

## Article

# Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association

Wei-Jen Chen <sup>1,†</sup>, Mao-Jhen Jhou <sup>1,†</sup>, Tian-Shyug Lee <sup>1</sup> and Chi-Jie Lu <sup>1,2,3,\*</sup> 

<sup>1</sup> Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City 242062, Taiwan; justice1207@gmail.com (W.-J.C.); aaa73160@gmail.com (M.-J.J.); 036665@mail.fju.edu.tw (T.-S.L.)

<sup>2</sup> Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City 242062, Taiwan

<sup>3</sup> Department of Information Management, Fu Jen Catholic University, New Taipei City 242062, Taiwan

\* Correspondence: 059099@mail.fju.edu.tw

† These authors contributed equally to this work.

**Abstract:** The sports market has grown rapidly over the last several decades. Sports outcomes prediction is an attractive sports analytic challenge as it provides useful information for operations in the sports market. In this study, a hybrid basketball game outcomes prediction scheme is developed for predicting the final score of the National Basketball Association (NBA) games by integrating five data mining techniques, including extreme learning machine, multivariate adaptive regression splines, k-nearest neighbors, eXtreme gradient boosting (XGBoost), and stochastic gradient boosting. Designed features are generated by merging different game-lags information from fundamental basketball statistics and used in the proposed scheme. This study collected data from all the games of the NBA 2018–2019 seasons. There are 30 teams in the NBA and each team play 82 games per season. A total of 2460 NBA game data points were collected. Empirical results illustrated that the proposed hybrid basketball game prediction scheme achieves high prediction performance and identifies suitable game-lag information and relevant game features (statistics). Our findings suggested that a two-stage XGBoost model using four pieces of game-lags information achieves the best prediction performance among all competing models. The six designed features, including averaged defensive rebounds, averaged two-point field goal percentage, averaged free throw percentage, averaged offensive rebounds, averaged assists, and averaged three-point field goal attempts, from four game-lags have a greater effect on the prediction of final scores of NBA games than other game-lags. The findings of this study provide relevant insights and guidance for other team or individual sports outcomes prediction research.

**Keywords:** sports outcomes prediction; basketball game; game score prediction; data mining; XGBoost; National Basketball Association



**Citation:** Chen, W.-J.; Jhou, M.-J.; Lee, T.-S.; Lu, C.-J. Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association. *Entropy* **2021**, *23*, 477. <https://doi.org/10.3390/e23040477>

Academic Editors: António M. Lopes and José A. Tenreiro Machado

Received: 15 March 2021

Accepted: 14 April 2021

Published: 17 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The sports market has grown rapidly over the last several decades with the development of technology, broadcasting, press, and social media. The prediction of sports outcomes is crucial in many sports markets, such as sports betting, club management and operations, and broadcast management, since precise sports outcomes prediction provides accurate betting reference, management and operations information, and increased viewer interests. For example, in sports lotteries, there will be more interest in predicting scores as the major bets focus on scores, such as spread points, handicaps, correct scores, or total over/under goals. Therefore, developing an effective sports outcomes prediction model that can achieve accurate and robust prediction results is one of the important and attractive challenges of sports analytics [1].

Sports outcomes prediction has attracted attention in different sports [1–6]. However, most of the existing studies in this regard are focused on basketball, in particular, on

the National Basketball Association (NBA) games, since the NBA is the most popular basketball league in the world [4,6–15]. Most existing research related to basketball games is aimed at predicting the win/lose (W/L) outcomes of a game [3,4,6,8–10,12,15]. Only a few studies have focused on the prediction of basketball game scores [11–15].

The W/L prediction provides only binary classification results; whereas, score prediction provides the intensity of a basketball game, i.e., more detailed information for stakeholders and relative application, such as sports betting, club management and operations, and broadcast management. In this study, we integrate data mining methods to propose a hybrid NBA game score prediction model.

Data mining, the process of automatically exploring potentially useful information in a large dataset, has been successfully used to construct effective forecast models for different applications in various fields [16]. There are few articles utilizing data mining methods for NBA game outcomes prediction [8–10,12,15]. Only the authors of [13] implemented data mining methods on NBA game score prediction. The five well-known data mining methods, including multivariate adaptive regression splines (MARS), k-nearest neighbors (KNN), extreme learning machine (ELM), eXtreme gradient boosting (XGBoost), and stochastic gradient boosting (SGB), are used in this study for building an NBA game score prediction model, as they have been widely used in various applications such as public health [17,18], finance [19,20] and civil engineering [21,22]. Moreover, the five methods are also successfully applied in the sports outcomes prediction research [6,23–26].

Regarding the five data mining methods, MARS, a powerful nonparametric regression method, can identify model relationships that are nearly additive or involve interactions with fewer variables [27]. KNN is a lazy algorithm that solves classification and regression problems by assigning weight to the contributions of the neighbors, where nearer neighbors contribute more than others [28]. ELM is a simple and efficient single-hidden-layer feedforward neural network [29]. XGBoost is a supervised learning algorithm based on a scalable end-to-end gradient tree boosting system [30]. SGB is a gradient-boosting-based algorithm that adds bagging and randomness into the tree building procedure, using full-dataset subsampling [31].

