (2016-2017 season), Various datasets evaluated
Poisson Factorization	Ruiz and Perez-Cruz (2015)	-	Negative log-likelihood, Profitability	Team-specific and conference-specific attack and defense coefficients	Regular season results from over 5000 games to predict outcomes of the 2014 NCAA tournament games, data sourced from Kaggle and OddsPortal
Heteroscedastic Models	Manner (2016)	-	Win probabilities	Team strengths, Home court advantage, Back-to-back game effects	Eight NBA seasons' data from 2006-2014
Markov Models	Strumbelj and Vračar (2012)	-	Probabilistic forecasts	Possession-based game progression	Play-by-play data from the NBA 2007-08 and 2008-09 seasons

Through the previous introduction, we can know the features of different models (see Table 2). The function of Linear Regression is the prediction of single indicators, such as player points or rebounds, analysis of linear influencing factors on team winning rate. The role is to clarify linear correlations between variables with extremely high interpretability, enabling quick identification of key influencing factors. Moreover, the function of Machine Learning, like Decision Trees, is the prediction of game outcomes, the comprehensive rating of player performance, evaluation of complex tactical effects. The character is to process high-dimensional non-linear data, integrate multi-dimensional information such as player data, tactical data, and opponent data, with higher prediction accuracy than traditional models. The function of the Markov Model is to predict player state transitions and round-by-round offensive and defensive outcomes. In terms of accuracy, logistic regression performed best, reaching 93.20% accuracy in the three NBA seasons (2018-2021), followed by XGBoost, which reached 91.82% accuracy in the 2018-2019 NBA season [23]. Overall, the choice of model depends on the objective: using linear regression for interpretable insights, using machine learning for high prediction accuracy, and using Markov models for short-term state analysis.

3.3. Determinants and limitations in prediction

3.3.1. Importance of data quality

Data quality is important in prediction modelling. Accurate and complete data is the base of a proper model. Wrong or incomplete data will make the prediction result deviate more. We usually use player and team metrics in predicting the outcome of basketball games. The accuracy of the data will affect the quality of the result. Player-related data (such as points, rebounds and assists, even the contribution of the core players to the team) are the basic data to do basketball prediction. If that data is wrong or out of date, we can not estimate the influence of players in the team. For example, comparing the contribution of two core players (Jordan and Malone) in a basketball game, although the difference in per-minute productivity between Jordan and Malone is small (0.00442 versus 0.00437), Jordan is the reason for 0.38 additional postseason victories, because only there is an additional 77 min Jordan played [8]. The above case illustrates that even a very small difference in the data can affect a big difference in the prediction result if the data is right. On the other hand, hidden metrics, such as players' injury status and recent players' fatigue, can be easily ignored by data modelling, but they may be the most important variables involved in the result of the game.

Similarly, Team metrics also depend on data quality. A team's offensive and defensive efficiency, tactics execution rate and home win rate can represent its overall level of coordination. If these metrics are not updated in real time and have errors, the prediction accuracy will decrease. As for metrics, such as the balance between offence and defence and clutch shot handling ability, they are more efficient in improving the credibility of the prediction than the strength in a single dimension, and the impact of metrics on prediction will be further enlarged by factors that are hard to quantify, such as team chemistry and tactical adjustments in games.

The Elo rating system reinforces that it is consistent and accurate data that are essential for reliable prediction. The Elo rating system is based on statistical theory and adjusts the rating of players by comparing the actual results of matches with expected results. It assumes that the differences in players' ratings can determine the outcome of matches, that higher-rated players will have a higher chance of winning. When a lower-rated player defeats a higher-rated one, the lower-rated player will receive a large number of points and the higher-rated player will receive a large loss. If the higher-rated player wins, the rating changes will be small for both sides. The long-term average Elo rating is 1,500. However, it may differ a little in any given year depending on the timeliness of the league's expansion (see more on that below). The ratings of more than 90 per cent of teams are between 1,300 (pretty awful) and 1,700 (really good), but historically great or terrible teams may lie outside this range. We can see that if we don't have accurate and timely data updates, level scores and predictions on the basis of level scores will become ineffective (see Table 3).

