

Article

# Regression Tree Model for Predicting Game Scores for the Golden State Warriors in the National Basketball Association

Mei-Ling Huang \*  and Yi-Jung Lin

Department of Industrial Engineering and Management National Chin-Yi University of Technology, Taichung 411, Taiwan; s4a815002@student.ncut.edu.tw

\* Correspondence: huangml@ncut.edu.tw

Received: 28 April 2020; Accepted: 16 May 2020; Published: 19 May 2020



**Abstract:** Data mining is becoming increasingly used in sports. Sport data analyses help fans to understand games and teams' results. Information provided by such analyses is useful for game lovers. Specifically, the information can help fans to predict which team will win a game. Many scholars have devoted attention to predicting the results of various sporting events. In addition to predicting wins and losses, scholars have explored team scores. Most studies on score prediction have used linear regression models to predict the scores of ball games; nevertheless, studies have yet to use regression tree models to predict basketball scores. Therefore, the present study analyzed game data of the Golden State Warriors and their opponents in the 2017–2018 season of the National Basketball Association (NBA). Strong and weak symmetry requirements were identified for each team. We developed a regression tree model for score prediction. After predicting the scores of each player on two teams, we summed and compared the predicted total scores to obtain the predicted results (lose or win) of the team of interest. The results of this study revealed that the regression tree model can effectively predict the score of each player and the total score of the team. The model achieved a predictive accuracy of 87.5%.

**Keywords:** National Basketball Association; regression tree; linear regression; game points prediction

## 1. Introduction

Advanced statistical methods were commonly used in various studies. A soft computing model used a learning approach for addressing data management over social networks [1]. Dulebenets et al. [2] applied regression models to estimate the effects of various factors on the driving ability of individuals. Andrée et al. [3] estimated a penalized non-parametric model of environmental output across economic development. A multivariate random parameter Tobit model was utilized to determine the factors that drive both the crash occurrence probability and the crash rate of 65+ roadway users [4]. Narasingam et al. [5] applied sparse regression to determine the structure of reduced-order model on a hydraulic fracturing process.

Since the 20th century, sport has grown globally [6,7]. Professional sporting games, such as basketball, baseball, tennis, and golf, and events such as the World Football Championship and the Olympic Games not only attract the attention of many fans but also continue to create extremely high output value for the sport industry. The Internet has enabled sport betting to develop rapidly. In most countries, sport betting is inextricably linked to professional sport. In sport betting, increasing the betting win rate requires—in addition to subjective judgements—predictions from historical data on the game, such as predictions of total results, total over or under, handicaps, and point spreads.

Scholars in most studies have applied linear regression models to predict the scores of various ball games. However, studies have yet to use the regression tree method to predict basketball scores.

Regression trees are similar to classification trees and are easy to understand and interpret. Unlike classification trees, the regression tree is suitable for continuous or ordered discrete dependent variables, and the prediction error of a regression tree is usually measured by the square difference between the actual value and the predicted value (see [8]). Regression trees have been used for predictions in several areas; for example, they have been used for predicting short-term algal blooms in the field of environmental engineering [9], for predicting student performance in teaching [10], for predicting the ultimate bearing capacity of shallow foundations in cohesive soils in civil engineering [11], and for demand analysis in economics [12]. Therefore, in the present study, we used the regression tree method to predict the scores of basketball games. Single-season (2017–2018) match data from the National Basketball Association (NBA) were used to predict the scores of each player on two teams. After the predicted scores of the two teams were added and compared, the team's predicted win or loss results could be obtained. The linear regression and support vector regression were also utilized in this study. Results from regression tree model, linear regression model, and support vector regression model were compared.

## 2. Related Studies

Several scholars have devoted attention to predicting the results of various sporting events and have conducted research on basketball, baseball, football, cricket, and other ball games [13–16]. Thabtah et al. [13] applied naive Bayes, neural network-like, and decision tree machine learning methods to various feature sets, in order to construct prediction models. By comparing the respective prediction accuracy rates, they could select the model with superior performance and determine the key factors affecting the results of the game. Valero [14] analyzed 10 years of Major League Baseball (MLB) regular season game data, using four data mining methods, namely lazy learners, artificial neural networks, support vector machines, and decision trees; the goal was to evaluate the abilities of the aforementioned classification- and regression-based methods in predicting game outcomes (home team win or lose) in MLB games. Razali et al. [15] used Bayesian networks to predict home victories, away victories, and draws in the English Premier League. Pathak et al. [16] applied modern classification techniques, namely naive Bayes, support vector machine, and Random Forest, to predict the outcome of the One Day International (ODI) cricket match.

Loeffelholz et al. [17] collected 620 NBA games and used neural networks to predict the success of basketball teams. The selection of features input to the neural networks as the most salient features for prediction from signal-to-noise ratios and expert opinions was also discussed in this study. Cao [18] collected the data of five regular NBA seasons and applied machine learning algorithms to build models for predicting the NBA game outcomes.

In addition to the prediction of victory or defeat, many scholars have studied team scores. For example, Harville [19] proposed a linear model for predicting differences in scores for college basketball or football games, using the difference in team effects plus or minus the home-field or field advantage. To fit a relevant linear model, Harville proposed an improved method of least-squares estimation and applied team estimates, in order to rank teams in the league. The study findings revealed that the results of the playoffs could be effectively predicted. Karlis and Ntzoufras [20] proposed bivariate Poisson models for analyzing goals scored by two teams and adjusted the models to increase the probability of a draw. Adam [21] attempted to extend the bivariate Poisson method through a generalized linear model. The score was modeled as the joint probability of a Poisson distribution representing the total number of goals and a binomial distribution representing a team goal.

Wheeler [22] first used the chi-square test to screen input variables; the threshold was set to 0.05, excluding features not exceeding the threshold, and 16 variables were finally obtained. Linear regression was then used to calculate the average score of each player according to the characteristic variables and to obtain the sum of the predicted scores for the two players on the field. Finally, the results of the two teams' matches (win/lose) were compared. In addition, to compare the performance of the linear regression model with other benchmark models, feature variables were input into naive Bayes

and support vector machine (SVM) classifiers, to obtain classification results. The linear regression output value was converted into two classification values for comparison. The linear regression model predicted a player's average score, and the converted win–loss classification result had an error rate of 53%. The naive Bayes and SVM classifier error rates were as low as 31%.

Singh et al. [23] proposed two separate models for predicting the match results for the ODI. They used data for the 2013 and 2014 ODI competitions as training and testing sets and conducted 10 cross-validations. Linear regression was used to predict the final score of the first round of ODI, and a naive Bayes classifier was used to estimate the probability of a team winning the second round. The results showed that, for prediction of the final score, the errors in the linear regression classifier were less than those for the current Run Rate method. As the game progressed, the accuracy of the naive Bayes prediction of the game results increased from 70% to 91%.

Wiseman [24] predicted the winning score of events on the PGA Tour, using first-round data. The author used linear regression, neural network regression, Bayesian linear regression, decision forest regression, and boosted decision tree regression models and compared the performance of the methods. Models were constructed by using data from 2004 to 2015 and validated by using the 2016 tournament. Correlation matrix analysis was conducted for various features. The first-round lead score, first-round average score, event, course yardage, and total prize money were selected as forecast indicators, and the R-squared and mean square error (MSE) values were used as evaluation indicators. The results revealed that the linear regression and Bayesian linear regression models were superior to the other models.

Lu et al. [25] analyzed games from 2012 to 2016 and established the least square fir model, using previous game results, team ability, and home advantages, based on data from five seasons, to predict the point difference for each team. Linear regression models depending on total over/under were fit to data before the all-star break and checked for adequacy, to predict final score difference between home and away NBA teams, for a regular season during 2011–2012 [26].

### 3. Materials and Methods

#### 3.1. Data

In this study, we considered the Golden State Warriors (GSW), one of the 30 NBA teams, for analysis. All teams competing with the GSW were regarded as opponents. The reason for choosing the GSW as the research object is that, in the traditional basketball concept, the closer to the basket score, the more solid the game will win, and the three-pointer is just a way to assist in scoring. However, the presence and rise of Stephen Curry subverts this traditional concept and opens up the modern basketball of “three-pointers”. In addition, the Golden State Warriors are the champions of the NBA's first season, and the team won a total of six league championships. Therefore, this study used the Golden State Warriors as the object of analysis. Other teams can also model and predict according to the proposed method of this article.

We executed the data collection step by capturing GSW player match data for the 2017–2018 season from the Basketball Reference website [27]. Because the final purpose was to obtain prediction results through prediction scores, GSW opponents' data must also be collected. After the collection of the data, the missing values were deleted. The missing data in this dataset comprised only two major items: Inactive and Player Suspended. We manually removed data fields indicating “Inactive” and “Player Suspended”. The number of records is not fixed for each player on the field. Some teams change players frequently. Each team has a total of 30 database fields, as shown in Table 1. The first eight items are related to the event and were not considered in the prediction model in this study. For the ninth item (i.e., Games Started), players in the starting lineup are usually the best players on the team (see [28]). If the best player acts as a starter but does not contribute to the team, the player is considered to have hindered the team. The corresponding variable was therefore selected to establish whether it is relevant to the score. Regarding the 10th item (i.e., Minutes Played), Martínez

and Martínez [29] indicated that no linear correlation exists between score and playing time. However, we used the M5Prime model tree algorithm (M5P) [30], which predicts nonlinear continuous data; therefore, this variable was selected and used in the prediction model. The 11th to 28th items pertain to personal data contributed by players to the team in the game. Among them, Field Goals, 3-Point Field Goals, and Free Throws are absolutely linearly related to the score ( $2 \times \text{FG} + 3\text{P} + \text{FT} = \text{PTS}$ ). Therefore, the above three items were excluded in the prediction model. In addition, the 28th item represents the player's score, which we used as an output item. Game Score (29th item) and Plus/Minus (30th item) are the player efficiency level and personal goal difference. Both items are calculated based on the player's personal data; therefore, they were not used in the prediction model. Finally, items considered to be removed from the dataset were player uniform number, rank, season game, date, age, team, home/away, opponent, field foals, 3-point field goals, free throws, game score, and plus/minus.

**Table 1.** Variables for each field in the database.

Items	Variable Name	Abbreviation	Whether to Choose
1	Player Uniform Number	No.	X
2	Rank	Rk	X
3	Season Game	G	X
4	Date	Date	X
5	Age	Age	X
6	Team	Tm	X
7	Home/Away	H/A	X
8	Opponent	Opp	X
9	Games Started	GS	V
10	Minutes Played	MP	V
11	Field Goals	FG	X
12	Field Goal Attempts	FGA	V
13	Field Goal Percentage	FG%	V
14	3-Point Field Goals	3P	X
15	3-Point Field Goal Attempts	3PA	V
16	3-Point Field Goal Percentage	3P%	V
17	Free Throws	FT	X
18	Free Throw Attempts	FTA	V
19	Free Throw Percentage	FT%	V
20	Offensive Rebounds	ORB	V
21	Defensive Rebounds	DRB	V
22	Total Rebounds	TRB	V
23	Assists	AST	V
24	Steals	STL	V
25	Blocks	BLK	V
26	Turnovers	TOV	V
27	Personal Fouls	PF	V
28	Points	PTS	V (Output)
29	Game Score	GmSc	X
30	Plus/Minus	+/-	X

Table 1 presents a summary of the variables considered in the regression tree model for this study, including the variable fields and variable abbreviations. The last column in the table indicates whether the variables were included in the model in this study. Sixteen input variables were used, and the output variables were the actual points.

