

# *Random Forest Prediction of NBA Regular Season MVP Winners Based on Metrics Optimization*

Jiuxi Han<sup>1</sup>, Zhipeng Yu<sup>2,3,\*</sup>

<sup>1</sup>*Beijing Huiwen Middle School (Peking Academy), Beijing, 100061, China*

<sup>2</sup>*The Open University of China, Jinan, 250014, China*

<sup>3</sup>*Hanting Administration for Market Regulation, Weifang, 261100, China*

*\*Corresponding author*

**Keywords:** Indicator Optimization, Correlation Coefficient, Random Forest, Data Mining, MVP

**Abstract:** In the National Basketball Association (NBA), there are various individual awards, but one of them, the Most Valuable Player Award, is considered crucial. This award represents a player's outstanding performance in a team sport, and its recipient not only plays an important role on the court, but also attracts widespread attention and discussion in society and the sports world. An effective MVP prediction model can synthesize a player's statistical performance and team success. This study proposes a random forest algorithm to predict seasonal MVPs. In this study, seasonal statistics of 300 players were analyzed using nearly 50 years of NBA seasonal data with an era split: early NBA vs. small ball era. Correlation analysis was used to eliminate interdependent criteria. As a result, the number of criteria was reduced from 20 to 7, which were defined as decision factors: (i) Points Per Game (PPG), (ii) Field Goal Percentage (FG%), (iii) Three-Point Percentage (3P%), (iv) Rebounds Per Game (RPG), (v) Assists Per Game (APG), and (vi) Turnovers Per Game (TOV). After correlation analysis, the Random Forest algorithm was applied in order to predict MVP for the years 2021-2023. The results of the study clearly show that the use of the Random Forest algorithm to predict MVP is a highly feasible method with excellent adaptability and foresight. The high feasibility of this approach, which was able to provide accurate predictions across seasons and tournament environments, further validates its potential for application in the NBA. The excellent adaptability of the Random Forest algorithm means that it is able to effectively deal with a wide range of decision factors and data characteristics, including in the context of data in the underdog era, which can change frequently. This allows us to better cope with evolving sports environments and data challenges and provide reliable support to decision makers. As basketball and sports data continue to grow and become richer, and as machine learning techniques continue to evolve, the Random Forest algorithm will hopefully further improve its performance and range of applications. It can be a powerful tool for decision makers and team managers to help them make more informed decisions, identify the players with the highest potential to win MVP awards, and contribute to team success.

## 1. Introduction

The NBA (National Basketball Association), one of the most prominent professional basketball leagues in the United States, not only attracts enthusiastic support from basketball fans around the globe, but also serves as a role model for many young basketball players in their pursuit of excellence. Each NBA team represents a sense of pride and belonging to a city or region, and fans go all out to cheer and support the team they support. This passion is evident not only on game day, but also in fan interactions, fan club events, and social media. NBA season games and playoffs are events that generate fervor among basketball fans around the globe, with team rivalries and player performances becoming widely discussed topics. Whether it's discussing which team is superior or admiring the skill of each player, the spirit of competition always generates heated debate and excitement. At the end of the NBA season, in addition to team honors, individual players have the opportunity to receive a variety of awards in recognition of their outstanding performance in the game. These awards include the season's Most Valuable Player (MVP), Best Rookie, Best Defensive Player, Best Scorer, Best Assists, and more. These awards were created to recognize outstanding players, motivate other players, and raise the level of competition in the game. The NBA is not just a sport, it is a cultural and social phenomenon. It brings together fans from all over the world and promotes social interaction and community cohesion. The league provides an experience of fun, passion and teamwork. On and off the court, the NBA has become part of what shapes our lifestyles and values. This league is not only a temple of basketball, but also a mirror of society, reflecting the core values of unity, collaboration, and the pursuit of excellence [1].