In the proposed hybrid basketball game score prediction model, we collected the most commonly used statistics of an NBA game as variables; then, relevant features derived from different game-lag information of a variable were generated. Since feature selection is useful for reducing the complexity of computation, the three embedded data mining feature selection techniques, including MARS, SGB, and XGBoost, are used to select relevant features for our hybrid model. After integrating and selecting the most important features, which are used as predictors, the MARS, KNN, ELM, SGB, and XGBoost, were implemented to generate the final models for predicting the final score of the NBA game. Finally, by comparing the prediction performances of the developed models under different game-lags and several selected important features, the best prediction model was identified, and the corresponding game-lag information and features used are considered the most suitable game-lag information and features for the NBA game score prediction.

The rest of this paper is organized as follows: Section 2 present the related literature review. Section 3 gives a brief introduction to the five algorithms used in this paper. Section 4 demonstrates the details of the proposed scheme. Results are presented in Section 5, followed by the Conclusions section.

## 2. Literature Review

Regarding the sports-market-related research, the research domain appears dynamic. Some studies paid attention to NBA policy such as players' labor rights [32,33] and the decision-making process of the NBA draft [34,35]. Some research aims to determine the factors that affect the outcomes of sports games such as home advantage [36,37] and individual players' performance [38,39].

There are many different prediction techniques have been applied in different areas, including civil engineering [40], industrial engineering [41,42], healthcare [43,44], safety [45]

and data mining [46,47]. Sports outcomes prediction has attracted attention in different sports [1–6]. For example, the author of [7] built a simple, weighted and penalized regression model using the match-up, date and final score records to predict baseball, basketball, American football and hockey outcomes. However, most of the existing studies in this regard are focused on basketball, in particular, on the National Basketball Association (NBA) games, since NBA is the most popular basketball league in the world [4,6,8–15].

Part of sports outcomes prediction aims to predict the winners and losers of specific NBA games. The author of [8] proposed a fusion model using multiple neural network methods to predict the outcomes of NBA games. In [9], the author generated a scheme based on the maximum entropy principle and k-means clustering to predict the winner of NBA playoff games. The author of [10] presented a model based on the support vector machine, merging the decision tree and correlation-based feature selection algorithm to predict the outcomes of the NBA games. The author of [12] proposed a framework that used historical data of NBA finals games to build up a machine learning (ML) method model to predict the outcomes of NBA games. The author of [15] presented a model that used both basketball statistics on players and teams. They used multivariate logistic regression analysis to simulate players' rotation and use these simulation results, i.e., the historical data of players and teams, as input data for predicting the outcomes of NBA games.

Some studies focus on the prediction of basketball game scores. The author of [11] established a regression model based on bivariate normal to investigate the relationship between an NBA team's score and the team's performance statistics. The author of [13] built a model based on regression tree, linear regression and support vector regression to predict the final score of the Golden State Warriors (an NBA team) in the 2017–2018 season. The author of [14] proposed a model based on the gamma process to predict the total points of NBA games, predicting the final total score of both teams.

### 3. Methods

#### 3.1. MARS Algorithm

MARS, an adaptable algorithm used to discover the optimal transformations and interactions of variables, has been widely used to recognize model relationships that are additive or include interactions with fewer variables. It is a nonparametric statistical algorithm developed from the divide-and-conquer concept for segmenting training-data subsets into divided groups with their regression equations. The MARS model implements its nonlinearity using separate linear regression slopes in distinct intervals of the independent variable space.

MARS consists of a series of weighted sum of the basis functions (BFs), which are splines piecewise polynomial functions, and are demonstrated in the following equation [18,27]:

$$f(x) = \alpha_0 + \sum_m^M a_m B_m(x) \quad (1)$$

where  $\alpha_0$  and  $a_m$  are constant coefficients that can be estimated using a least-squares method.  $M$  is the total number of BFs.  $B_m(x)$  represents the BFs. The hinge functions,  $\max(0, x - k)$  or  $\max(0, k - x)$ , with a knot defined at value  $t$ , are used in MARS modeling [48].

#### 3.2. ELM Algorithm

ELM, a single-hidden-layer feedforward neural network that randomly determines the input weights and systematically computes the output weights of the network [28], has a faster modeling time than the conventional feedforward network learning algorithms. It reduces usual disadvantages found in gradient-based methods, such as stopping criteria, learning rate and epochs [18].

Input weights and hidden layer biases in the ELM algorithm can be randomly generated, and the determination of output weights is as simple as finding the least-square solution for a specific linear system. Therefore, a linear system achieves its minimum norm least-square solution by  $\hat{A} = \tilde{H}$ , where  $\tilde{H}$  represents the Moore–Penrose generalized in-

verse of the original matrix  $H$ . The minimum norm least-squares solution has the smallest norm among all other solutions [18,29].

### 3.3. XGBoost Algorithm

XGBoost, one of the widely used tree-based learning methods, is a supervised ML algorithm developed from a scalable end-to-end gradient tree boosting concept [30]. Boosting is an ensemble learning technique that develops many models concurrently, with each new model aiming to improve the disadvantage of the previous model. A weak learner (ML model) is developed to be maximally correlated with the negative gradient of the loss function related to the entire scheme for each iteration in gradient boosting [18,49].