The dataset was divided into a training set, validation set, and test set, and the ratio of the three sets was 6:2:2. In 82 games, the total number of players playing in each game is different. To avoid a scenario in which data of the 3rd player in the 50th game are assigned to the training set and data of

the 4th player in the same game are assigned to the verification set, we divided the dataset according to Season Game, with the first 50 fields representing the training set, the 51st–66th fields representing the validation set, and the 67th–82nd fields representing the test set.

### 3.2. Methods

The flowchart of the study procedure is illustrated in Figure 1. We considered three regression methods, namely regression tree, linear regression, and support vector regression models, for modeling, prediction, and comparison. After training three regression models through the training set, we used the three constructed models to predict the validation dataset and used the root mean square error (RMSE) as an error index (loss function) for the models. We determined the superior of the three aforementioned models and used it to predict player scores. This step was executed by using the test set. After predicting and summing up the scores of players in each field, we obtained the team's predicted total score. By comparing the two teams' predicted total scores, we could obtain the predicted match result; finally, we could compare the result with the actual result and calculate the accuracy rate of the predicted match result.

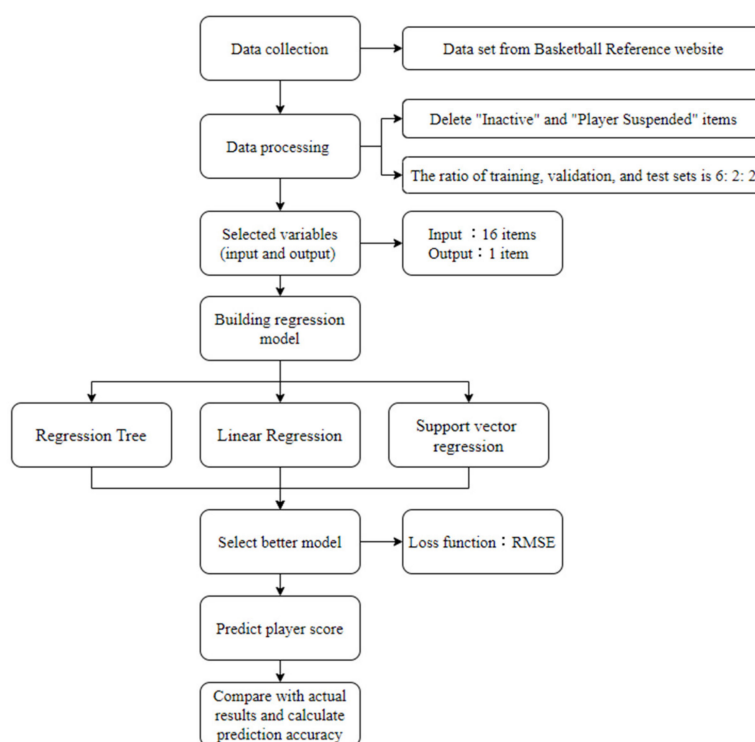


Figure 1. Research flowchart.

### 3.3. Regression Tree

The overall process of the regression tree method is similar to that of the classification tree method, and a prediction value is obtained at each node. Classification trees are used to process discrete data, whereas regression trees are used to process continuous data.

We constructed the model employed in this study by using the M5P tree regression algorithm in Weka software. M5P is a machine learning algorithm published by Wang and Witten in 1996 [31]. Its predecessor was M5, which was developed by Quinlan in 1992. Compared with traditional linear regression algorithms, M5P can accurately predict nonlinear data, and the rules and regression models are easy to interpret.

M5P is a binary regression tree model. The last node in the regression tree is a linear regression function that produces continuous numerical attributes. The M5P algorithm includes four main steps:

The first step entails dividing the input space into several subspaces, to create a tree. The variability in the subspace from root to node can be minimized by using segmentation criteria. The standard deviation of the value reaching this node is used to measure variability. The construction of the tree is completed by using a reduced standard deviation range (SDR) factor, which maximizes the expected reduction in errors on the nodes, as expressed in the following equation:

$$SDR = sd(S) - \sum_i \frac{S_i}{|S|} \times sd(S_i) \quad (1)$$

where  $S$  is the set of data records arriving at the node,  $S_i$  is the set obtained by dividing the node according to a given attribute, and  $sd$  is the standard deviation. The second step entails developing a linear regression model in each subspace, using the data associated with that subspace. The third step involves applying pruning techniques to overcome the problem of overtraining. However, the pruning process may cause a sharp interruption between adjacent linear models. The final step entails performing a smoothing process to compensate for the sharp interruption. The smoothing process combines all models from leaf to root to create the final model of the leaf. In the process, the predicted values of the leaves are filtered when they return to the root. The filtered values are combined with the predicted values through a linear regression of the node, as follows:

$$E' = \frac{ne + ka}{n + k} \quad (2)$$

where  $E'$  is the estimated value passed to the next highest node,  $e$  is the estimated value passed from below to the current node,  $a$  is the predicted value of the model at this node,  $n$  is the number of training examples that have reached the node, and  $k$  is a constant (see [31,32]).

### 3.4. Linear Regression

Linear regression is the simplest and most commonly used prediction model. A linear regression model predicts the linear relationship between continuous target variables and predicted variables, and many data items fulfil the basic assumptions of normal distribution and linear relationship.

Linear regression models can be divided into simple linear regression and multiple linear regression models. Simple linear regression models entail the use of a single independent variable ( $X$ ) to predict a dependent variable ( $Y$ ). The regression equation can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n \quad (3)$$

where  $Y_i$  is the actual observation value (variable) for the  $i$ th observation value of the dependent variable,  $Y$ ;  $X_i$  is the  $i$ th observation (variable) of the independent variable,  $X$ ;  $\beta_0$  is the parameter of the regression mode (termed the intercept or constant term);  $\beta_1$  is the parameter of the regression mode (termed the regression coefficient or slope);  $n$  is the number of observations; and  $\varepsilon_i$  is a random variable of the  $i$ th observation and belongs to a random error.

Multiple linear regression models entail the use of two or more independent variables to predict a dependent variable ( $Y$ ). The regression equation can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, \dots, n \quad (4)$$

where  $Y_i$  is the actual observation value of the  $i$ th observation value for the dependent variable,  $Y$ ;  $X_{ki}$  is the  $i$ th observation for the  $k$ th independent variable,  $X$ ;  $\beta_0$  is a parameter of the regression mode (termed the intercept);  $\beta_1, \dots, \beta_k$  is a parameter of the multiple regression mode (termed the regression coefficient);  $\varepsilon_i$  is a random variable of the  $i$ th observation value;  $n$  is the number of observations;  $k$  is the number of independent variables; and  $k > 0$  is a positive integer.

### 3.5. Support Vector Regression

Support vector regression (SVR) solves binary classification problems and has been proven to be an effective tool in real-value function estimation [33]. Like a regression method, the output of SVR is a real number. SVR finds an optimal hyperplane that balancing the model complexity and prediction error. The main advantages of SVR include that its computational complexity does not depend on the dimensionality of the input space, and it has excellent generalization capability, with high prediction accuracy. The prediction function of SVR is defined as follows [34]:

$$f(x) = (w, x) + b \tag{5}$$

where  $X$  denotes the space of the input patterns; and  $(w, x)$  denotes the dot product in  $X$ . If we minimize  $w$  and  $b$ , the optimization problem is defined as follows:

$$R_{svm}(C) = C \frac{1}{n} \sum_{i=1}^n L\varepsilon(d_i, y_i) + \frac{1}{2} \|w\|^2 \tag{6}$$

$$L\varepsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \varepsilon, & |d_i - y_i| \geq \varepsilon \\ 0 & , otherwise \end{cases} \tag{7}$$

where  $C \frac{1}{n} \sum_{i=1}^n L\varepsilon(d_i, y_i)$  is empirical error risk, which could be obtained from an  $\varepsilon$ -insensitive loss function in Equation (7);  $\frac{1}{2} \|w\|^2$  is a regularization term; and  $C$  is a regularization.

By introducing positive slack variables  $\xi_i$  and  $\xi_i^*$  into Equation (6), we get the following:

$$\begin{aligned} \text{Minimize} \quad & R_{svm}(w, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{Subject to} \quad & d_i - w\phi(x_i) - b_i \leq \varepsilon + \xi \\ & w\phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^*, \xi_i^* \geq 0 \end{aligned} \tag{8}$$

Lagrange multiplier is used to solve the optimization problem, and Equation (5) becomes the following:

$$f(x, a_i, a_i^*) = \sum_{i=1}^n (a_i - a_i^*) K(x, x_i) + b \tag{9}$$

where  $a_i$  and  $a_i^*$  are Lagrange multipliers, satisfying  $a_i^* a_i = 0$ ,  $a_i \geq 0$  and  $a_i^* \geq 0$  for  $i = 1, \dots, n$ , and the dual optimization problem is as follows:

$$\begin{aligned} \text{Max } R(a_i, a_i^*) &= \sum_{i=1}^n d_i (a_i - a_i^*) - \varepsilon \sum_{i=1}^n (a_i + a_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*) (a_j - a_j^*) K(x_i - x_j) \\ \text{Subject to } \sum_{i=1}^n (a_i - a_i^*) &= 0, \begin{cases} 0 \leq a_i \leq C, & i = 1, 2, \dots, n \\ 0 \leq a_i^* \leq C, & i = 1, 2, \dots, n \end{cases} \end{aligned} \tag{10}$$

### 3.6. Performance Evaluation

Several loss functions are used for regression, with commonly used functions being the MSE and mean absolute error (MAE). The MAE is an absolute value of the deviation between target and output value; therefore, positive and negative phases cannot cancel out. The MAE can thus effectively reflect the reality of the prediction error. Nevertheless, the MAE value is not differentiable at 0, and no method exists for determining the correction direction of a model through differentiation. The MSE overcomes this disadvantage but cannot easily be used to interpret data to obtain interpretable units. The solution is to use the RMSE to obtain an interpretable unit.

Accordingly, we used the RMSE to evaluate predictive performance. The RMSE is the square root of the ratio of the sum of all squared deviations of the predicted value from the actual values

to the number of observations,  $n$ . To explain the degree of dispersion of a sample, the RMSE can be minimized for nonlinear fitting. The RMSE formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (11)$$

where  $\hat{Y}_i$  is the predicted value, and  $Y_i$  is the actual value.

#### 4. Results

In this study, we used the training set to construct three regression models—regression tree, linear regression, and support vector regression models—through Weka software. We subsequently used the three regression models to predict the validation data and then calculated the RMSE values of the models, in order to determine the optimal model. Finally, the optimal model obtained from the training and validation sets was employed for prediction, using the test set to measure model performance. Results obtained by using data for the GSW as an example are described in the following sections.