While the NBA organization awards a variety of individual awards, there is still a passionate discussion amongst them about the method of selecting the winner of the MVP award. On this issue, it is widely recognized that selecting the MVP award winner requires a combination of factors, including statistical performance as well as the team's successful performance throughout the season. This approach aims to ensure that the MVP award does not just reward outstanding players, but also considers their overall impact on the team and the game. Sports scientists identify the key factors that influence a basketball player's performance and assist coaches in developing more data-supported criteria for evaluating basketball tactics and performance by constructing sophisticated decision-analytic models to assess a player's overall ability. This data-driven approach is expected to play an active role in improving player training, developing tactical strategies, recruiting players and evaluating overall team performance. Through a scientific approach, the game of basketball can evolve more efficiently, paving the way for player and team success. Sarlis & Tjorjjs [2] explored basic and advanced basketball metrics used in the NBA and European leagues in their study and provided a thorough literature review and performance analysis methodology in order to assess team and player performance. Mertz [3] and his co-authors used a data-driven approach in their study. authors in their study used a statistical model to rank the top NBA players of all time and utilized a linear regression model to create a trustworthy list of the top 150 players in NBA history. Constructing this type of model is indeed a great challenge, as it requires the synthesis of a large number of individual player statistics and accomplishments, while also having to incorporate the impact of changes in the rules of the game and the dynamics of the game over the years on the analysis of individual player performance. In short, constructing this type of model requires in-depth data analysis and statistical modeling skills, as well as constant adaptation to the rapidly changing field of basketball. It is only through a combination of factors that individual player performance can be more accurately assessed and strong support and advice provided to teams.

There are a number of scholars who have proposed different approaches in sports-related data prediction. Chen [4] used a statistical model to predict who would win the 2017 NBA Most Valuable Player award and applied data mining discriminant analysis to group players. Watanabe et

al [5] used a streaming network model for visual analytics to model the set of low-dimensional latent variables and their generation process using National Basketball Association basketball team data. Hubacek and his team [6] proposed a new prediction system using machine learning techniques aiming to profit from the sports betting market. Ballı et al [7] used an artificial neural network to select the best team player in a basketball game.

However, the above methods are still deficient in the selection of predictions for MVP:

(1) For the selection of indicators, the indicators screened by the literature [4-5] contain the data of outstanding athletes in the past 50 years, but they do not consider that the pattern of basketball games has changed greatly over time, and the small-ball era in the past 20 years is very different from the pattern of rivalry and competition in the earlier era, which means that the direction in which we have to extract the characteristics of the player's information to determine the indicators has to be changed as well.

(2) The literature [6] models this problem using traditional statistical class models, an approach that often faces the challenge of providing a general model with excellent generalization. This is due to the fact that decision-making factors and player performance in basketball are affected by a variety of dynamics and complexities, many of which may vary from season to season and from team to team. Statistical models are usually based on known data, but the evolving and diverse nature of the data makes it difficult for these models to accurately capture these complexities, thus compromising their ability to generalize.

(3) Artificial neural networks used in the literature [7] typically exhibit excellent performance when faced with prediction and classification tasks on large-scale datasets, high-dimensional data, and complex patterns. For the two periods of sports data, the early NBA and the small-ball era, the data we have is usually limited and only contains information from the last 20 years, i.e., neural networks are not applicable to the prediction of MVP.

Therefore, this paper employs the Random Forest algorithm, a machine learning technique that can effectively deal with limited datasets. By integrating multiple decision tree models, Random Forest is able to better handle the correlation between multiple decision factors, which improves generalization and performs well even in data contexts from the early NBA and small ball era. This integrated approach helps overcome the limitations of statistical models and traditional neural networks with small datasets, thus providing reliable MVP predictions.

In this study, we propose to predict the performance of players through Random Forest Algorithm to select the MVP of a basketball league. The main objective of this study is to perform a correlation analysis of multiple decision factors and select the most weighted factor to be used as an input to the Random Forest Algorithm for seasonal MVP prediction. The paper is structured as follows: in Section 2, we introduce the proposed methodology and provide the context for predicting MVP using Random Forest; Section 3 presents a case study of two categories of NBA seasons (early NBA era vs. small-ball era) and summarizes our findings. The results from the case studies demonstrate how the methodology proposed in this paper effectively handles individual decision factors to accurately predict seasonal MVPs. Finally, we present a comprehensive discussion and provide a conclusion that encapsulates the key insights derived from our study. In addition, we outline and summarize opportunities for future research in Section 4.

## 2. Methodology

Researchers have long been fascinated by problems involving predictive ranking because in real life, ranking these problems occur frequently. Many practices are encountered in our daily lives, such as identifying candidates to be interviewed, prioritizing projects, and ranking the best performing choices. Although systems that use Random Forests to predict MVPs can make more

thorough and multidimensional predictions, they have significant drawbacks such as scaling and aggregation, where "scaling" refers to the need to deal with a large number of input features or data, and "aggregation" refers to the need to consider multiple features together to make a final MVP prediction. Random forest prediction methods have been widely used in the literature to determine the most accurate rankings by jointly evaluating multiple dimensions.