XGBoost is an application of a generalized gradient boosting decision tree that is implemented by a new distributed tree searching method that reduces tree construction time. XGBoost moderates overfitting and provides support for arbitrary adaptable loss functions by regularization term [18,50].

### 3.4. KNN Algorithm

KNN, a supervised data mining and ML method, is useful for solving classification and regression problems. Both classification and regression methods design weights based on the influence of neighbors, where the nearest neighbors have more influence than the others. The fundamental of KNN is the metric distance value, for which various metrics have been developed; the most common metric is the Euclidean distance [28].

KNN finds a group of  $k$  objects in the training set that are closest to the test object and facilitates the assignment of a label based on the prevalence of a specific class in this neighborhood [16,51]. For model-based methods, which learn from training datasets and then predict test datasets with the trained models, the KNN method reduces the training stage and performs classification tasks by computing the distance between test data points and all training data points to obtain the nearest neighbors and then proceed with the KNN classification [52].

### 3.5. SGB Algorithm

In SGB, a hybrid method that comprises boosting and bagging techniques [53,54], data are selected by random sampling at each stage of the steepest gradient algorithm-based boosting procedure. Smaller trees are developed instead of developing a full classification tree at each stage of the boosting process [55]. Optimal data fractionation is calculated by referring to a consequential process, and the residual of each fraction is determined. The next step in finding a new fraction, which is expected to reduce the variance of the residual of the data from the tress sequence, is to fit the residual tree. The results are merged to reduce the sensitivity of these methods for target datasets [56]. SGB does not require pre-select or transform predictor variables and is resistant to outliers since the steepest gradient methods concentrate on points that are similar to their correct classification [57].

### 3.6. Model Implementation

In this study, all methods were implemented in R version 3.6.2 [58]. MARS was implemented by the *earth* package version 5.3.0 [59]. The default setting of this package was set. XGBoost was implemented by the *XGBoost* package version 1.3.2.1 [60]. To estimate the best parameter set for developing effective XGBoost models, the *caret* package version 6.0-84 was used for tuning the relevant hyperparameters [61]. SGB was constructed by the *gbm* package version 2.1.8 [62]. ELM was computed by the *elmNN* package version 1.0 [63]. The default activation function used in this package is radial bias. The *caret* package version 6.0-84 was also implemented to search for the best number of hidden neurons that can generate promising ELM models [61]. KNN was implemented by the *kknm* package version 1.3.1 [64].

### 4. Proposed Basketball Game Score Prediction Scheme

In this study, the five data mining techniques described above were integrated to develop a novel scheme for predicting the final score of an NBA game. The flowchart of the proposed scheme is shown in Figure 1.

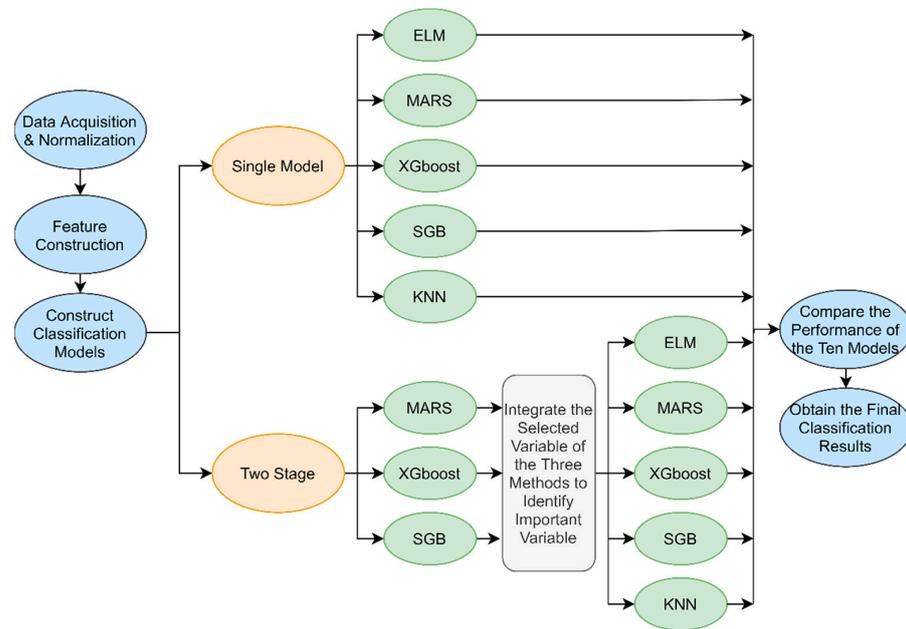


Figure 1. Flowchart of the proposed basketball game score prediction scheme.

The first step of the proposed scheme was data acquisition and normalization. We collected data from the basketball-reference website (<https://www.basketball-reference.com>, accessed on 15 March 2021) [65] for every single NBA game in the 2018–2019 season. That NBA season comprised 1230 games, and each game is categorized into home/away team statistics. Each game generates two datasets, one from the home and another from the away team. Therefore, 2460 game scores were collected and used in our research.