##### 4.1. Information on Opponents

Table 2 lists the opponent data corresponding to Season Game for the test set. The total scores of the last 16 GSW games were predicted; subsequently, the total scores of the 11 teams in Table 2 were also predicted. By comparing the predicted total scores of the two teams, we obtained the predicted outcomes. Finally, we compared the actual results with the predicted results and calculated the accuracy of win or loss predictions.

**Table 2.** Opponent data corresponding to GSW test set data.

Season Game	Date	Opponent	Short Name
67	2018/3/11	Minnesota Timberwolves	MIN
68	2018/3/14	LA Lakers	LAL
69	2018/3/16	Sacramento Kings	SAC
70	2018/3/17	Phoenix Suns	PHX
71	2018/3/19	SA Spurs	SAS
72	2018/3/23	Atlanta Hawks	ATL
73	2018/3/25	Utah Jazz	UTH
74	2018/3/27	Indiana Pacers	IND
75	2018/3/29	Milwaukee Bucks	MIL
76	2018/3/31	Sacramento Kings	SAC
77	2018/4/1	Phoenix Suns	PHX
78	2018/4/3	Oklahoma City Thunder	OCT
79	2018/4/5	Indiana Pacers	IND
80	2018/4/7	New Orleans Pelicans	NOP
81	2018/4/8	Phoenix Suns	PHX
82	2018/4/10	Utah Jazz	UTH

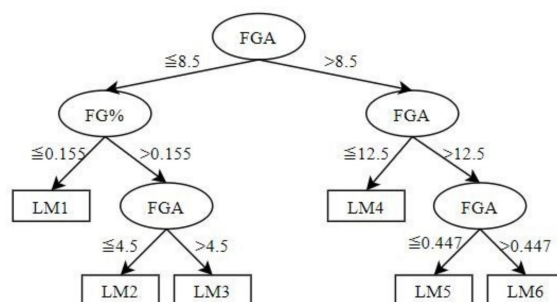
##### 4.2. Model Validation Results

We applied the training set to construct the model and used the validation set to establish the best model; we then obtained the prediction results for the two models, as well as the equations for the regression tree, linear regression, and support vector regression models. The modeling results are presented in Table 3. Figure 2 illustrates the regression tree, and Table 4 presents the regression equation. The linear regression equation is represented by Equation (12).



**Table 3.** Comparison of predictive performance of regression models, using GSW data.

Regression Models	Model Building Time (Seconds)	RMSE
Regression tree (M5P)	0.210	0.9645
Linear regression	0.001	1.3081
Support vector regression	0.310	2.4904



**Figure 2.** Regression tree diagram of GSW.

**Table 4.** Six regression equations for regression tree model for prediction of GSW scores.

Rule	Regression Equation
LM1	$PTS = 0.0696 * GS - 0.0003 * MP + 0.4028 * FGA + 0.8397 * FG\% - 0.0269 * 3PA + 0.5822 * 3P\% + 0.3586 * FTA + 1.5292 * FT\% - 0.0075 * ORB + 0.0149 * DRB - 0.0168 * TRB - 0.018 * BLK - 0.0056 * PF - 0.4345$
LM2	$PTS = 0.0375 * GS - 0.0001 * MP + 1.5187 * FGA + 5.7312 * FG\% - 0.0045 * 3PA + 2.1884 * 3P\% + 0.5562 * FTA + 0.8675 * FT\% - 0.0075 * ORB - 0.0407 * DRB - 0.0048 * TRB - 0.0156 * BLK - 0.0056 * PF - 4.5329$
LM3	$PTS = 0.0375 * GS - 0.0003 * MP + 1.1779 * FGA + 11.5318 * FG\% + 0.1554 * 3PA + 1.7568 * 3P\% + 0.5515 * FTA + 0.8305 * FT\% - 0.0075 * ORB + 0.0399 * DRB - 0.0581 * TRB + 0.0477 * AST - 0.0837 * BLK + 0.0804 * TOV - 0.0056 * PF - 6.7484$
LM4	$PTS = 0.3764 * GS - 0.0004 * MP + 1.0899 * FGA + 22.5574 * FG\% + 0.3394 * 3PA + 3.1752 * 3P\% + 0.755 * FTA + 1.0507 * FT\% - 0.0176 * ORB - 0.0132 * PF - 13.0728$
LM5	$PTS = 0.2379 * GS - 0.0007 * MP + 0.8717 * FGA + 32.494 * FG\% + 0.1633 * 3PA + 6.1986 * 3P\% + 0.9576 * FTA + 0.0606 * FT\% - 0.0176 * ORB - 0.0954 * DRB - 0.0132 * PF - 13.2189$
LM6	$PTS = 0.2379 * GS - 0.0015 * MP + 1.0394 * FGA + 32.1141 * FG\% + 0.4884 * 3PA + 5.535 * 3P\% + 0.8918 * FTA + 0.5658 * FT\% - 0.0176 * ORB - 0.0132 * PF - 17.2302$

#### 4.2.1. Regression Tree

According to Table 3, the regression tree model with the lowest RMSE was the optimal model. Although constructing the model would require a longer time than that required for the linear regression model, the RMSE value of the regression tree model was the lowest among the three models. Therefore, the regression tree model was used to predict player scores in the subsequent step.

#### 4.2.2. Linear Regression

The linear regression model for conducting predictions for the GSW can be expressed as follows:

$$PTS = 0.716 * GS + (-0.001) * MP + 1.049 * FGA + 6.03 * FG\% + 0.133 * 3PA + 4.635 * 3P\% + 0.933 * FTA + (-0.219) * ORB + (-0.164) * PF + (-2.055) \tag{12}$$

#### 4.3. Model Test Results

Consider, for example, the test set data of GSW players for March 11, 2018; each row in Figure 3 contains information on players who played for the GSW on that day. According to the rules derived

from the training set in Figure 2, we used the program to determine the rules (LM1–LM6) for each row of test set data. After judgement, the input variables were entered into the corresponding equations, to obtain the predicted PTS in the last column of Figure 4. The actual total score of the GSW team on March 11, 2018, was 103 points, and the team’s total score based on predictions was 108 points (Figure 4).

Date	GS	MP(s)	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
2018/3/11	0	448	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2018/3/11	0	1318	0	4	0	0	4	0	0	0	0	1	1	2	0	1	1	0	2	0
2018/3/11	0	423	1	1	1	0	0	0	1	2	0.5	1	1	2	0	0	0	1	2	3
2018/3/11	1	866	1	1	1	0	0	0	0	0	0	2	1	3	0	1	0	0	2	2
2018/3/11	0	1256	3	6	0.5	0	0	0	2	2	1	1	3	4	1	0	0	1	0	8
2018/3/11	1	1741	2	7	0.28	0	2	0	0	0	0	0	4	4	2	0	0	3	3	4
2018/3/11	1	2109	4	9	0.44	2	6	0.33	0	0	0	0	8	8	7	3	1	2	1	10
2018/3/11	0	1527	7	9	0.77	0	0	0	2	3	0.66	4	7	11	3	1	2	0	3	16
2018/3/11	1	2396	8	22	0.36	3	12	0.25	2	2	1	0	4	4	5	0	0	4	0	21
2018/3/11	1	2316	11	32	0.34	5	12	0.41	12	12	1	3	9	12	4	0	3	2	3	39

Figure 3. Regression tree rules derived through Excel.

Date	GS	MP(s)	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	rule	Predicted PTS		
2018/3/11	0	448	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	LM1	0		
2018/3/11	0	1318	0	4	0	0	4	0	0	0	0	1	1	2	0	1	1	0	2	0	LM1	1		
2018/3/11	0	423	1	1	1	0	0	0	1	2	0.5	1	1	2	0	0	0	1	2	3	LM2	4		
2018/3/11	1	866	1	1	1	0	0	0	0	0	0	2	1	3	0	1	0	0	2	2	LM2	3		
2018/3/11	0	1256	3	6	0.5	0	0	0	2	2	1	1	3	4	1	0	0	1	0	8	LM3	8		
2018/3/11	1	1741	2	7	0.28	0	2	0	0	0	0	0	4	4	2	0	0	3	3	4	LM3	5		
2018/3/11	1	2109	4	9	0.44	2	6	0.33	0	0	0	0	8	8	7	3	1	2	1	10	LM4	9		
2018/3/11	0	1527	7	9	0.77	0	0	0	2	3	0.66	4	7	11	3	1	2	0	3	16	LM4	17		
2018/3/11	1	2396	8	22	0.36	3	12	0.25	2	2	1	0	4	4	5	0	0	4	0	21	LM5	21		
2018/3/11	1	2316	11	32	0.34	5	12	0.41	12	12	1	3	9	12	4	0	3	2	3	39	LM5	40		
total																						103		108

Figure 4. Predictions of GSW total team scores on 11 March 2018.

We used the data in the remaining 15 test sets for the GSW and the data in the test sets for the opponent teams, to obtain the total score from each prediction. Each team input into the M5P training model was subject to several rules, with each rule having a corresponding regression equation. The relevant form of the opponent is presented in the Appendix A to Appendix C. After obtaining the predicted scores of both parties in all test sets, we could predict the GSW match results by comparing the predicted GSW scores with the opponent scores. Finally, we compared the predicted scores with the actual win or loss results and then calculated the accuracy of the predicted match results. Table 5 lists the predicted results, on the basis of the test set, of the last 16 games of the GSW, and their opponents in the 2017–2018 season. Figure 5 presents a line chart comparing the actual and predicted scores of the GSW and their opponents.

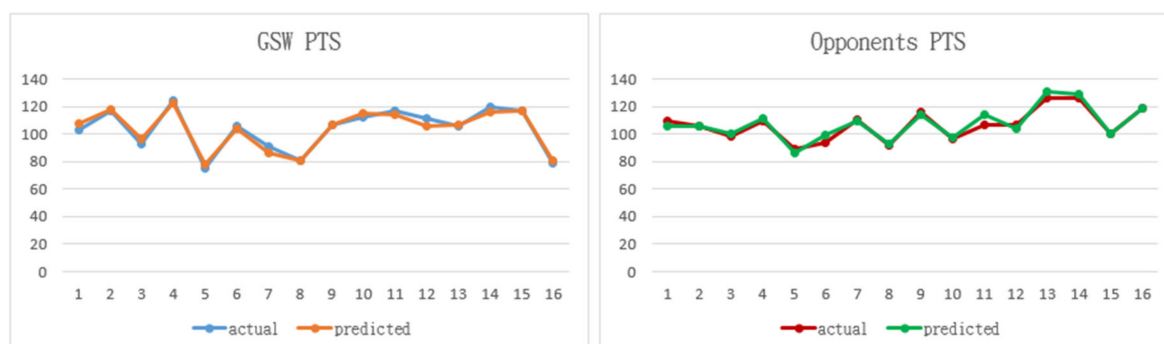


Figure 5. GSW actual and predicted score (left) vs. opponent’s actual and predicted score (right).