To determine the most accurate ranking by evaluating multiple dimensions together, to solve the aggregation problem, we can utilize the correlation coefficient, which is a useful tool for measuring the correlation between multiple dimensions. By calculating the correlation coefficient between these dimensions, it is possible to determine how much they influence each other. If there is a high degree of positive correlation between dimensions, then they are likely to have a similar influence in the rankings, and therefore we can combine them into a single composite factor. On the contrary, if the correlation between the dimensions is low, we can keep them as independent dimensions to ensure that we do not overlook any important information

In the following, we explain the steps of the proposed approach to solve the problem of predicting MVP.

Step 1. Data collection and screening, in this paper, we collect various game indicators of each athlete in the last 50 years from the NBA official website. The data are initially screened to exclude duplicated, missing or abnormal data to ensure the quality and completeness of the data.

Step 2. Split the NBA into eras according to the changes in game styles, player characteristics and league rules in each era: the early NBA and the small-ball era.

Step 3. Correlate the indicators of each player, analyze the correlation matrix, and based on the results of the correlation analysis and paying attention to the impact of indicators with high correlation coefficients on the performance of the players, select the decision factors that will ultimately be used to predict the MVP.

Step 4. Further analyze the selected decision factors to understand their weights and impacts in the MVP prediction model to ensure that they reasonably reflect the overall performance of the players

Step 5. Random forest algorithm

Step 5.1. Build the base learner for Random Forest using CART regression tree: where the training set, we use is denoted as  $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$ , the input sample  $x_i = (x_i^1, x_i^2, \dots, x_i^Z)$  contains  $Z$  attribute variables, and the output  $Y = (y_1, y_2, \dots, y_n)$  is a continuous value.

Step 5.2. The nodes of the regression tree set a cut point  $s$  for the attribute variable  $j$  of sample  $x_i (1 < i < n)$ . This input variable will be partitioned into one region for pairs greater than  $s$  and into another region otherwise. The region obtained from the division is further divided using different attribute variables, based on the node's cut-off point will be divided into  $m$  regions, respectively, denoted as  $R_1, R_2, \dots, R_m$ . Define the output value of each region as  $c_1, c_2, \dots, c_m$ , respectively.

Then CART is modeled as Equation (1):

$$f(x) = \sum_{m=1}^m c_m I(x \in R_m) \quad (1)$$

$$\text{where, } I(x \in R_m) = \begin{cases} 1(x \in R_m) \\ 0(x \notin R_m) \end{cases}$$

The squared error of the regression tree model is Equation (2):

$$E = \sum_{x_i \in R_m} (y_i - f(x_i))^2 = \sum_{x_i \in R_m} (y_i - \sum_{m=1}^m c_m I(x \in R_m))^2 \quad (2)$$

Step 5.3. Suppose that the variable  $x^{(j)}$  in the sample is selected as the cutoff variable and the node takes the value  $s$  as the cutoff point. Comparison of variable  $J$  in the input sample with the cut-off point  $s$  yields region  $R_1(j, s) = \{x | x^{(j)} \leq s\}$  and region  $R_2(j, s) = \{x | x^{(j)} > s\}$ . When  $j$  and  $s$  are set to definite values, the samples contained in regions  $R_1(j, s)$  and  $R_2(j, s)$  are also definite. Therefore, it is necessary to determine the output values  $c_1$  and  $c_2$  for each region to minimize the squared difference on the respective intervals as in Equation (3):

$$\min_{j,s} [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \quad (3)$$

Then  $\hat{c}_1 = \text{ave}(y_i | x_i \in R_1(j, s))$ ,  $\hat{c}_2 = \text{ave}(y_i | x_i \in R_2(j, s))$ .

Step 5.4. Iterate through all the variables in the sample, the optimal cut-off point  $s$  of different cut-off variables gets the smallest squared error recorded as the optimal cut-off variable  $J$ . Similarly, the cut-off region is further divided to find the optimal cut-off variable and cut-off point,

$$f(x) = \sum_{m=1}^m c_m I(x \in R_m)$$

and finally the regression tree is obtained.