A total of 14 variables were collected and used in this study. One is the final score of a team; the remaining 13 are the most commonly used statistics of a game, such as the team’s defensive performance and game-related offenses [4,6,8–15]. Table 1 shows variable definitions; variable  $V_{i,t}$  is the  $i$ -th variable at the  $t$ -th game and variable  $Y_t$  is the final score at the  $t$ -th game, which is used as the target variable of this study. Since each team play 82 games in a season, variable  $V_{i,t}$  can be defined as  $1 \leq i \leq 13, 1 \leq t \leq 82$ .

Table 1. Variables description.

Variables	Definition	Description
$V_{1,t}$	2PA	2-Point Field Goal Attempts of a team in $t$ -th game
$V_{2,t}$	2P%	2-Point Field Goal Percentage of a team in $t$ -th game
$V_{3,t}$	3PA	3-Point Field Goal Attempts of a team in $t$ -th game
$V_{4,t}$	3P%	3-Point Field Goal Percentage of a team in $t$ -th game
$V_{5,t}$	FTA	Free Throw Attempts of a team in $t$ -th game
$V_{6,t}$	FT%	Free Throw Percentage of a team in $t$ -th game
$V_{7,t}$	ORB	Offensive Rebounds of a team in $t$ -th game
$V_{8,t}$	DRB	Defensive Rebounds of a team in $t$ -th game
$V_{9,t}$	AST	Assists of a team in $t$ -th game
$V_{10,t}$	STL	Steals of a team in $t$ -th game
$V_{11,t}$	BLK	Blocks of a team in $t$ -th game
$V_{12,t}$	TOV	Turnovers of a team in $t$ -th game
$V_{13,t}$	PF	Personal Fouls of a team in $t$ -th game
$Y_t$	Score	Team Score of a team in $t$ -th game

Data normalization shall be implemented before data analysis since different variables have different scales. The min-max normalization method was used to convert a value  $v$  of each variable  $V$  to  $v'$  in the range  $[0, 1]$  by calculating using the following equation:

$$V'_{i,t} = \frac{V_{i,t} - \min V_i}{\max V_i - \min V_i} \tag{2}$$

where  $\max V_i$  and  $\min V_i$  are the maximum and minimum values for the attribute  $V_i$ . Data normalization was performed to ensure that large input variable values do not influence smaller input values, thus reducing prediction errors.

The second step is the feature construction to generate input features for the data mining models based on the variables shown in Table 1. We define the game-lag of a game variable as “the  $n$ -th game before game  $t$ ”. For example, the third game-lag of game 65 is game 62. In most related research, researchers used only the game-lag information of up to six games for model construction [8,9]. To consider more complete game-lag information, the game-lag information of 1–6 games is used in this study. However, a variable value in a single game may not be sufficient for evaluating a team’s performance. Therefore, we calculate the mean value of a variable within  $l$  game-lags to evaluate a team’s performance during a specific period. Variable  $\bar{V}^l_{i,t}$  is the designed  $i$ -th predictor variable at the  $t$ -th game with  $l$  game-lags.

$$\bar{V}^l_{i,t} = \frac{\sum_{n=1}^l V'_{i,t-n}}{l}, \forall i, t, n, l \in \mathbb{N}, \tag{3}$$

where  $n$  is the  $n$ -th game-lag,  $1 \leq i \leq 13, 1 \leq l \leq 6, 1 \leq t \leq 82, n \leq l$ .

For instance, for the first variable ( $i = 1$ ), if we want to design a feature considering three game-lags’ information ( $l = 3$ ) for the game No. 10 (or 10-th game) ( $t = 10$ ) of a team, the values of the first variable in the previous three games are averaged as the designed feature. That is,  $\bar{V}^3_{1,10} = \frac{V_{1,9} + V_{1,8} + V_{1,7}}{3}$ . Therefore, using the same concept, a variable in one game can be extended to six designed features under the consideration of one game-lag to six game-lags’ information. Figure 2 shows designed feature examples for variable  $V_{i,t}$  in different game-lags.

	$V_{i,t-1}$	$V_{i,t-2}$	$V_{i,t-3}$	$V_{i,t-4}$	$V_{i,t-5}$	$V_{i,t-6}$
$l = 1$	$\bar{V}^1_{i,t}$					
$l = 2$	$\bar{V}^2_{i,t}$					
$l = 3$	$\bar{V}^3_{i,t}$					
$l = 4$	$\bar{V}^4_{i,t}$					
$l = 5$	$\bar{V}^5_{i,t}$					
$l = 6$	$\bar{V}^6_{i,t}$					

Figure 2. Example of the designed features for variable  $V_{i,t}$  in different game-lags.

This research aims to construct the prediction model using the designed features ( $\bar{V}^l_{i,t}$ ) to predict the final score of a game ( $Y_t$ ), as expressed using Equation (4):

$$Y_t = f(\bar{V}^l_{i,t}), \tag{4}$$

where  $1 \leq i \leq 13, 1 \leq l \leq 6, 7 \leq t \leq 82, \forall i, t, l \in \mathbb{N}$ .