Table 3. NBA team elo rating and corresponding record description [10]

ELO	EQUIVALENT RECORD	TEAM DESCRIPTION
1,800	67-15	All-time great
1,700	60-22	Title contender
1,600	51-31	Playoff bound
1,500	41-41	Average
1,400	31-51	In the lottery
1,300	22-60	LOL
1,200	15-67	Historically awful

3.3.2. Contextual and qualitative factors

Not only data quality, but also some contextual and qualitative aspects are relevant for prediction results. Injury is always an unavoidable part. Players can collide and get serious injuries. The events of injuries have a measurable impact on player and team performance. Thus, predictive models that do not consider injuries are likely to under-estimate the volatility [25]. Also, some injuries affect salaries. For instance, chest injuries produce the largest average salary reduction (−25.6%). Surprisingly, the average salary change is positive (+13.9% and +12.6%) for injuries in the pelvic and upper arm and forearm areas, respectively. This could be explained by contracts, insurance and other external factors that are not reported in the table [11]. The Tornado diagram shown in Figure 3 is a visual representation of the variance in basketball performance metrics following an injury. Each row is a different performance metric. The percentages on the left are "Lesser Post-Injury" performances, whereas the percentages on the right are "Greater Post-Injury" in Figure 3. This diagram suggests a variance in basketball performance metrics following an injury. Certain metrics, such as possession, defensive and offensive rating and usage percentage, exhibited the greatest variance, indicating that players performed differently following an injury.

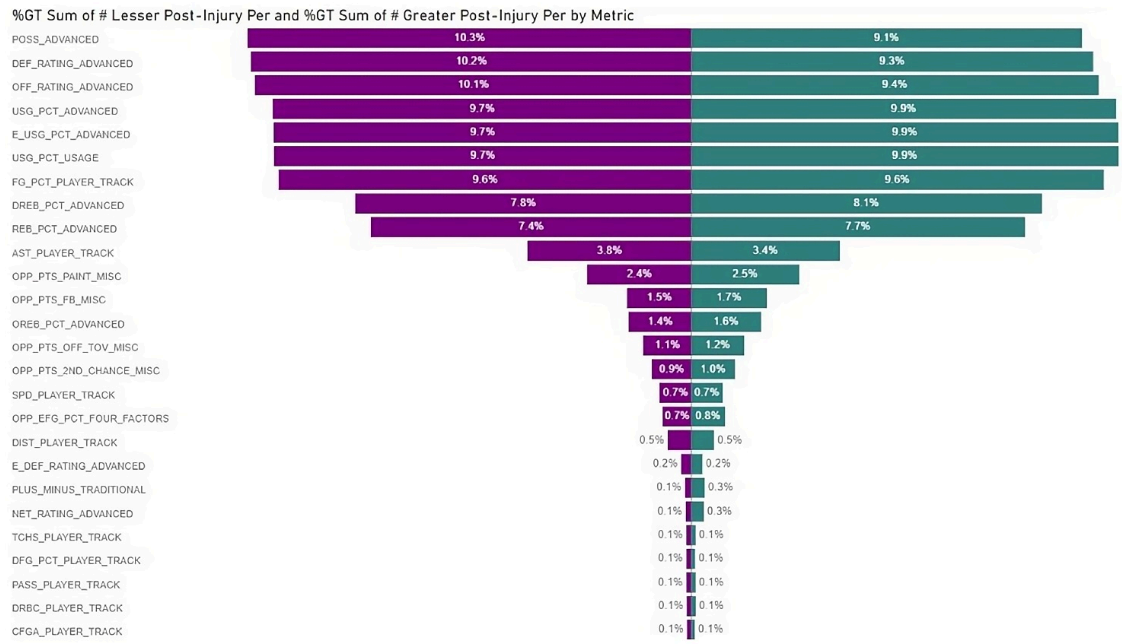

Figure 3. Tornado diagram that analyses the percentage variance in basketball performance analytics in lesser/greater post-injury cases [11]