Except for the prediction errors for the 67th and 77th fields, all predictions were accurate, and the prediction accuracy was 87.5% (Table 5). In addition, the actual scores for some games were not significantly different from the corresponding predicted scores. For some games, the actual scores were

accurately predicted; for example, the predictions for the GSW in game 74, for both teams in game 81, and for the opponents in game 82 were accurate. Therefore, the regression tree model can effectively predict team scores.

**Table 5.** Final prediction of test set.

Season Game	Data	GSW PTS		Opponents	Opponents PTS		GSW WIN/LOSE	
		Actual	Predicted		Actual	Predicted	Actual	Predicted
67	2018/3/11	103	108	MIN	109	106	<b>LOSE</b>	<b>WIN</b>
68	2018/3/14	117	118	LAL	106	106	WIN	WIN
69	2018/3/16	93	96	SAC	98	100	LOSE	LOSE
70	2018/3/17	124	122	PHX	109	111	WIN	WIN
71	2018/3/19	75	78	SAS	89	86	LOSE	LOSE
72	2018/3/23	106	104	ATL	94	99	WIN	WIN
73	2018/3/25	91	86	UTH	110	109	LOSE	LOSE
74	2018/3/27	81	81	IND	92	93	LOSE	LOSE
75	2018/3/29	107	107	MIL	116	114	LOSE	LOSE
76	2018/3/31	112	115	SAC	96	97	WIN	WIN
77	2018/4/1	117	114	PHX	107	114	<b>WIN</b>	<b>TIE</b>
78	2018/4/3	111	106	OCT	107	104	WIN	WIN
79	2018/4/5	106	107	IND	126	131	LOSE	LOSE
80	2018/4/7	120	116	NOP	126	129	LOSE	LOSE
81	2018/4/8	117	117	PHX	100	100	WIN	WIN
82	2018/4/10	79	81	UTH	119	119	LOSE	LOSE

Accuracy: 87.5%

Note: Bold font represents that the predicted game result is different from the actual game result.

## 5. Discussion

Table 6 provides results from relevant studies. Several scholars have devoted attention to predicting game wins or losses. The use of machine learning models in earlier studies, to predict competition results, and the development of models based on other principles in recent years have engendered an increase in the accuracy of predictions. Miljkovic et al. [35] used four machine learning algorithms to predict the competition results and reported that the naive Bayes classifier had the best prediction accuracy rate (67%). Moreover, Cao [18] used four machine learning methods for prediction, including a naive Bayes classifier, and revealed that the logistic regression model achieved higher prediction accuracy than did the naive Bayes classifier. Cheng et al. [36] developed an NBAME model based on the principle of maximum entropy, to predict the outcome of games. They compared the performance of the NBAME model with traditional machine learning classifiers and reported that the NBAME model achieved a higher prediction accuracy rate (74.4%). To solve the shortcomings of SVMs that lack rule generation, Pai et al. [37] and Kaur et al. [38] used SVMs to combine decision rules and fuzzy rules, respectively, to develop new predictive match outcome models. The results demonstrated that the models achieved higher accuracy than did the conventional SVMs. Linear regression was used in another study [22], and the accuracy achieved was 47%. Although it is much lower than our result from our linear regression model, it is important to know that the proposed methodology is not comparable with this study, due to the use of a different database.

Due to the uncertainty of game results, linear regression cannot be used to generate a regression equation that illustrates the linear relationship between variables and scores. The M5P regression tree algorithm used in this study could establish multiple regression models based on the distribution of data, and the prediction accuracy was determined to be higher than those of the linear regression model and support vector regression model. The present study differs from other studies in that it applied a regression tree model to predict the scores of players in two opposing teams for each match and then summed and compared the scores, to obtain the game results of the team of interest.

**Table 6.** Comparison with results in the relevant literature.

Author, Year	Features	Method	Accuracy
Miljković et al., 2010 [35]	FG, FGA, FG%, 3P, 3PA, 3PA%, FT, FTA, FT%, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, PTS, W, L, Pct, Homewon, Homelost, Roadwon, Roadlost, Divwon, Divlost, confwon, conflost, streak, L10won, L10lost	<b>Naive Bayes</b> Decision tree KNN SVM	67.00%
Cao, 2012 [18]	G, Opp, MP, FG, FGA, 3P, 3PA, FT, FTA, ORB, TRB, AST, STL, BLK, TOV, PF, PTS, PER, TS%, eFG%, ORB%, DRB%, TRB%, AST%, STL%, BLK%, TOV%, USG%, Ortg, DRtg, OWS, DWS, WS, WS/48	<b>Logistic regression</b> ANN SVM Naive Bayes	69.67%
Wheeler, 2012 [22]	Pace, OPTS, OFG%, OTOR, DRB%, O%Rim, O%Short, OXeFG%, OeFG%, TRB%, MP, FGA, FTA, eFG%, TS%, ORtg	Linear regression	47.00%
Cheng et al., 2016 [36]	FG, FGA, 3P, 3PA, FT, FTA, ORB, DRB, AST, STL, BLK, TOV, PF, PTS	NBAME model (Using the principle of maximum entropy)	74.40%
Pai et al., 2016 [37]	2P%, 3P%, FT, DRB, TRB, STL, AST	HSVMDT (SVM + decision rules)	85.25%
Kaur and Jain, 2017 [38]	FG, FGA, FG%, 3P, 3PA%, FT, FT%, DRB, TRB, AST, TOV, PF, PTS, TS%, eFG%, ORB%, TRB%, BLK%, TOV%, ORtg, DRtg	HFSVM model (SVM + fuzzy rules)	88.26%
This study	GS, MP(s), FGA, FG%, 3PA, 3P%, FTA, FT%, ORB, DRB, TRB, AST, STL, BLK, TOV, PF	<b>Regression tree (M5P)</b>	87.50%

Note: The accuracy listed in this table was the highest one in each related study from the method in the bold font.

## 6. Conclusions

We conducted this study to develop regression tree and linear regression models by using data from two competing teams in a single NBA season. We predicted the scores of each player on the two teams and summed and compared the predicted scores of the two competing teams; thus, the win or loss results of the team of interest could be obtained. The results reveal that the regression tree model could predict player scores more accurately when compared with the linear regression model.

Any game is a complex system. The model proposed in this study yields favorable results for the prediction of the outcome of NBA games. This model can thus provide valuable prediction information for NBA team leaders and players. The limitation and future study of our method includes the following:

- (1) The procedures to determine the data rules and to obtain the corresponding equations for obtaining the prediction scores for each team were manual and must be debugged to avoid errors, which was time-consuming.
- (2) Other factors may not have been considered in this study. For example, if a team's key player does not play due to injury, the team may score lower in the relevant match. Factors that have been overlooked in the present study can be looked at by future studies, to determine whether the inclusion of such factors can further improve predictive accuracy.
- (3) The proposed models were established and tested for GSW and its opponents. More teams should be tested to evaluate the generality of the proposed model.
- (4) The dataset included GSW player match data for the 2017–2018 season only; we encourage researchers to analyze a larger dataset in the future.
- (5) The proposed models were established and tested for NBA games only. It can be used in other ball leagues; nevertheless, we do not guarantee that its predictive accuracy can be transferred to other ball leagues. Application of the proposed models on other sport events needs to be verified.

**Author Contributions:** M.-L.H. and Y.-J.L. conceived and planned the experiments; Y.-J.L. carried out the experiment; M.-L.H. and Y.-J.L. wrote the paper; M.-L.H. directed the project and verified the experimental results. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix A**

Figure A1 presents the comparison of the prediction performance of two regression models of 11 opponents of GSW.

Regression model	RMSE	Build time (seconds)
M5P	0.8908	0.06
Linear regression	2.6572	0.00

**MIN**

Regression model	RMSE	Build time (seconds)
M5P	1.0098	0.20
Linear regression	2.5208	0.01

**PHX**

Regression model	RMSE	Build time (seconds)
M5P	1.1683	0.05
Linear regression	2.6824	0.00

**UTH**

Regression model	RMSE	Build time (seconds)
M5P	1.1089	0.08
Linear regression	2.6479	0.00

**OCT**

Regression model	RMSE	Build time (seconds)
M5P	0.9464	0.05
Linear regression	2.6861	0.00

**LAL**

Regression model	RMSE	Build time (seconds)
M5P	0.8145	0.07
Linear regression	2.0288	0.00

**SAS**

Regression model	RMSE	Build time (seconds)
M5P	0.9370	0.04
Linear regression	2.2257	0.00

**IND**

Regression model	RMSE	Build time (seconds)
M5P	0.8539	0.03
Linear regression	2.5480	0.00

**NOP**

Regression model	RMSE	Build time (seconds)
M5P	0.7700	0.07
Linear regression	1.1684	0.00

**SAC**

Regression model	RMSE	Build time (seconds)
M5P	0.8225	0.04
Linear regression	2.0131	0.02

**ATL**

Regression model	RMSE	Build time (seconds)
M5P	0.6832	0.20
Linear regression	1.8760	0.01

**MIL**

Figure A1. Comparison of the prediction performance of two regression models of 11 opponents of GSW.

**Appendix B**

Figures A2 and A3 show the regression tree diagram of 11 opponents of GSW.

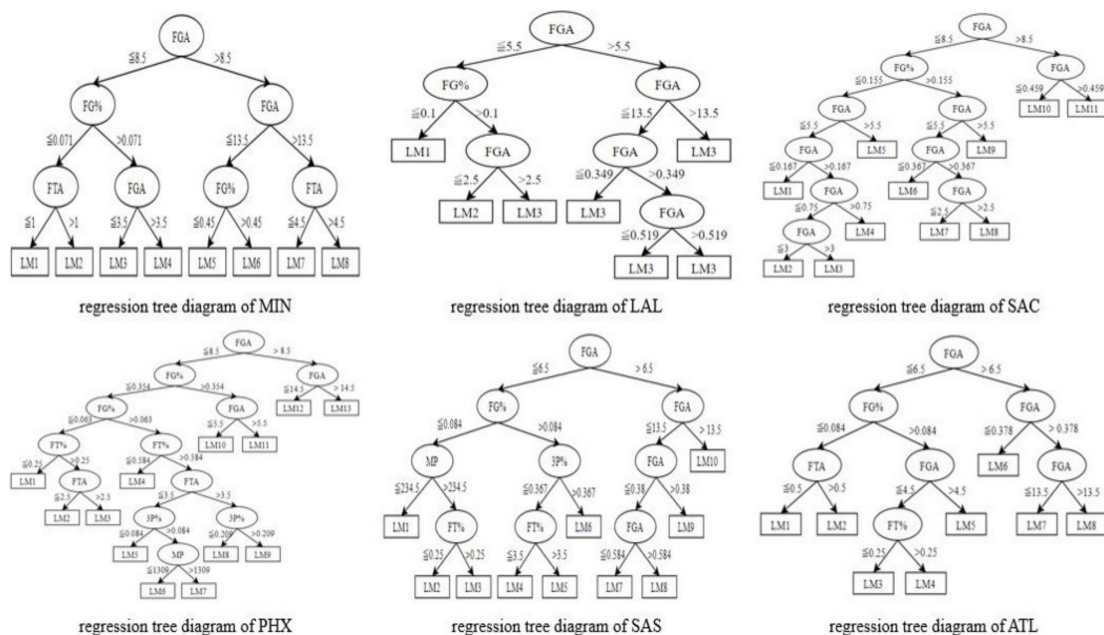


Figure A2. Regression tree diagram of GSW opponents (A).