Step 6. Based on the collected and analyzed data, the six-game metrics of Points Per Game (PPG), Field Goal Percentage (FG%), Three-Point Percentage (3P%), Rebounds Per Game (RPG), Assists Per Game (APG), and Turnovers Per Game (TOV) are used as our decision-making factors in order to predict the MVP winners for the years 2021-2023.

Step 7. Run principal component analysis to visualize the data metrics by dimensionality reduction, give reasonable explanations and analyze the MVP players.

### 3. Case Study and Results

#### 3.1 Problem Definition and Data

In this study, regular season statistics from 1980-2020 were obtained from the official NBA page to determine the MVP for the 2021-2023 NBA regular season. These statistics are: Games played (GP), Minutes Per Game (MPG), Points Per Game (PPG), Field Goals Made (FGM), Field Goal Attempts (FGA), and Field Goal Percentage (FG%), 3-Points Made (3PM), 3-Point Attempts (3PA), 3-Point Percentage (3P%), Free Throws Made (FTM), Free Throw Attempts (FTA), Free Throw Percentage (FT%), Offensive Rebounds (ORB), Defensive Rebounds (DRB), Rebounds Per Game (RPG), Assists Per Game (APG), Steals Per Game (SPG), Blanks Per Game (BPG), Field Goal Turnovers (TOV), and Field Goal Percentage (PF). In order to better examine the factors affecting the predicted MVP decision below, we visualize the main data of MVP winners from 2010-2020 as shown in Figure 1.

Over time, the NBA (National Basketball Association) has gone through a number of different stages of development, with the two most notable periods being the early NBA and the small ball era. These two periods had significant differences in the style, tactics, and culture of the game of basketball, and therefore led to changes in the decision-making factors for selecting MVPs. In this article, we will explore the key differences between the early NBA and the small-ball era to help understand how these two periods have shaped the modern NBA. Here are some of the key

differences.

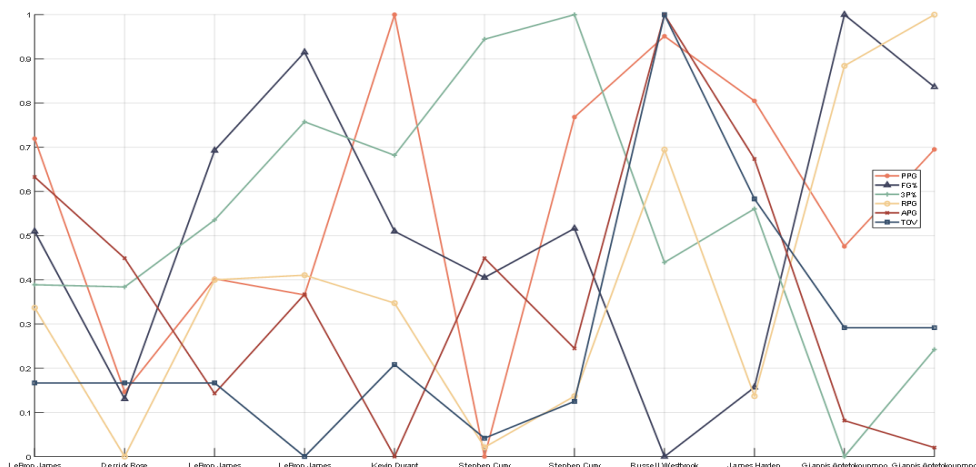


Figure 1: Data visualization

(1) Team styles:

Early NBA: In the early NBA, the style of basketball played was more interior offense and physical, with teams usually focusing on finding scoring opportunities down low, with big men being more prominent. Games were usually more physical and lower scoring.

Small-ball era: The small-ball era emphasized fast-paced, outside shooting and teamwork. Teams preferred small-ball lineups with an increased percentage of three-point shots and fast-paced drives to the hoop. This resulted in increased scoring and more entertaining basketball games.

(2) Positional Flexibility:

Early NBA: In the early days, players had relatively fixed positional divisions such as center, forward and guard. Each position had clear roles and responsibilities.

Small Ball Era: Small Ball Era emphasizes positional flexibility, players are no longer limited to the traditional positional division, they can take different roles on the court, for example, big players can also shoot three points, small players can also move in the interior.

(3) The importance of the three-point shot:

Early NBA: In the early days, three-pointers were not an important part of the game's scoring and teams relied more on mid-range and inside scoring.