In contrast, home advantage is also a crucial external factor in basketball prediction. Players can be influenced by the actions of audiences, even by noise in the environment. Offensive variables and defensive variables were analysed between eight seasons across 4337 players. Home and away matches differed significantly in goals ($p < 0.001$), assists ($p < 0.001$) and season key passes ($p < 0.001$), all of which were higher at home. In this sense, Home advantage contributes to the winning rate of some teams. These results suggest that, while contextual and qualitative factors are difficult to quantify, they are essential elements of data quality.

Meanwhile, background and qualitative factors such as injuries and home advantage are often difficult to measure with specific numerical values accurately. If the calculation of these factors is not accurate enough, it will bring additional errors to the predicted results of the competition. However, if these factors are completely ignored, the predicted results will exhibit systematic bias instead. Especially in games with tight scores and intense competition, injuries or home advantage can often directly affect the outcome of the game. In fact, even with some simple substitute indicators or adjustments based on the specific situation of the competition, the prediction model can be more in line with the real competition situation. This also indicates that the purpose of predicting basketball games is to improve the accuracy of predictions in situations full of uncertainty, rather than completely eliminating all uncertain factors.

3.4. Practical implications and future directions

3.4.1. Applications in team performance optimisation

In recent years, many statistical models have been used to predict the outcome of basketball games. Methods for quantifying and characterising basketball gameplay were introduced by a review [17]. This growth has coincided with a rise in data availability and innovative methodology that has inspired fans to study basketball through a statistical lens. The weighted likelihood method is also applied to analyse the team performance of the Chicago Bulls and Utah Jazz in the 1996-97 NBA Finals, and use the game data of the 1996-1997 season to determine the key factors affecting the playoff results [26].

Furthermore, Real-time collects players' data such as position, running distance, sprint speed, heart rate, and muscle load through high-definition cameras installed on the court and wearables. For example, NBA teams use systems like Second Spectrum and Sport VU, which track players' and the ball's real-time positions over 20 frames per second via overhead cameras to produce extensive trajectory data. This data helps evaluate player efficiency and calculate advanced metrics such as Real Plus-Minus and defensive coverage area. Additionally, utilise tactical analysis systems to generate comprehensive reports shortly after matches. These systems automatically record players' running distances, pass accuracy, and defensive coverage, and identify opponents' weaknesses, providing a basis for coaches to adjust tactics [27]. Meanwhile, coaches employ analytics to significantly enhance athlete performance by offering strategic guidance, personalised training plans, and ongoing feedback to help athletes reach their full potential and develop skills. Therefore, in the context of sports competition, there are various applications aimed at optimising team performance.

3.4.2. Applications in sports betting and market strategy

In NBA prediction, the target is betting on the winner of each game. The betting has only two results, directly corresponding to the most basic coin-tossing setup [28]. Many prediction models are used in sports betting to help fans or businessmen reap benefits. As shown in Figure 4, from 2012 to 2023, articles on basketball betting prediction models were published each year. Some excellent basketball prediction models were summarised. For example, the focus was on predicting the outcomes of NBA games using algorithms such as the Simple Logistics Classifier, ANN, SVM, and Naïve Bayes, with the Simple Logistics Classifier achieving the highest accuracy of 69.67% [29]. Later, his models were used for field goal shooting percentage, three-point shooting percentage, free-throw shooting percentage, offensive rebounds, assists, turnovers, and attempted free throws, achieving an accuracy of 88-94% [30]. In addition, a probabilistic model using modified Poisson factorisation was applied to predict NCAA tournament game outcomes, demonstrating superior profitability and prediction accuracy [31].