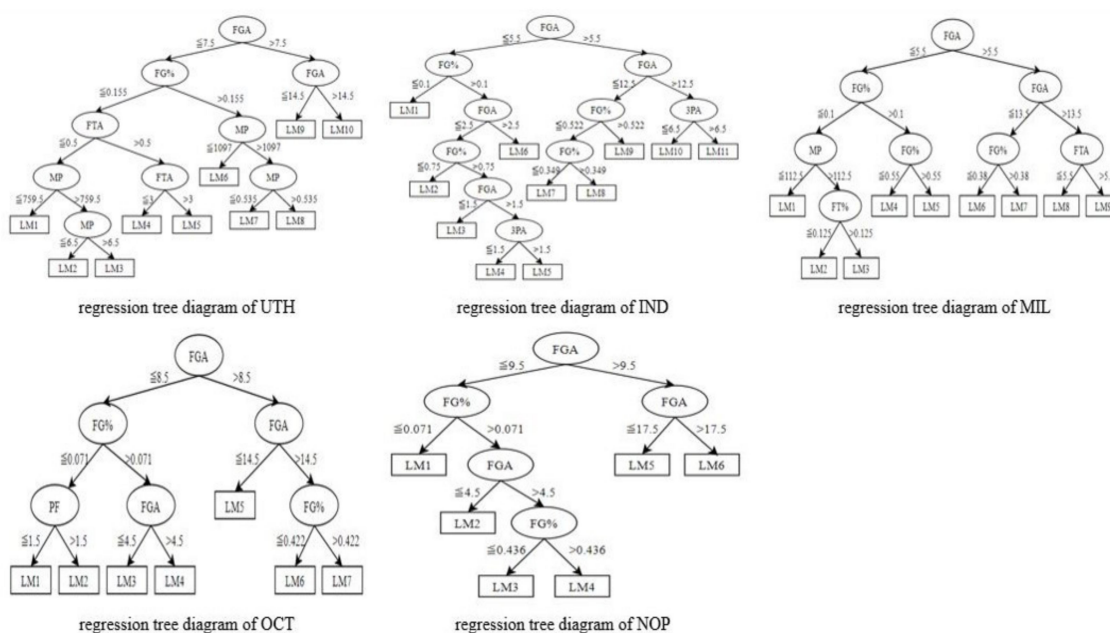


Figure A3. Regression tree diagram of GSW opponents (B).

Appendix C

Tables A1–A11 present the regression equations of several regression trees of 11 GSW opponents.

Table A1. Eight regression equations of the regression tree model of MIN.

Rule	Regression Equation
LM1	$PTS = 0.1136 * GS - 0.0001 * MP + 0.1766 * FGA + 1.0854 * FG\% + 0.5169 * 3P\% + 0.2683 * FTA + 0.2759 * FT\% + 0.0278 * DRB - 0.0262 * AST - 0.0096 * STL - 0.058 * BLK - 0.3728$
LM2	$PTS = 0.1136 * GS - 0.0001 * MP + 0.1766 * FGA + 1.0854 * FG\% + 0.5169 * 3P\% + 0.8043 * FTA + 1.2538 * FT\% + 0.0278 * DRB - 0.0262 * AST - 0.0096 * STL - 0.058 * BLK - 1.2327$
LM3	$PTS = 0.0641 * GS - 0.0001 * MP + 1.3903 * FGA + 5.1844 * FG\% + 0.1773 * 3PA + 1.2693 * 3P\% + 0.2176 * FTA + 1.6953 * FT\% + 0.0386 * DRB - 0.0281 * AST - 0.0096 * STL - 0.0723 * BLK - 0.0305 * TOV - 3.8398$
LM4	$PTS = 0.0641 * GS - 0.0001 * MP + 0.9832 * FGA + 10.0824 * FG\% + 0.1815 * 3PA + 1.7658 * 3P\% + 0.7227 * FTA + 0.7961 * FT\% + 0.1687 * ORB + 0.0299 * DRB - 0.0211 * AST - 0.0096 * STL - 0.0523 * BLK - 0.1202 * TOV - 5.3827$
LM5	$PTS = 0.0231 * GS - 0.0002 * MP + 0.8761 * FGA + 22.6725 * FG\% + 0.2429 * 3PA + 2.3976 * 3P\% + 0.6056 * FTA + 1.36 * FT\% - 0.0084 * ORB + 0.017 * DRB + 0.0226 * AST - 0.0093 * STL - 0.0145 * BLK + 0.0078 * PF - 10.3598$
LM6	$PTS = 0.0231 * GS - 0.0006 * MP + 1.1301 * FGA + 21.8965 * FG\% + 0.2616 * 3PA + 1.9132 * 3P\% + 0.7687 * FTA + 0.8672 * FT\% - 0.0084 * ORB + 0.017 * DRB + 0.0757 * AST - 0.0093 * STL - 0.0145 * BLK + 0.0078 * PF - 12.1217$
LM7	$PTS = 0.0231 * GS - 0.0002 * MP + 0.85 * FGA + 30.7728 * FG\% + 0.4108 * 3PA + 2.1747 * 3P\% + 0.2773 * FTA + 2.3993 * FT\% - 0.0114 * ORB + 0.0194 * DRB + 0.0342 * AST - 0.0093 * STL - 0.0717 * BLK + 0.0424 * PF - 14.1381$
LM8	$PTS = 0.0231 * GS - 0.0002 * MP + 0.9548 * FGA + 32.9434 * FG\% + 0.3487 * 3PA + 2.3784 * 3P\% + 0.8931 * FTA + 4.6337 * FT\% - 0.0114 * ORB + 0.0194 * DRB + 0.0371 * AST - 0.0093 * STL - 0.079 * BLK + 0.0468 * PF - 20.7739$

**Table A2.** Seven regression equations of the regression tree model of LAL.

Rule	Regression Equation
LM1	$PTS = -0.041 * GS + 0.1584 * FGA + 0.9103 * FG\% - 0.0292 * 3PA + 0.5522 * 3P\% + 0.1344 * FTA + 1.9649 * FT\% - 0.0137 * DRB + 0.0263 * TRB - 0.0058 * AST - 0.3066$
LM2	$PTS = -0.1506 * GS + 0.0001 * MP + 0.8048 * FGA + 2.1766 * FG\% - 0.0222 * 3PA + 1.782 * 3P\% + 0.2419 * FTA + 0.4397 * FT\% + 0.1166 * DRB + 0.0199 * TRB - 0.0058 * AST + 0.0508 * STL - 0.1423 * TOV - 0.8115$
LM3	$PTS = -0.0976 * GS + 0.9177 * FGA + 6.8786 * FG\% + 0.0752 * 3PA + 1.7277 * 3P\% + 0.5734 * FTA + 0.9142 * FT\% - 0.0088 * DRB + 0.0199 * TRB - 0.0058 * AST + 0.1641 * STL - 3.534$
LM4	$PTS = 0.0332 * GS + 0.6652 * FGA + 17.0546 * FG\% + 0.1436 * 3PA + 2.9707 * 3P\% + 0.5066 * FTA + 0.968 * FT\% + 0.0039 * DRB - 0.0168 * AST - 0.0097 * STL - 0.0193 * BLK - 5.9762$
LM5	$PTS = 0.0518 * GS + 0.8974 * FGA + 20.9579 * FG\% + 0.2801 * 3PA + 2.3938 * 3P\% + 0.6731 * FTA + 1.2043 * FT\% + 0.0039 * DRB - 0.0094 * AST - 0.0097 * STL - 0.0287 * BLK - 0.0092 * PF - 10.2445$
LM6	$PTS = 0.0527 * GS + 1.125 * FGA + 15.3185 * FG\% + 0.3924 * 3PA + 1.7757 * 3P\% + 0.6069 * FTA + 1.4276 * FT\% + 0.0039 * DRB - 0.0094 * AST - 0.0097 * STL - 0.1476 * BLK - 0.0093 * PF - 9.355$
LM7	$PTS = -0.0238 * GS + 0.8634 * FGA + 33.1303 * FG\% + 0.3647 * 3PA + 2.4466 * 3P\% + 0.6734 * FTA + 1.9331 * FT\% + 0.0039 * DRB - 0.0033 * AST - 0.025 * STL - 16.0197$

**Table A3.** Eleven regression equations of the regression tree model of SAC.

Rule	Regression Equation
LM1	$PTS = -0.0129 * GS + 0.2146 * FGA + 1.5615 * FG\% + 0.0132 * 3PA + 1.6046 * 3P\% + 0.1703 * FTA + 0.8129 * FT\% + 0.0034 * TRB - 0.0036 * AST - 0.0062 * STL - 0.0228 * PF - 0.7113$
LM2	$PTS = -0.0129 * GS + 0.2146 * FGA + 1.5615 * FG\% + 0.0132 * 3PA + 1.6046 * 3P\% + 0.2985 * FTA + 1.4463 * FT\% + 0.0034 * TRB - 0.0036 * AST - 0.0062 * STL - 0.0228 * PF - 0.7983$
LM3	$PTS = -0.0129 * GS + 0.2146 * FGA + 1.5615 * FG\% + 0.0132 * 3PA + 1.6046 * 3P\% + 0.3021 * FTA + 1.4463 * FT\% + 0.0034 * TRB - 0.0036 * AST - 0.0062 * STL - 0.0228 * PF - 0.7939$
LM4	$PTS = -0.0129 * GS + 0.2146 * FGA + 1.5615 * FG\% + 0.0132 * 3PA + 1.6046 * 3P\% + 0.278 * FTA + 1.4463 * FT\% + 0.0034 * TRB - 0.0036 * AST - 0.0062 * STL - 0.0228 * PF - 0.6649$
LM5	$PTS = -0.0129 * GS + 0.2609 * FGA + 1.5615 * FG\% + 0.0132 * 3PA + 5.3769 * 3P\% + 0.1705 * FTA + 2.0271 * FT\% + 0.0034 * TRB - 0.0036 * AST - 0.0062 * STL - 0.2509 * PF + 0.0913$
LM6	$PTS = -0.0129 * GS + 0.5289 * FGA + 3.4292 * FG\% + 0.0257 * 3PA + 2.3564 * 3P\% + 0.2095 * FTA + 1.5175 * FT\% - 0.0371 * ORB + 0.0034 * TRB - 0.0036 * AST - 0.0189 * STL + 0.0472 * BLK - 1.0459$
LM7	$PTS = -0.0129 * GS + 1.3501 * FGA + 5.7764 * FG\% + 0.0831 * 3PA + 0.8934 * 3P\% + 0.5419 * FTA + 1.0467 * FT\% - 0.016 * ORB + 0.0034 * TRB - 0.0036 * AST - 0.0189 * STL + 0.0673 * BLK - 4.0174$
LM8	$PTS = -0.0129 * GS + 1.2549 * FGA + 8.1146 * FG\% + 0.2546 * 3PA + 1.3603 * 3P\% + 0.4042 * FTA + 1.3464 * FT\% - 0.016 * ORB + 0.0367 * DRB + 0.0034 * TRB - 0.0036 * AST - 0.1139 * STL + 0.1569 * BLK - 5.5553$
LM9	$PTS = -0.0129 * GS + 0.8811 * FGA + 12.7822 * FG\% + 0.2248 * 3PA + 1.92 * 3P\% + 0.5038 * FTA + 1.066 * FT\% + 0.0431 * DRB + 0.0034 * TRB - 0.0036 * AST - 0.0196 * STL - 6.0382$
LM10	$PTS = -0.0177 * GS + 0.7554 * FGA + 21.0721 * FG\% + 0.3239 * 3PA + 2.1617 * 3P\% + 0.5608 * FTA + 1.2428 * FT\% + 0.0124 * ORB + 0.0046 * TRB - 0.0049 * AST - 0.0271 * STL - 8.648$
LM11	$PTS = -0.0177 * GS + 1.0667 * FGA + 23.761 * FG\% + 0.519 * 3PA + 1.9583 * 3P\% + 0.6918 * FTA + 1.0464 * FT\% + 0.0106 * ORB - 0.0938 * DRB + 0.087 * TRB - 0.0049 * AST - 0.1599 * STL - 13.9301$