Note that all 13 designed features ( $1 \leq i \leq 13$ ) were used with 1–6 game-lags’ information ( $1 \leq l \leq 6$ ) for each  $Y_t$ . Since we use up to six games’ information as our game-lag information, the first six games of the season are skipped ( $7 \leq t \leq 82$ ).

In the third step, we construct predictive models for predicting final scores of the NBA games considering different game-lags. The predictive models were constructed using two types of modeling processes. One, a single modeling process, and the other, a two-stage modeling process. In the single modeling process, all 13 designed features were directly used as predictors for developing ELM, MARS, XGBoost, SGB, and KNN as five single-predictive models. These were termed single ELM (S-ELM), single MARS (S-MARS), single XGBoost (S-XGBoost), single SGB (S-SGB) and single KNN (S-KNN) models.

Developing a two-stage model began with the implementation of a feature selection method, as some important basketball variables have a greater influence on predicting the outcomes of basketball games. This study used an embedded feature selection method with the implementation of MARS, XGBoost, and SGB since these methods are equipped with feature selection functions. The three algorithms generate their best subsets of features. This study uses ensemble techniques to merge the selected features subsets of the three algorithms in order to provide stable and effective feature selection result. Ensemble technique is a paradigm, where several intermediate selected features are generated and combined using combination rules to obtain a simple selection result. It can be used to avoid unstable selection results and improve the performance of feature selection [66].

For example, Table 2 illustrates the feature importance ranking generated by MARS, XGBoost and SGB algorithms under game-lag = 4 ( $l = 4$ ). Note that a feature with a rank of 1 is considered the most important, while one with a rank of 13 is considered less important than other features. The average ranking is obtained by calculating each feature by its ranking in the MARS, XGBoost and SGB. Table 2 shows the average rank of each feature. It can be observed that  $\bar{V}_{8,t}^4$  with an average rank of 1.67, is the most important feature, followed by  $\bar{V}_{2,t}^4$  with value 2.00 and  $\bar{V}_{6,t}^4$  with value 3.67.

**Table 2.** Feature rank by MARS, XGBoost and SGB methods under game-lag = 4.

Designed Feature	MARS	XGBoost	SGB	Average Rank
$\bar{V}_{1,t}^4$	6	8	10	8.00
$\bar{V}_{2,t}^4$	2	1	3	2.00
$\bar{V}_{3,t}^4$	5	6	8	6.33
$\bar{V}_{4,t}^4$	8	7	5	6.67
$\bar{V}_{5,t}^4$	7	9	7	7.67
$\bar{V}_{6,t}^4$	3	4	4	3.67
$\bar{V}_{7,t}^4$	4	5	6	5.00
$\bar{V}_{8,t}^4$	1	2	2	1.67
$\bar{V}_{9,t}^4$	13	3	1	5.67
$\bar{V}_{10,t}^4$	13	13	13	13.00
$\bar{V}_{11,t}^4$	13	10	11	11.33
$\bar{V}_{12,t}^4$	13	11	12	12.00
$\bar{V}_{13,t}^4$	9	12	9	10.00

Table 2 illustrates the importance ranking of each designed feature. To reduce the number of less important features selected, this study uses the significant predictive feature selection rule proposed by the author of [67]. Their method selects important features based on the total number of features. If the total number of features is between 10 and 75, researchers can select 40% of the features as relevant from the overall features. Therefore, since there are 13 designed features in this paper, we select 6 designed features as relevant features.

According to the ranking results in Table 2, under game-lag = 4,  $\bar{V}_{8,t}^4$ ,  $\bar{V}_{2,t}^4$ ,  $\bar{V}_{6,t}^4$ ,  $\bar{V}_{7,t}^4$ ,  $\bar{V}_{9,t}^4$  and  $\bar{V}_{3,t}^4$  are selected as the relevant features. These six relevant features served as the input variables for the ELM, MARS, XGBoost, SGB and KNN methods for predicting the final score of an NBA game. The five two-stage methods were termed two-stage ELM

(T-ELM), two-stage MARS (T-MARS), two-stage XGBoost (T-XGBoost), two-stage SGB (T-SGB), and two-stage KNN (T-KNN).

Next, we compare the performance of the models after obtaining prediction results from the five single and two-stage models under a specific game-lag. This study used the mean absolute percentage error (MAPE) as the indicator to evaluate the performance of the prediction models and determine the best game-lag selection.

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left( \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100 \quad (5)$$

where  $\hat{y}_i$  represents the actual game score of the  $i$ -th sample,  $y_i$  represents the predicted game score of the  $i$ -th sample, and  $m$  is the number of samples.

MAPE has been widely used as a performance indicator for evaluating forecasting/predicting methods [68]. When  $\text{MAPE} < 10\%$ , a model is considered to have “high accurate prediction ability”. When  $11\% < \text{MAPE} < 20\%$ , a model has a “good prediction ability”. When  $21\% < \text{MAPE} < 50\%$ , a model has a “reasonable prediction ability”. When  $\text{MAPE} > 51\%$ , a model has an “inaccurate prediction ability” [69]. A 10-fold cross-validation method is used in this study to evaluate the performances of the 10 proposed models.