Small-ball era: The small-ball era elevated the three-point shot to a key position, with many teams focusing on outside shooting and three-point attempts becoming one of the primary ways to score in a game.

(4) Defensive Strategies:

Early NBA: Early NBA defenses focused more on physical play, including strong man-to-man defense and defensive rebounding. Defensive teams focused more on interior protection.

Small Ball Era: Defensive strategies in the small ball era focused more on team defense, including quick transition defense and blocking outside shots. Teams typically pursued more steals and fast break scoring.

Due to the high number of decision factors and the possibility of duplication, we calculate the correlation coefficients of all the decision factors of the early NBA and the small ball era with each other, and the results are shown in Fig. 2 and Fig. 3.



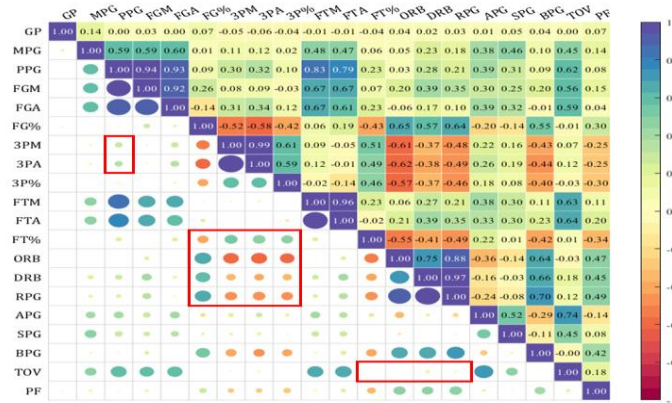


Figure 2: Early NBA

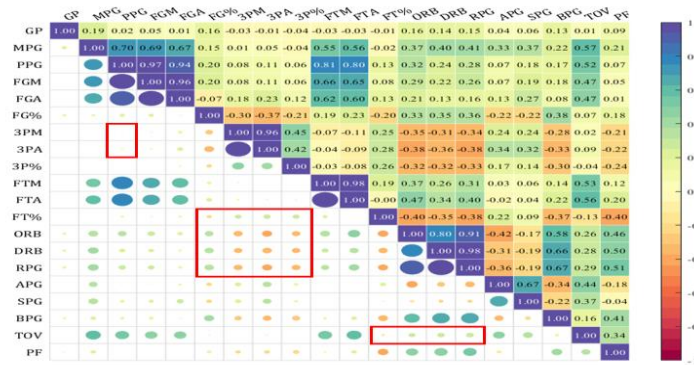


Figure 3: The era of small ball

The areas boxed out in Figure 2, as well as Figure 3, are the parts of the data that have changed the most in the comparison of the early NBA to the small-ball era, and we can reduce the impact of the final discrepancy on the projected results simply by analyzing these areas.

(1) The correlation between points per game (PPG) and three-point field goal attempts (3PM) and three-point field goal attempts (3PA) is significantly higher due to the fact that three-point attempts became one of the game's primary ways of scoring in the small-ball era. In the early days of the NBA, scoring relied heavily on mid-range shooting and inside scoring, so the correlation between points per game (PPG) and three-point field goal attempts (3PM) and three-point field goal attempts (3PA) was relatively low. However, in the small-ball era, teams have focused more on outside shooting, and the number of three-point attempts and hits has increased dramatically. This has led to the following two scenarios:

1) Increased positive correlation: due to the increased scoring opportunities from three-point range, points per game (PPG) will typically be positively correlated with three-point field goals made (3PM), i.e., as players make more three-pointers, their scoring will increase.

2) Negative correlation may decrease: although the number of three-point attempts (3PA) increases, due to the high percentage of three-pointers made, this may lead to a decrease in the negative correlation between points per game and three-pointers made. This means that players can take more three-point shots without significantly decreasing their scoring.

The small-ball era of basketball did change the correlation between field goal percentage and three-point attempts and hits to a closer and more positive correlation. This trend continues to persist in the modern NBA, reflecting significant changes in basketball tactics and styles.

(2) In the small-ball era, the correlation between turnovers per game (TOV) and free throw percentage (FT%), offensive rebounds (ORB), defensive rebounds (DRB), and rebounds per game (RPG) has decreased significantly, likely due to changes in tactics, techniques, and player roles.