**Table A4.** Ten regression equations of the regression tree model of SAS.

Rule	Regression Equation
LM1	$PTS = 0.1117 * FGA + 0.9506 * FG\% + 0.0409 * 3PA + 0.4507 * 3P\% + 0.2321 * FTA + 0.4317 * FT\% - 0.0134 * AST - 0.0288 * TOV - 0.2691$
LM2	$PTS = 0.1117 * FGA + 0.9506 * FG\% + 0.0409 * 3PA + 0.4507 * 3P\% + 0.2924 * FTA + 0.4856 * FT\% - 0.0134 * AST - 0.0288 * TOV - 0.2767$
LM3	$PTS = 0.1117 * FGA + 0.9506 * FG\% + 0.0409 * 3PA + 0.4507 * 3P\% + 0.6312 * FTA + 1.2838 * FT\% - 0.0134 * AST - 0.0288 * TOV - 0.7792$
LM4	$PTS = 0.0002 * MP + 1.1678 * FGA + 5.1841 * FG\% + 0.052 * 3PA + 0.5047 * 3P\% + 0.247 * FTA + 1.4588 * FT\% - 0.0079 * AST - 0.0279 * BLK - 0.0431 * TOV + 0.0142 * PF - 3.4654$
LM5	$PTS = 0.0007 * MP + 0.8117 * FGA + 8.8607 * FG\% + 0.052 * 3PA + 0.5047 * 3P\% + 0.5133 * FTA + 1.0659 * FT\% - 0.0079 * AST - 0.0279 * BLK - 0.035 * TOV + 0.0142 * PF - 4.4419$
LM6	$PTS = 0.0008 * MP + 1.024 * FGA + 8.438 * FG\% + 0.4651 * 3PA + 0.6746 * 3P\% + 0.4815 * FTA + 1.244 * FT\% - 0.0079 * AST - 0.0553 * BLK - 0.0199 * TOV + 0.0281 * PF - 5.5065$
LM7	$PTS = 0.0642 * GS + 0.6585 * FGA + 17.1977 * FG\% + 0.123 * 3PA + 2.8478 * 3P\% + 0.5512 * FTA + 0.3376 * FT\% - 0.0062 * TOV - 5.7993$
LM8	$PTS = -0.1892 * GS + 0.6895 * FGA + 19.3371 * FG\% + 0.1374 * 3PA + 1.1288 * 3P\% + 0.9205 * FTA + 3.8532 * FT\% - 0.0062 * TOV - 10.0091$
LM9	$PTS = 0.9738 * FGA + 18.7541 * FG\% + 0.3873 * 3PA + 1.8752 * 3P\% + 0.6744 * FTA + 0.9939 * FT\% - 0.1228 * TOV + 0.0941 * PF - 9.9508$
LM10	$PTS = 0.9505 * FGA + 33.2296 * FG\% + 0.4135 * 3PA + 0.5176 * 3P\% + 0.6855 * FTA + 1.3068 * FT\% - 0.0062 * TOV - 16.1274$

**Table A5.** Thirteen regression equations of the regression tree model of PHX.

Rule	Regression Equation
LM1	$PTS = 0.1417 * FGA + 2.3515 * FG\% + 0.0438 * 3PA + 0.5717 * 3P\% + 0.3088 * FTA + 0.4956 * FT\% + 0.0123 * AST - 0.0114 * STL - 0.0186 * PF - 0.5119$
LM2	$PTS = 0.1417 * FGA + 2.3515 * FG\% + 0.0438 * 3PA + 0.5717 * 3P\% + 0.5335 * FTA + 1.5342 * FT\% + 0.0123 * AST - 0.0114 * STL - 0.0186 * PF - 0.9602$
LM3	$PTS = 0.1417 * FGA + 2.3515 * FG\% + 0.0438 * 3PA + 0.5717 * 3P\% + 0.7157 * FTA + 1.8756 * FT\% + 0.0123 * AST - 0.0114 * STL - 0.0186 * PF - 1.6306$
LM4	$PTS = 0.544 * FGA + 10.6236 * FG\% + 0.2015 * 3PA + 1.5694 * 3P\% + 0.5165 * FTA + 0.3576 * FT\% + 0.0086 * AST - 0.0114 * STL - 0.0186 * PF - 3.1167$
LM5	$PTS = 0.0007 * MP + 0.2844 * FGA + 5.1577 * FG\% + 0.1501 * 3PA + 1.7466 * 3P\% + 0.505 * FTA + 0.4932 * FT\% + 0.0269 * ORB + 0.0086 * AST - 0.0114 * STL - 0.0186 * PF - 0.6677$
LM6	$PTS = 0.0009 * MP + 0.2844 * FGA + 5.1577 * FG\% + 0.1501 * 3PA + 1.7207 * 3P\% + 0.505 * FTA + 0.4932 * FT\% + 0.0086 * AST - 0.0114 * STL - 0.0186 * PF - 0.7115$
LM7	$PTS = 0.0009 * MP + 0.2844 * FGA + 5.1577 * FG\% + 0.1501 * 3PA + 1.7207 * 3P\% + 0.505 * FTA + 0.4932 * FT\% + 0.0086 * AST - 0.0114 * STL - 0.0186 * PF - 0.6864$
LM8	$PTS = 0.0004 * MP + 0.2844 * FGA + 5.1577 * FG\% + 0.1638 * 3PA + 2.3301 * 3P\% + 0.5467 * FTA + 0.4932 * FT\% + 0.0269 * AST - 0.0114 * STL - 0.0186 * PF + 0.0274$
LM9	$PTS = 0.0004 * MP + 0.2844 * FGA + 5.1577 * FG\% + 0.1638 * 3PA + 2.5019 * 3P\% + 0.5467 * FTA + 0.4932 * FT\% + 0.0086 * AST - 0.0114 * STL - 0.0186 * PF + 0.1507$
LM10	$PTS = -0.3212 * GS + 0.0005 * MP + 1.2472 * FGA + 7.1835 * FG\% + 0.3778 * 3PA + 1.2633 * 3P\% + 0.569 * FTA + 1.0027 * FT\% - 0.01 * AST - 0.0114 * STL + 0.022 * TOV - 0.0175 * PF - 5.3952$
LM11	$PTS = -0.4592 * GS + 0.0007 * MP + 0.9651 * FGA + 12.8293 * FG\% + 0.3184 * 3PA + 1.8388 * 3P\% + 0.594 * FTA + 1.3671 * FT\% - 0.0122 * AST - 0.0114 * STL + 0.0267 * TOV - 0.0175 * PF - 7.5166$
LM12	$PTS = 0.0004 * MP + 0.8434 * FGA + 23.068 * FG\% + 0.3742 * 3PA + 1.8547 * 3P\% + 0.6369 * FTA + 1.2095 * FT\% - 0.0187 * STL - 0.0246 * PF - 11.2644$
LM13	$PTS = 0.9557 * FGA + 35.1348 * FG\% + 0.4416 * 3PA + 2.0119 * 3P\% + 0.7654 * FTA + 1.7094 * FT\% - 0.0187 * STL - 0.0293 * PF - 18.4038$



**Table A6.** Eight regression equations of the regression tree model of ATL.

Rule	Regression Equation
LM1	$PTS = -0.0001 * MP + 0.1764 * FGA + 1.2818 * FG\% + 0.0302 * 3PA + 0.3982 * 3P\% + 0.2116 * FTA + 0.2799 * FT\% - 0.0063 * ORB - 0.0039 * AST + 0.0066 * TOV - 0.4257$
LM2	$PTS = -0.0001 * MP + 0.1764 * FGA + 1.2818 * FG\% + 0.0302 * 3PA + 0.3982 * 3P\% + 0.6367 * FTA + 1.3491 * FT\% - 0.0063 * ORB - 0.0039 * AST + 0.0066 * TOV - 1.08$
LM3	$PTS = 0.0004 * MP + 1.2723 * FGA + 5.901 * FG\% + 0.0727 * 3PA + 1.4136 * 3P\% + 0.1406 * FTA + 0.3975 * FT\% - 0.0063 * ORB - 0.0101 * DRB - 0.0039 * AST + 0.0066 * TOV - 4.2519$
LM4	$PTS = 0.0002 * MP + 1.3525 * FGA + 5.4025 * FG\% + 0.2633 * 3PA + 1.8194 * 3P\% + 0.1406 * FTA + 0.5444 * FT\% - 0.0063 * ORB - 0.1234 * DRB - 0.0039 * AST + 0.1613 * TOV - 3.1576$
LM5	$PTS = 0.2564 * FGA + 10.7834 * FG\% + 0.3757 * 3PA + 1.4551 * 3P\% + 0.576 * FTA + 1.2406 * FT\% - 0.0063 * ORB - 0.0039 * AST + 0.0066 * TOV - 2.0274$
LM6	$PTS = -0.0558 * GS + 0.6383 * FGA + 19.7827 * FG\% + 0.268 * 3PA + 2.5233 * 3P\% + 0.7214 * FTA + 0.6825 * FT\% - 0.0058 * ORB + 0.0199 * DRB - 0.017 * TRB - 0.0036 * AST - 0.0131 * STL + 0.1189 * TOV - 7.0797$
LM7	$PTS = -0.3244 * GS + 1.0839 * FGA + 19.3213 * FG\% + 0.4449 * 3PA + 2.1591 * 3P\% + 0.6807 * FTA + 1.0698 * FT\% - 0.0058 * ORB + 0.0231 * DRB - 0.0163 * TRB - 0.0036 * AST - 0.0079 * STL + 0.1128 * TOV - 11.7177$
LM8	$PTS = -0.1213 * GS + 0.9832 * FGA + 28.6595 * FG\% + 0.4575 * 3PA + 3.2159 * 3P\% + 0.6941 * FTA + 1.0059 * FT\% - 0.0058 * ORB + 0.1812 * DRB - 0.2023 * TRB - 0.0036 * AST - 0.0079 * STL + 0.0431 * TOV - 15.7175$