In the final step of the proposed scheme, after comparing the prediction performances of the 10 models, including S-ELM, S-MARS, S-XGBoost, S-SGB, S-KNN, T-ELM, T-MARS, T-XGBoost, T-SGB and T-KNN, under different game-lags, the model with the best prediction performance is identified. Moreover, the most suitable basketball game-lag information and corresponding important basketball features are selected.

## 5. Empirical Results

In this paper, the NBA teams’ statistics in each game in the 2018–2019 season were used to verify the proposed basketball prediction scheme for predicting the final score of the NBA games. The performance of the 10 models, including S-ELM, S-MARS, S-XGBoost, S-SGB, S-KNN, T-ELM, T-MARS, T-XGBoost, T-SGB and T-KNN, can be evaluated using the proposed scheme mentioned in Section 4.

Table 3 shows the performance of the five single models, including S-ELM, S-MARS, S-XGBoost, S-SGB and S-KNN, under 1–6 game-lags. S-XGBoost obtains the best performance under game-lag = 4 with an MAPE value of 0.0842, followed by S-SGB under game-lag = 4 with an MAPE value of 0.0845, and S-MARS under game-lag = 4 with an MAPE value 0.0846.

**Table 3.** Performance of the five single models under six game-lags.

Methods	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$
S-ELM	0.1020	0.0960	0.0915	0.0870	0.0931	0.0928
S-MARS	0.0910	0.0909	0.0897	0.0846	0.0917	0.0907
S-XGBoost	0.0919	0.0907	0.0911	<b>0.0842</b>	0.0927	0.0920
S-SGB	0.0910	0.0925	0.0913	0.0845	0.0923	0.0908
S-KNN	0.0992	0.1011	0.0947	0.0873	0.0934	0.0941

Note: The bold indicates the best prediction performance.

As aforementioned, six features were selected as the important features and served as the crucial predictor variables for developing the two-stage models, including T-ELM, T-MARS, T-XGBoost, T-SGB and T-KNN. Table 4 shows the performance of the five two-stage models under six game-lags. T-XGBoost obtains the best performance under game-lag = 4 with an MAPE value of 0.0818, followed by T-SGB under game-lag = 4 with an MAPE value of 0.0829, and T-MARS under game-lag = 4 with an MAPE value of 0.0845.

**Table 4.** Performance of the five two-stage models under six game-lags.

Methods	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$
T-ELM	0.1206	0.0924	0.0951	0.0863	0.0972	0.0902
T-MARS	0.0917	0.0911	0.0912	0.0845	0.0928	0.0900
T-XGBoost	0.0918	0.0930	0.0916	<b>0.0818</b>	0.0929	0.0920
T-SGB	0.0909	0.0918	0.0912	0.0829	0.0930	0.0908
T-KNN	0.0998	0.0984	0.0973	0.0872	0.0993	0.0970

Note: The bold indicates the best prediction performance.

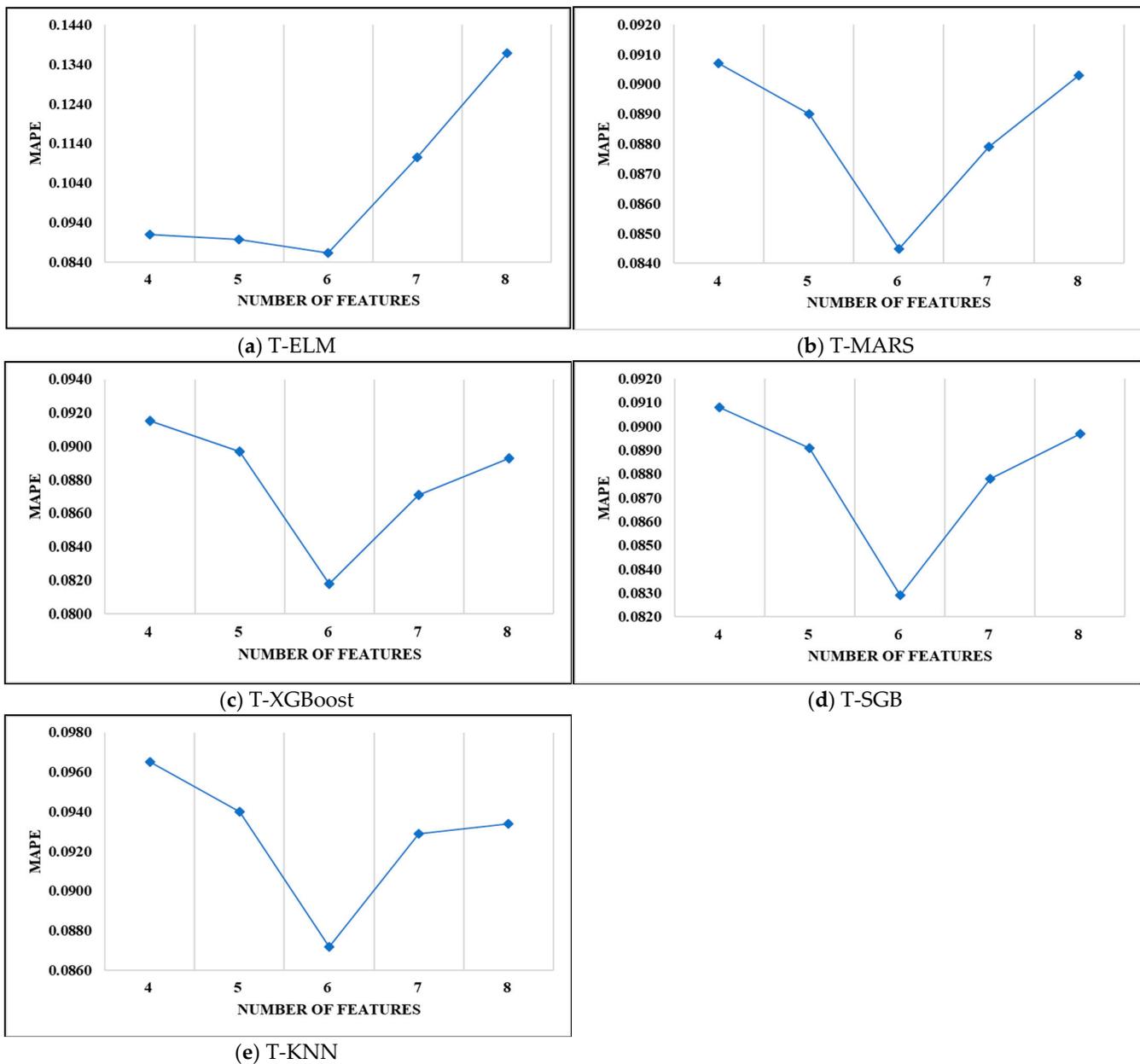
From Tables 3 and 4 it can be observed that T-XGBoost under four game-lags obtains the best performance among the models, including the single and two-stage models, under every game-lag information. It can also be seen that the 10 models obtain their lowest MAPE values, along with their best prediction performance, under game-lag = 4. Therefore, game-lag = 4 is the most suitable game-lag for NBA game score prediction.