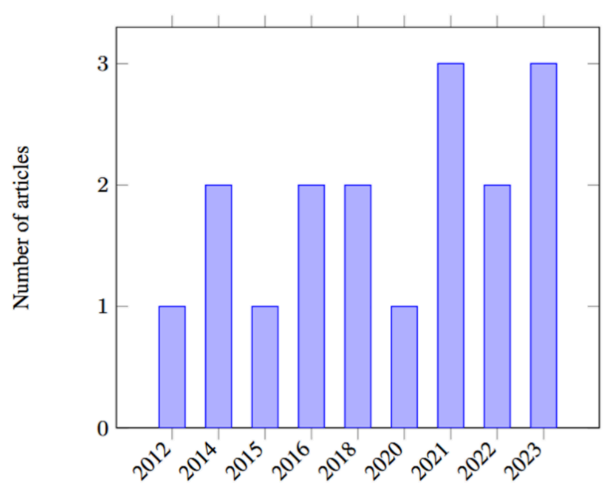

Figure 4. Histogram showing the number of articles per year in basketball betting [20]

Machine Learning (ML) has played a key role in the transformation of basketball betting, and it can provide more accurate predictions, dynamic odds-setting, and enhanced risk management for both bookmakers and bettors [20]. It can be seen that predicting basketball game results has a wide range of applications in sports betting and market strategy.

3.4.3. Combining models and new approaches

In the future, how can we use statistical models to predict basketball game results more accurately? I suggest that the model can be approached from the following 3 perspectives: model combination, data innovation, and method improvement. A hybrid basketball game outcomes prediction scheme was developed for predicting the final score of the National Basketball Association (NBA) games by integrating five data mining techniques, including extreme learning machine, multivariate adaptive regression splines, *k*-nearest neighbours, Extreme Gradient Boosting (XGBoost), and stochastic gradient boosting. And the scheme achieved high prediction [23]. What's more, real-time data was used to predict the potential score of each ball possession, so that more accurately analyse players' performance [32]. As modes become increasingly complex, interpretability becomes more important and necessary. Coaches, management, and even players need easy-to-read prediction outcomes and analysis. So, it is requisite to consider models of interpretability. In summary, we can improve our model accuracy by combining models, data innovation, and new approaches.

4. Conclusion

In answering the question of how to improve the outcome of basketball prediction accurately, this dissertation has explored different kinds of ideas and theories from many mathematical models and the data of basketball games, and found that both fields lead to the same direction: basketball prediction will be improved with some statistical models, each offering its own strengths. Linear regression models are useful because they make it easy to see which variables are the strongest predictors of winning, and for be-ginners are easy to understand. Logistic regression models transform team statistics into probabilities of winning, and provide an effective and straightforward baseline for a classification problem. Machine learning models, such as random forests or support vector machines, can be used on larger and more complex data sets, capture non-linear relations, and typically provide higher accuracy when the quality of data is high. Markov and state transition models have a special strength in modeling the game flow possession by possession and are very useful for deepening understanding of short-term dynamics and tactical patterns.

It should be mentioned that basketball game also is dynamic to some un-predictable factors and different modelling methods have differences in computational cost and burden, data requirement, and interpretability. Also the prediction accuracy depends largely on the quality of data and it is hard for real data to get complete data. However, these limitations do not reduce the application value of statistical models. The aim of the model is not to get 100% accurate prediction but to provide a more reliable and scientific judgment basis than intuition under uncertainty and reveal potential structure that is hard for the eye to see. Hence, it is essential to update data regularly, choose methods in a required way, and combine multiple technological routes to improve prediction performance.

After that, the dissertation has examined the wider implications of basketball prediction in different fields. In team management, basketball prediction can provide data on the number of assists for individual players and help coaches coordinate the lineup. In this way, it can help teams get more scores easily and make the competition run more smoothly. This can help teams get better grades in seasons. In sports betting, basketball prediction can provide cutting-edge data on basketball games. Companies and fans can rely on these data when making decisions, which increases both interest and media attention. These examples show that basketball prediction has practical, strategic, and commercial value.