1) Changed Offensive Strategies: The small-ball era emphasized fast-paced, outside shooting and quick transition offense. Teams are focusing more on outside shooting and fast breaks rather than relying on offensive rebounds. This reduces the correlation between turnovers per game and offensive rebounds as players focus more on outside shooting and fast breaks rather than fighting for offensive rebounds. iii.

2) Increased Outside Shooting: An increase in outside shooting could lead to a decrease in the correlation between free throw shooting and turnovers per game. Free throws are usually associated with inside scoring and contact, while outside shooting is usually less associated with those situations, so the impact of free throw shooting on field goal percentage may be diminished.

3) Greater Positional Flexibility: The small-ball era encourages positional flexibility, where players can play different roles on the court. This may lead to a lower correlation between turnovers per game and defensive rebounds and rebounds per game, as players become more diverse in their responsibilities and positions and are no longer limited by traditional positional divisions of labor.

4) More three-point attempts: The small-ball era has seen a significant increase in the number of three-point attempts, which may also reduce the correlation between turnovers per game and rebounding statistics. Because three-point attempts do not usually lead to rebound contests, the link with rebounding data is weakened.

These changes reflect the significant impact of the small-ball era on the overall tactics and style of basketball play. As a result, the correlation between turnovers per game and free throw percentage, offensive rebounds, defensive rebounds, and rebounds per game is significantly lower in the small-ball era, unlike in the early NBA.

(3) Longitudinal Comparison:

The correlation between field goal percentage (FG%) and free throw percentage (FT%), offensive rebounds (ORB), defensive rebounds (DRB), and rebounds per game (RPG) is significantly higher in the Small Ball Era.

1) Offensive Rebounds (ORB) vs. Field Goal Percentage (FG%): In the Small Ball Era, teams have focused more on outside shooting, which has led to more offensive rebounding opportunities, as outside shooting typically generates more rebounding opportunities. ii. The increased aggressiveness of players in offensive rebounding further increases the correlation with shooting percentage.

2) Free Throw Shooting (FT%) vs. Field Goal Percentage (FG%): In the Small Ball Era, outside shooting opportunities have increased, and free throw attempts are likely to increase as well, which leads to an increased correlation between free throw shooting and field goal percentage. Outside shots usually do not include free throws, so the link between hits and free throw attempts may have increased.

(4) Horizontal Comparison:

In the small-ball era, the correlation between offensive rebounds (ORB) and field goal percentage (FG%), three-point field goals made (3PM), three-point field goal attempts (3PA), and three-point field goal percentage (3P%) has increased significantly; the correlation between rebounds per game (RPG) and field goal percentage (FG%), three-point field goals made (3PM), three-point field goal attempts (3PA), and three-point percentage (3P%) is slightly higher.

1) Increased Correlation between Offensive Rebounds (ORB) and Three-Point Shooting: In the Small Ball Era, the importance of the offensive rebound (ORB) has increased because it provides a second chance to score. This has led to an increase in the correlation between offensive rebounding and three-point attempts (3PA) and three-points made (3PM) as players compete more aggressively



for the rebound to provide opportunities for outside shots.

2) a slight increase in the correlation between rebounds per game (RPG) and threes: in the small ball era, teams focus more on rebounding to protect the rim to prevent opponents from scoring on fast breaks. This increases the correlation between rebounds per game (RPG) and three-point attempts (3PA) and three-points made (3PM), as rebounding statistics correlate with outside shooting. iii.

The more open style of play of the small-ball era may have reduced the intensity of the matchups, which affected free throw attempts. At the same time, less favorable matchups than in the early NBA may have led to more outside shooting, further enhancing the correlation between shooting percentage (FG%) and other statistics. Together, this set of changes reflects significant differences in tactics, technique, and style of play between the small-ball era and the early NBA.

Combining these analyses, we ultimately selected the six game metrics of Points Per Game (PPG), Field Goal Percentage (FG%), Three-Point Percentage (3P%), Rebounds Per Game (RPG), Assists Per Game (APG), and Turnovers Per Game (TOV) as our decision-making factors for predicting the 2021-2023 MVP.