Since the two-stage model T-XGBoost, with game-lag = 4, is the best model in this study, the six selected designed features, including  $\bar{V}_{8,t}^4$  (defensive rebounds),  $\bar{V}_{2,t}^4$  (two-point field goal percentage),  $\bar{V}_{6,t}^4$  (free throw percentage),  $\bar{V}_{7,t}^4$  (offensive rebounds),  $\bar{V}_{9,t}^4$  (assists) and  $\bar{V}_{3,t}^4$  (three-point field goal attempts), are the important features.

To evaluate the robustness of the feature selection results of the proposed scheme, we select different numbers of important features for modeling the two-stage models according to the feature ranking results mentioned in Section 4. We consider the selection of more or fewer features than the six selected features. We used 4, 5, 6, 7 and 8 features to develop the five two-stage prediction models. Figure 3 demonstrates the MAPE value obtained by each model with a different selection of features. It shows that the MAPE value gradually decreases with an increase in the number of features selected from 4–6. The converse happens if more than six features are selected. Therefore, the selection of six features as relevant features produced better performance than any other number of features selected.

Based on the finding discussed in this manuscript, it can be inferred that the proposed basketball game outcomes prediction scheme is a promising method for the final score of NBA games. This study examines the impact of different game-lag information while related studies arbitrarily select game-lag information either by 3 or 6 [8,9]. This research suggests that game-lag = 4 is a proper selection by appropriate feature designation. The suitable important features have been selected by using feature ensemble [66] and selection techniques [67] in the proposed scheme which are simple and effective methods. These selected important features are related to offensive factors since this study is focused on predicting the final score of basketball games and this finding is supported by [13].

To further validate the effectiveness of the best model, i.e., T-XGBoost with lag = 4, of the proposed basketball game score prediction scheme, linear regression, M5P regression tree and support vector regression (SVR), which are used in related research [13], are implemented and compared.



**Figure 3.** Evaluation results of the selection of different numbers of important features for modeling the two-stage models: (a) T-ELM, (b) T-MARS, (c) T-XGBoost, (d) T-SGB, (e) T-KNN.

With implementation of these three methods in the proposed scheme in Figure 1, each method will generate its own single-stage and two-stage models. The prediction performance among the S-Linear (single-stage linear regression), S-M5P (single-stage regression tree), S-SVR (single-stage SVR), T-Linear (two-stage linear regression), T-M5P (two-stage regression tree), T-SVR (two-stage SVR) and the best model T-XGBoost models under lag = 4 are compared. Alongside MAPE, we also involve root-mean-square error (RMSE) and sum of squared error (SSE) as additional performance indicators since these indicators are effectively used in sports outcomes prediction [11,13]. This study uses MAPE, RMSE and SSE as performance indicators to compare prediction performance among seven models.

The results of model comparison are shown in Table 5. As shown in Table 5, the MAPE, RMSE and SSE values of T-XGBoost are 0.0818, 11.4753 and 61,627.37, respectively. T-XGBoost still has the best prediction performance among the six competing models.