**Table A7.** Ten regression equations of the regression tree model of UTH.

Rule	Regression Equation
LM1	$PTS = 0.0879 * GS + 0.2072 * FGA + 1.2399 * FG\% + 0.0347 * 3PA + 0.5013 * 3P\% + 0.2015 * FTA + 0.3589 * FT\% + 0.0115 * DRB - 0.4674$
LM2	$PTS = -0.0195 * GS + 0.305 * FGA + 1.2399 * FG\% + 0.0347 * 3PA + 0.5013 * 3P\% + 0.2015 * FTA + 0.3589 * FT\% + 0.0115 * DRB - 0.5628$
LM3	$PTS = -0.0534 * GS + 0.3246 * FGA + 1.2399 * FG\% + 0.0347 * 3PA + 0.5013 * 3P\% + 0.2015 * FTA + 0.3589 * FT\% + 0.0115 * DRB - 0.438$
LM4	$PTS = 0.0879 * GS + 0.2789 * FGA + 1.2399 * FG\% + 0.0347 * 3PA + 0.5013 * 3P\% + 0.355 * FTA + 1.5366 * FT\% + 0.0115 * DRB - 0.736$
LM5	$PTS = 0.2084 * GS + 0.2692 * FGA + 1.2399 * FG\% + 0.0347 * 3PA + 0.5013 * 3P\% + 0.4238 * FTA + 1.3183 * FT\% + 0.0115 * DRB - 0.514$
LM6	$PTS = 0.0519 * GS + 1.076 * FGA + 6.512 * FG\% + 0.1729 * 3PA + 1.8056 * 3P\% + 0.7639 * FTA + 0.1474 * FT\% + 0.0049 * DRB - 0.0194 * STL - 4.1035$
LM7	$PTS = 0.0519 * GS + 0.9832 * FGA + 9.8844 * FG\% + 0.0892 * 3PA + 2.7334 * 3P\% + 0.6035 * FTA + 1.0813 * FT\% + 0.0132 * DRB + 0.0488 * AST - 0.0168 * STL + 0.0159 * TOV - 0.0707 * PF - 5.3049$
LM8	$PTS = 0.0519 * GS + 1.1435 * FGA + 9.7413 * FG\% + 0.4609 * 3PA + 1.436 * 3P\% + 0.7036 * FTA + 0.5026 * FT\% + 0.0148 * DRB - 0.0168 * STL + 0.0189 * TOV - 6.3336$
LM9	$PTS = 0.0552 * GS - 0.0005 * MP + 0.933 * FGA + 19.8717 * FG\% + 0.3163 * 3PA + 2.8292 * 3P\% + 0.6613 * FTA + 0.7231 * FT\% + 0.1377 * ORB + 0.0706 * AST + 0.0995 * PF - 9.7557$
LM10	$PTS = 0.0949 * GS - 0.0001 * MP + 0.8701 * FGA + 34.7682 * FG\% + 0.4849 * 3PA + 0.7035 * 3P\% + 0.8098 * FTA + 0.854 * FT\% - 16.5234$

**Table A8.** Eleven regression equations of the regression tree model of IND.

Rule	Regression Equation
LM1	$PTS = 0.0373 * GS + 0.0001 * MP + 0.1732 * FGA + 1.3719 * FG\% + 0.0186 * 3PA + 0.5554 * 3P\% + 0.4754 * FTA + 0.1408 * FT\% - 0.0137 * ORB - 0.4115$
LM2	$PTS = 0.0373 * GS - 0 * MP + 1.1834 * FGA + 4.4244 * FG\% + 0.2057 * 3PA + 0.8084 * 3P\% + 0.2278 * FTA + 1.1137 * FT\% - 0.0137 * ORB - 0.0526 * DRB + 0.0372 * TRB - 2.7073$
LM3	$PTS = 0.0373 * GS + 1.636 * FGA + 4.5897 * FG\% + 0.4005 * 3PA + 0.8084 * 3P\% + 0.2278 * FTA + 0.7968 * FT\% - 0.0137 * ORB - 0.0526 * DRB + 0.0372 * TRB - 3.3137$
LM4	$PTS = 0.0373 * GS + 1.7152 * FGA + 4.5897 * FG\% + 0.3666 * 3PA + 0.8084 * 3P\% + 0.2278 * FTA + 0.7968 * FT\% - 0.0137 * ORB - 0.0526 * DRB + 0.0372 * TRB - 3.2551$
LM5	$PTS = 0.0373 * GS + 1.7152 * FGA + 4.5897 * FG\% + 0.3666 * 3PA + 0.8084 * 3P\% + 0.2278 * FTA + 0.7968 * FT\% - 0.0137 * ORB - 0.0526 * DRB + 0.0372 * TRB - 3.2495$
LM6	$PTS = 0.0373 * GS + 0.9708 * FGA + 7.3572 * FG\% + 0.2765 * 3PA + 1.4484 * 3P\% + 0.4488 * FTA + 1.1871 * FT\% - 0.0137 * ORB - 0.026 * DRB + 0.0184 * TRB - 3.8555$
LM7	$PTS = 0.0576 * GS + 0.0004 * MP + 0.6212 * FGA + 15.8628 * FG\% + 0.2172 * 3PA + 2.6106 * 3P\% + 0.6639 * FTA + 0.7769 * FT\% - 0.0066 * ORB - 0.0034 * TRB - 6.2503$
LM8	$PTS = 0.0428 * GS - 0.0003 * MP + 0.917 * FGA + 17.8927 * FG\% + 0.3595 * 3PA + 1.7654 * 3P\% + 0.6567 * FTA + 0.9399 * FT\% - 0.0066 * ORB - 0.0034 * TRB - 8.5359$
LM9	$PTS = 0.018 * GS + 1.2138 * FGA + 18.004 * FG\% + 0.4471 * 3PA + 1.971 * 3P\% + 0.6639 * FTA + 0.6885 * FT\% - 0.0066 * ORB - 0.0398 * TRB - 11.5673$
LM10	$PTS = 0.018 * GS + 0.86 * FGA + 28.3833 * FG\% + 0.2731 * 3PA + 3.349 * 3P\% + 0.8045 * FTA + 0.0848 * FT\% - 0.0066 * ORB - 0.0084 * TRB + 0.0604 * AST - 13.2198$
LM11	$PTS = 0.018 * GS + 0.9991 * FGA + 34.5189 * FG\% + 0.1635 * 3PA + 6.3365 * 3P\% + 0.8812 * FTA + 0.0848 * FT\% - 0.0066 * ORB - 0.0084 * TRB - 18.7932$

**Table A9.** Nine regression equations of the regression tree model of MIL.

Rule	Regression Equation
LM1	$PTS = 0.1275 * FGA + 0.744 * FG\% - 0.0091 * 3PA + 0.3932 * 3P\% + 0.2197 * FTA + 0.5097 * FT\% - 0.0094 * ORB - 0.0116 * STL + 0.0017 * PF - 0.2317$
LM2	$PTS = 0.1275 * FGA + 0.744 * FG\% - 0.0091 * 3PA + 0.3932 * 3P\% + 0.1947 * FTA + 0.4477 * FT\% - 0.0094 * ORB - 0.0116 * STL + 0.0017 * PF - 0.2315$
LM3	$PTS = 0.1275 * FGA + 0.744 * FG\% - 0.0091 * 3PA + 0.3932 * 3P\% + 0.5317 * FTA + 1.5753 * FT\% - 0.0094 * ORB - 0.0116 * STL + 0.0017 * PF - 0.8865$
LM4	$PTS = 0.1211 * GS + 0.8761 * FGA + 7.1345 * FG\% + 0.1493 * 3PA + 1.5771 * 3P\% + 0.2827 * FTA + 1.2688 * FT\% - 0.0094 * ORB - 0.0116 * STL + 0.0561 * PF - 3.4599$
LM5	$PTS = 1.5315 * FGA + 6.1029 * FG\% + 0.5123 * 3PA + 0.8355 * 3P\% + 0.3973 * FTA + 1.1581 * FT\% - 0.0094 * ORB - 0.0116 * STL + 0.0222 * PF - 5.1388$
LM6	$PTS = 0.6704 * FGA + 16.1226 * FG\% + 0.0548 * 3PA + 4.5211 * 3P\% + 0.7321 * FTA + 0.4145 * FT\% + 0.0239 * ORB + 0.0612 * AST - 0.0107 * STL - 0.0074 * PF - 5.732$
LM7	$PTS = 1.0576 * FGA + 17.4133 * FG\% + 0.2653 * 3PA + 3.2639 * 3P\% + 0.7026 * FTA + 1.1675 * FT\% + 0.0023 * ORB - 0.0696 * AST - 0.1496 * STL - 0.0074 * PF - 9.8628$
LM8	$PTS = 0.9827 * FGA + 32.7185 * FG\% + 0.3273 * 3PA + 2.747 * 3P\% + 0.5761 * FTA + 1.4136 * FT\% + 0.122 * ORB + 0.0165 * DRB - 0.0086 * AST - 0.0107 * STL - 0.0074 * PF - 17.2515$
LM9	$PTS = 1.0699 * FGA + 34.8959 * FG\% + 0.3141 * 3PA + 1.5965 * 3P\% + 0.7351 * FTA + 7.6551 * FT\% - 0.0086 * ORB + 0.0149 * DRB - 0.0086 * AST - 0.0107 * STL - 0.0074 * PF - 24.755$

**Table A10.** Seven regression equations of the regression tree model of OCT.

Rule	Regression Equation
LM1	$PTS = 0.0413 * GS - 0.0001 * MP + 0.1552 * FGA + 0.8205 * FG\% + 0.3668 * 3P\% + 0.7021 * FTA + 0.2154 * FT\% - 0.0063 * DRB - 0.0047 * AST + 0.0219 * TOV - 0.261$
LM2	$PTS = 0.0413 * GS - 0.0001 * MP + 0.1552 * FGA + 0.8205 * FG\% + 0.3668 * 3P\% + 0.3342 * FTA + 0.2154 * FT\% - 0.0063 * DRB - 0.0047 * AST + 0.0219 * TOV - 0.2706$
LM3	$PTS = 0.0413 * GS - 0.0001 * MP + 1.4643 * FGA + 5.6642 * FG\% + 0.1354 * 3PA + 1.5398 * 3P\% + 0.3504 * FTA + 1.6443 * FT\% - 0.0063 * DRB - 0.0047 * AST + 0.1064 * STL + 0.0216 * TOV - 4.5824$
LM4	$PTS = 0.0413 * GS - 0.0001 * MP + 0.9679 * FGA + 11.2032 * FG\% + 0.2884 * 3PA + 1.7817 * 3P\% + 0.5292 * FTA + 1.1448 * FT\% - 0.0063 * DRB - 0.0047 * AST + 0.0105 * STL + 0.1496 * TOV - 5.9879$
LM5	$PTS = 0.0733 * GS - 0.0008 * MP + 0.8959 * FGA + 24.0439 * FG\% + 0.4825 * 3PA + 0.8487 * 3P\% + 0.7238 * FTA + 1.4474 * FT\% - 0.0343 * ORB - 0.0111 * DRB - 0.0656 * TRB - 0.0163 * AST + 0.0239 * STL - 10.3867$
LM6	$PTS = 0.0733 * GS - 0.0002 * MP + 0.6956 * FGA + 35.8693 * FG\% + 0.2857 * 3PA + 3.7184 * 3P\% + 0.7753 * FTA + 0.6452 * FT\% - 0.3115 * ORB - 0.0111 * DRB - 0.099 * AST + 0.0205 * STL - 12.4609$
LM7	$PTS = 0.0733 * GS - 0.0002 * MP + 0.9346 * FGA + 38.5068 * FG\% + 0.4206 * 3PA + 3.3224 * 3P\% + 0.8132 * FTA + 0.59 * FT\% - 0.0863 * ORB - 0.0971 * DRB - 0.0354 * AST + 0.0205 * STL - 19.0252$