Lastly, the dissertation has made some suggestions for further research. To improve prediction accuracy further, we should show the whole figures of basketball players and teams, combining this with certain models in order to improve accuracy. With regard to model design and model combination, real time data, interpretability and contextual factors such as data quality, player injury and home perspective should be considered. In general, statistical models are useful for understanding and predicting basketball games, but their usefulness is limited by data quality and the nature of the sport, which is inherently unpredictable.

References

[1] Kasimov, E. (2025, November 13). *NBA global takeover: Does America's fade even matter? SportsEpreneur.* Retrieved November 13, 2025, from https: //sportsepreneur.com/nba-global-takeover

[2] Oliver, D. (2004). *Basketball on paper.* University of Nebraska Press.

[3] Di, W., Yuan, B., & Wen, H. (2021). How to win the house: NBA odds analysis and predictions. *Open Journal of Social Sciences, 9*(2), 177–187. https: //doi.org/10.4236/jss.2021.92013

[4] Roshan Chandru, Kaushik, A., & Jaiswal, P. (2025). Enhancing basketball team strategies through predictive analytics of player performance. *Electronics, 14*(11), 2177. https: //doi.org/10.3390/electronics14112177

[5] Cuesta, I. M., Jiménez-Sáiz, S. L., Kelly, A. L., & Bustamante-Sánchez, Á. (2024). The influence of home-court advantage in elite basketball: A systematic review. *Journal of Functional Morphology and Kinesiology, 9*(4), 192. https: //doi.org/10.3390/jfmk9040192

[6] Vere, H. D. (2023). *The gamble of the first basket: A study on betting the first scorer in NBA games*. SSRN. https: //doi.org/10.2139/ssrn.4540820

[7] Alonso Pérez-Chao, E., Nieto-Acevedo, R., Scanlan, A. T., Lopez-García, A., Lorenzo, A., & Gómez, M. Á. (2024). Is there no place like home? Home-court advantage and home win percentage vary according to team sex and ability in Spanish basketball. *International Journal of Performance Analysis in Sport, 24*(5), 510–517. https: //doi.org/10.1080/24748668.2024.2325842

[8] Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics, 20*(8), 411–427. https: //doi.org/10.1002/(SICI)1099-1468(199912)20: 8< 411: : AID-MDE957> 3.0.CO; 2-G

[9] Zhang, S., Gomez, M. Á., Yi, Q., Dong, R., Leicht, A., & Lorenzo, A. (2020). Modelling the relationship between match outcome and match performances during the 2019 FIBA Basketball World Cup: A quantile regression analysis. *International Journal of Environmental Research and Public Health, 17*(16), 5722. https: //doi.org/10.3390/ijerph17165722

[10] Fischer-Baum, R. (2015, June 2). *How we calculate NBA Elo ratings. FiveThirtyEight.* Retrieved November 13, 2025, from https: //fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/

[11] Sarlis, V., Papageorgiou, G., & Tjortjis, C. (2024). Injury patterns and impact on performance in the NBA league using sports analytics. *Computation, 12*(2), 36. https: //doi.org/10.3390/computation12020036

[12] Leicht, A. S., Gomez, M. A., & Woods, C. T. (2017). Explaining match outcome during the men's basketball tournament at the Olympic Games. *Journal of Sports Science and Medicine, 16*(4), 468–474.