### 3.2 Results & Analysis

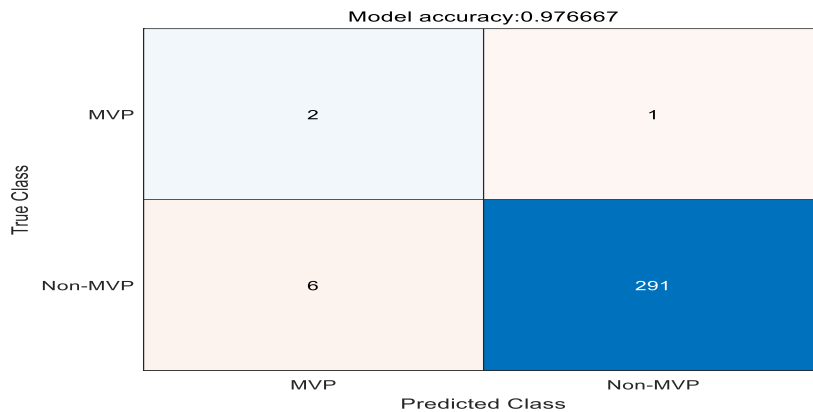


Figure 4: Predictive confusion matrix

The results from Fig. 4 show that out of the total 300 players predicted, the following conclusions were made: out of the 8 players predicted to be MVPs, 6 were actually not really MVPs, which is a case of misclassification. Of those 8 players, 2 were actually MVPs, which the model predicted correctly. And of the 292 non-MVPs, 1 was actually the MVP, which the model missed.

Table 1: Indicators for the analysis of forecast results

Evaluation indicators	Score
True positive examples	2
False positive examples	6
False negative examples	1
Accuracy rate	25%
Completeness rate	66.7%

In Table 1, true positive examples are the samples that predicted MVP players correctly, false positive examples refer to the samples of non-MVP players predicted to be MVPs, true negative examples refer to the samples that predicted non-MVP players correctly, and false negative examples refer to the samples of MVP players predicted to be non-MVPs.

## 4. Conclusions

This study proposes a random forest algorithm to predict seasonal MVP in the NBA. In this study, seasonal statistics of 300 players were analyzed using nearly 50 years of NBA seasonal data, divided by era: early NBA and small ball era. Correlation analysis was used to eliminate interdependent criteria. The number of criteria was reduced from 20 to 7, which were defined as decision factors: field goal percentage, field goal percentage, three-point percentage, rebounds per game, assists per game, and turnovers per game. Through correlation analysis, the Random Forest algorithm was used to predict MVP from 2021-2023. The results of the study clearly show that predicting MVP using the Random Forest algorithm is a highly feasible method with good adaptability and foresight. The high feasibility of this method, which provides accurate predictions across seasons and tournament settings, further validates its potential for NBA applications. The excellent adaptability of the Random Forest algorithm means that it is able to effectively deal with a wide range of decision-making factors and data characteristics, including in the context of data in an era of disadvantage, which can change frequently. This allows us to better cope with changing sports environments and data challenges and provide reliable support to decision makers. As basketball and sports data continue to grow and become richer, and as machine learning techniques continue to evolve, the Random Forest algorithm is expected to further improve its performance and range of applications. It is a powerful tool for decision makers and team managers to help them make more informed decisions, identify the players most likely to win MVP awards, and contribute to team success.

## References

- [1] Lacey R., & Kennett-Hensel P. (2016). *How Expectations and Perceptions of Corporate Social Responsibility Impact NBA Fan Relationships*. *Sport Marketing Quarterly*, 25(1).
- [2] Sarlis V., & Tjortjis C. (2020). *Sports analytics—Evaluation of basketball players and team performance*. *Information Systems*, 93, 101562.
- [3] Mertz J., Hoover L. D., Burke J. M., Bellar D., Jones M. L., Leitzelar B., & Judge W. L. (2016). *Ranking the greatest NBA players: A sport metrics analysis*. *International Journal of Performance Analysis in Sport*, 16(3), 737-759.
- [4] Chen M. (2017, June). *Predict NBA Regular Season MVP Winner*. In *IEOM South American Conference Proceedings* (pp. 44-51).
- [5] Watanabe R., Ishibashi H., & Furukawa T. (2022). *Visual analytics of set data for knowledge discovery and member selection support*. *Decision Support Systems*, 152, 113635.
- [6] Hubáček O., Šourek G., & Železný F. (2019). *Exploiting sports-betting market using machine learning*. *International Journal of Forecasting*, 35(2), 783-796.
- [7] Ballı S., Karasulu B., Uğur A., & Korukoğlu S. (2011). *Basketbolda oyuncu se çimi i çin sinirsel-bulanık karar destek sistemi*. *İTÜDERGİSİ/d*, 8(1).