**Table A11.** Six regression equations of the regression tree model of NOP.

Rule	Regression Equation
LM1	$PTS = 0.2048 * FGA + 1.25 * FG\% + 0.7276 * 3P\% + 0.4824 * FTA + 0.1768 * FT\% - 0.0656 * ORB - 0.0031 * DRB - 0.0102 * PF - 0.4649$
LM2	$PTS = -0.0339 * GS + 1.5162 * FGA + 6.3018 * FG\% + 0.0185 * 3PA + 1.7818 * 3P\% + 0.5911 * FTA + 0.7567 * FT\% - 0.0523 * ORB - 0.0031 * DRB + 0.0019 * PF - 4.9028$
LM3	$PTS = -0.2885 * GS + 0.7331 * FGA + 11.9836 * FG\% + 0.1983 * 3PA + 2.6125 * 3P\% + 0.4927 * FTA + 1.1207 * FT\% - 0.0639 * ORB - 0.0097 * DRB + 0.0252 * AST + 0.0126 * TOV - 0.0026 * PF - 4.5582$
LM4	$PTS = -0.0506 * GS + 1.1112 * FGA + 13.576 * FG\% + 0.3853 * 3PA + 2.3329 * 3P\% + 0.7029 * FTA + 0.2593 * FT\% - 0.066 * ORB - 0.0103 * DRB - 0.0247 * AST + 0.0137 * TOV - 0.0026 * PF - 8.2262$
LM5	$PTS = 0.9919 * FGA + 25.6773 * FG\% + 0.2983 * 3PA + 2.7463 * 3P\% + 0.6605 * FTA + 1.2326 * FT\% - 0.0118 * DRB - 0.1516 * STL - 0.0327 * PF - 13.4103$
LM6	$PTS = 0.0013 * MP + 0.9668 * FGA + 39.4122 * FG\% + 0.3625 * 3PA + 0.8249 * 3P\% + 0.7043 * FTA + 1.8209 * FT\% - 0.0169 * DRB - 0.0445 * PF - 22.6992$

## References

- Ruan, L.; Li, C.; Zhang, Y.; Wang, H. Soft computing model based financial aware spatiotemporal social network analysis and visualization for smart cities. *Comput. Environ. Urban Syst.* **2019**, *77*, 101268. [[CrossRef](#)]
- Dulebenets, M.A.; Abioye, O.F.; Ozguven, E.E.; Moses, R.; Boot, W.R.; Sando, T. Development of statistical models for improving efficiency of emergency evacuation in areas with vulnerable population. *Reliab. Eng. Syst. Saf.* **2019**, *182*, 233–249. [[CrossRef](#)]
- Andrée, B.P.J.; Chamorro, A.; Spencer, P.; Koomen, E.; Dogo, H. Revisiting the relation between economic growth and the environment; A global assessment of deforestation, pollution and carbon emission. *Renew. Sustain. Energy Rev.* **2019**, *114*, 109221. [[CrossRef](#)]
- Ulak, M.B.; Ozguven, E.E.; Vanli, O.A.; Dulebenets, M.A.; Spainhour, L. Multivariate random parameter Tobit modeling of crashes involving aging drivers, passengers, bicyclists, and pedestrians: Spatiotemporal variations. *Accid. Anal. Prev.* **2018**, *121*, 1–13. [[CrossRef](#)]
- Narasingam, A.; Kwon, J.S.-I. Data-driven identification of interpretable reduced-order models using sparse regression. *Comput. Chem. Eng.* **2018**, *119*, 101–111. [[CrossRef](#)]
- Nauright, J.; Zipp, S. The Complex World of Global Sport. *Sport Soc. Cult. Commerce Media Politics* **2018**, *21*, 1113–1119.

7. Allison, L. *The Global Politics of Sport: The Role of Global Institutions in Sport*; Psychology Press, Routledge: London, UK, 2005.
8. Loh, W.-Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
9. Yi, H.-S.; Lee, B.; Park, S.; Kwak, K.-C.; An, K.-G. Prediction of short-term algal bloom using the M5P model-tree and extreme learning machine. *Environ. Eng. Res.* **2019**, *24*, 404–411. [[CrossRef](#)]
10. Kumar, S.; Chowdary, D.; Venkatramaphanikumar, S.; Kishore, K. M5P model tree in predicting student performance: A case study. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; pp. 1103–1107.
11. Kohestani, V.; Hassanlourad, M.; Bazargan, L.M. Prediction the Ultimate Bearing Capacity of Shallow Foundations on the Cohesionless Soils Using M5P Model Tree. *J. Civ. Eng. Educ.* **2016**, *27*, 99–109.
12. Ozhegov, E.; Ozhegova, A. *Regression Tree Model for Analysis of Demand with Heterogeneity and Censorship*; HSE Working Papers WP BRP 174/EC/2017; National Research University Higher School of Economics: Moscow, Russia, 2017. [[CrossRef](#)]
13. Thabtah, F.; Zhang, L.; Abdelhamid, N. NBA game result prediction using feature analysis and machine learning. *Ann. Data Sci.* **2019**, *6*, 103–116. [[CrossRef](#)]
14. Valero, C.S. Predicting Win-Loss outcomes in MLB regular season games—A comparative study using data mining methods. *Int. J. Comput. Sci. Sport* **2016**, *15*, 91–112. [[CrossRef](#)]
15. Razali, N.; Mustapha, A.; Yatim, F.A.; Ab Aziz, R. Predicting football matches results using bayesian networks for English Premier League (EPL). *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *226*, 012099. [[CrossRef](#)]
16. Pathak, N.; Wadhwa, H. Applications of Modern Classification Techniques to Predict the Outcome of ODI Cricket. *Procedia Comput. Sci.* **2016**, *87*, 55–60. [[CrossRef](#)]
17. Loeffelholz, B.; Bednar, E.; Bauer, K.W. Predicting NBA games using neural networks. *J. Quant. Anal. Sports* **2009**, *5*, 1–17. [[CrossRef](#)]
18. Cao, C. Sports Data Mining Technology Used in Basketball Outcome Prediction. Bachelor’s Thesis, Technological University Dublin, Dublin, Ireland, 2012.
19. Harville, D.A. The selection or seeding of college basketball or football teams for postseason competition. *J. Am. Stat. Assoc.* **2003**, *98*, 17–27. [[CrossRef](#)]
20. Karlis, D.; Ntzoufras, I. Analysis of sports data using bivariate Poisson models. *J. R. Stat. Soc. Ser. D Stat.* **2003**, *52*, 381–393. [[CrossRef](#)]
21. Adam, A. Generalised linear model for football matches prediction. In Proceedings of the MLSA@PKDD/ECML, Riva del Garda, Italy, 19 September 2016; p. 1842.
22. Wheeler, K. Predicting NBA Player Performance. Available online: <http://cs229.stanford.edu/proj2012/Wheeler-PredictingNBAPlayerPerformance.pdf> (accessed on 7 June 2019).
23. Singh, T.; Singla, V.; Bhatia, P. Score and Winning Prediction in Cricket Through Data Mining. In Proceedings of the 2015 International Conference on Soft Computing Techniques and Implementations (ICSTI), Faridabad, India, 8–10 October 2015; pp. 60–66. [[CrossRef](#)]
24. Wiseman, O. Using Machine Learning to Predict the Winning Score of Professional Golf Events on the PGA Tour. Master’s Thesis, National College of Ireland, Dublin, Ireland, 2016.
25. Lu, J.; Chen, Y.; Zhu, Y. Prediction of future NBA games’ point difference: A statistical modeling approach. In Proceedings of the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 8–10 November 2019; pp. 252–256.
26. Revin, A.; Chimka, R.J. NBA game results versus sports gaming information. *Int. J. Perform. Anal. Sport* **2013**, *13*, 885–896. [[CrossRef](#)]
27. Basketball-Reference. Available online: <https://www.basketball-reference.com/> (accessed on 7 June 2019).
28. Casals, M.; Martinez, J. Modelling player performance in basketball through mixed models. *Int. J. Perform. Anal. Sports* **2013**, *13*, 64–82. [[CrossRef](#)]
29. Martínez, J.; Martínez, L. El uso de indicadores de desempeño normalizados para la valoración de jugadores: El caso de las estadísticas por minuto en baloncesto. *Motricidad. Eur. J. Hum. Mov.* **2010**, *24*, 39–62.
30. Quinlan, J.R. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Tasmania, 16–18 November 1992; pp. 343–348.

31. Wang, Y.; Witten, I.H. *Induction of Model Trees for Predicting Continuous Classes*; University of Waikato, Department of Computer Science: Hamilton, New Zealand, 1996.
32. Behnood, A.; Behnood, V.; Gharehveran, M.M.; Alyamac, K.E.J.C.; Materials, B. Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Constr. Bulid. Mater.* **2017**, *142*, 199–207. [[CrossRef](#)]
33. Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines*; Springer: Berlin, Germany, 2015; pp. 67–80.
34. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin, Germany, 2013.
35. Miljković, D.; Gajić, L.; Kovačević, A.; Konjović, Z. The use of data mining for basketball matches outcomes prediction. In Proceedings of the IEEE 8th International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 10–11 September 2010; pp. 309–312.
36. Cheng, G.; Zhang, Z.; Kyebambe, M.N.; Kimbugwe, N.J.E. Predicting the outcome of NBA playoffs based on the maximum entropy principle. *Entropy* **2016**, *18*, 450. [[CrossRef](#)]
37. Pai, P.-F.; ChangLiao, L.-H.; Lin, K.-P. Analyzing basketball games by a support vector machines with decision tree model. *Neural Comput. Appl.* **2016**, *28*. [[CrossRef](#)]
38. Jain, S.; Kaur, H. Machine learning approaches to predict basketball game outcome. In Proceedings of the 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), Dehradun, India, 15–16 September 2017; pp. 1–7.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).