[13] Torres, E. (2022, March 4). *Linear regression: Predicting NBA regular season wins. Medium.* Retrieved April 24, 2026, from https: //medium.com/@entorres93/linear-regression-predicting-nba-regular-season-wins-bd1acd33e8e5

[14] Kvam, P., & Sokol, J. S. (2006). A logistic regression/Markov chain model for NCAA basketball. *Naval Research Logistics, 53*(8), 788–803. https: //doi.org/10.1002/nav.20170

[15] Dobson, N. (2019). *FiveThirtyEight's Elo ratings and logistic regression*. GitHub. Retrieved April 24, 2026, from https: //nicidob.github.io/nba_elo

[16] Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports, 3*(3), Article 1. https: //doi.org/10.2202/1559-0410.1070

[17] Terner, Z., & Franks, A. (2021). Modeling player and team performance in basketball. *Annual Review of Statistics and Its Application, 8*(1), 1–23. https: //doi.org/10.1146/annurev-statistics-040720-015536

[18] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https: //doi.org/10.1147/rd.33.0210

[19] Kohli, I. S. (2016). *Finding common characteristics among NBA playoff teams: A machine learning approach*. SSRN. https: //doi.org/10.2139/ssrn.2764396

[20] Galekwa, R. M., Tshimula, J. M., Tajeuna, E. G., & Kyandoghere, K. (2024). *A systematic review of machine learning in sports betting: Techniques, challenges, and future directions*. arXiv preprint. https: //doi.org/10.48550/arXiv.2410.21484

[21] Khatibi, A., King, D. M., & Jacobson, S. H. (2015). Modeling the winning seed distribution of the NCAA Division I men's basketball tournament. *Omega, 50*, 141–148. https: //doi.org/10.1016/j.omega.2014.08.004

[22] Sandholtz, N., & Bornn, L. (2020). Markov decision processes with dynamic transition probabilities: An analysis of shooting strategies in basketball. *The Annals of Applied Statistics, 14*(3), 1511–1532. https: //doi.org/10.1214/20-AOAS1348

[23] Chen, W.-J., Jhou, M.-J., Lee, T.-S., & Lu, C.-J. (2021). Hybrid basketball game outcome prediction model by integrating data mining methods for the National Basketball Association. *Entropy, 23*(4), 477. https: //doi.org/10.3390/e23040477

[24] Štrumbelj, E., & Vračar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting, 28*(2), 532–542. https: //doi.org/10.1016/j.ijforecast.2011.01.004

[25] Sarlis, V., Chatziilias, V., Tjortjis, C., & Mandalidis, D. (2021). A data science approach analysing the impact of injuries on basketball player and team performance. *Information Systems, 99*, Article 101750. https: //doi.org/10.1016/j.is.2021.101750

[26] Hu, J., & Zidek, J. V. (2004). Forecasting NBA basketball playoff outcomes using the weighted likelihood. In *Lecture Notes–Monograph Series* (Vol. 45, pp. 385–395). Institute of Mathematical Statistics.

[27] Estrella, E. O. (2024). Exploring sports analytics of coaches: An optimization for team performance and strategy development. *Pakistan Journal of Life and Social Sciences, 22*(2), 3044–3053. https: //doi.org/10.57239/pjlss-2024-22.2.00384

[28] Matej, U., Gustav, Š., Ondřej, H., & Filip, Ž. (2021). Optimal sports betting strategies in practice: An experimental review. *IMA Journal of Management Mathematics, 32*(4), 465–489. https: //doi.org/10.1093/imaman/dpaa029

[29] Cao, C. (2012). *Sports data mining technology used in basketball outcome prediction* [Master's thesis, Technological University Dublin].

[30] Jones, E. S., & Magel, R. C. (2016). Predicting outcomes of NBA basketball games. *Journal of Advance Research in Business Management and Accounting, 2*(5), 1–13. https: //doi.org/10.53555/nnbma.v2i5.99

[31] Ruiz, F. J. R., & Perez-Cruz, F. (2015). A generative model for predicting outcomes in college basketball. *Journal of Quantitative Analysis in Sports, 11*(1), 1–12. https: //doi.org/10.1515/jqas-2014-0055

[32] Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association, 111*(514), 585–599. https: //doi.org/10.1080/01621459.2016.1141685