**Table 5.** Comparison of prediction performance of T-XGBoost and the six competing models.

Models (Lag = 4)	MAPE	RMSE	SSE
S-Linear	0.0897	12.7324	75,868.89
T-Linear	0.0883	12.0904	68,410.85
S-M5P	0.0922	13.0613	79,839.95
T-M5P	0.0931	12.9102	78,003.64
S-SVR	0.0914	13.0213	79,351.58
T-SVR	0.0889	12.2547	70,283.06
T-XGBoost	0.0818	11.4753	61,627.37

In order to test whether selecting game-lags information of four is significantly superior to other game-lag, the Wilcoxon signed-rank test is applied. The Wilcoxon signed-rank test is a distribution-free, non-parametric technique which determines whether two models are different by comparing the signs and ranks of prediction values. The Wilcoxon signed-rank test is one of the most popular tests in evaluating the predictive capabilities of two different models [70].

We employ the test to evaluate the prediction performance of selecting game-lag as four with other game-lag information on T-XGBoost. Table 6 shows the Z statistic and *p*-values of the two-tailed Wilcoxon signed-rank test for MAPE values between the game-lag = 4 and other game-lags. It can be observed from Table 6 that the MAPE values of the game-lag = 4 is significantly different from other game-lag. Therefore, we can conclude that selecting game-lag = 4 is significantly better than other selections of game-lag information.

To further validate the superiority of the best model, i.e., T-XGBoost with lag = 4, we test the best model with T-Linear, T-MARS, T-SVR, T-SGB, T-KNN, T-ELM and T-M5P with lag = 4 using the Wilcoxon signed-rank test. Table 7 reports the test results between the best model to each of the seven competing models. It shows that the prediction error of T-XGBoost model is significantly lower than those of its competing models. Therefore, it can be concluded that the proposed T-XGBoost model significantly outperforms the alternatives in NBA game score prediction.

**Table 6.** Wilcoxon signed-rank test between six pieces of game-lag information on the T-XGBoost model.

T-XGBoost	Lag = 1	Lag = 2	Lag = 3	Lag = 5	Lag = 6
Lag = 4	−1.017 (0.007) **	−1.044 (0.008) **	−4.284 (0.001) **	−6.115 (0.001) **	−10.859 (0.001) **

Note: The numbers in parentheses are the corresponding *p*-value; \*\* *p* < 0.05.

**Table 7.** Wilcoxon signed-rank test between T-XGBoost, T-Linear, T-MARS, T-SVR, T-SGB, T-KNN, T-ELM and T-M5P models.

Lag = 4	T-Linear	T-MARS	T-SVR	T-SGB	T-KNN	T-ELM	T-M5P
T-XGBoost	−1.239 (0.001) **	−0.994 (0.010) **	−0.997 (0.010) **	−0.989 (0.011) **	−0.885 (0.021) **	−1.377 (0.000) **	−1.043 (0.008) **

Note: The numbers in parentheses are the corresponding *p*-value; \*\* *p* < 0.05.

## 6. Discussion and Conclusions

This research proposed a hybrid data-mining-based scheme for predicting the final score of an NBA game. We design features from original basketball statistics based on game-lag information. The proposed prediction scheme used five data mining algorithms, namely, ELM, MARS, XGBoost, SGB and KNN. The prediction scheme comprises five single and five two-stage prediction models. Empirical results showed that the T-XGBoost model using game-lag = 4 achieved the best prediction performance among the 10 competing models, using 2–6 game-lags' information. The most suitable count of game-lag information for NBA game score prediction is four. The six identified important statistics (features) based on four game-lags are averaged defensive rebounds, averaged two-point field goal

percentage, averaged free throw percentage, averaged offensive rebounds, averaged assists, and averaged three-point field goal attempts. The findings of this study may be applied to the development of several applications for other teams or even individual sports.

Since the NBA data adopted in this study are limited to one season, future research should investigate the performance of the proposed basketball game score prediction scheme with more NBA seasons. Moreover, using more seasons' data to generate more stable important feature selection results could also be a future research direction. Furthermore, this study collected individual NBA team data to predict individual teams' final score in NBA games. To further predict interesting or specific games' scores or win/loss, such as particular matchups, low-scoring games, or intensity of a team's game schedule and use these specific types of dataset to improve the performance of the proposed model, as well as making modification and extension of the proposed scheme such as considering opponent teams' information as features, could be one of future research directions.

**Author Contributions:** Conceptualization, C.-J.L. and W.-J.C.; methodology, C.-J.L. and M.-J.J.; software, M.-J.J.; validation, T.-S.L.; formal analysis, C.-J.L., W.-J.C. and M.-J.J.; investigation, T.-S.L.; resources, C.-J.L. and T.-S.L.; data curation, W.-J.C.; writing—original draft preparation, C.-J.L., W.-J.C. and M.-J.J.; writing—review and editing, C.-J.L., W.-J.C. and M.-J.J.; visualization, C.-J.L., W.-J.C. and M.-J.J.; supervision, T.-S.L.; project administration, C.-J.L.; funding acquisition, C.-J.L. and T.-S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is partially supported by the Fu-Jen Catholic University under grant number A0109150 and Ministry of Science and Technology, Taiwan, under grant numbers: 109-2221-E-030-010-; 109-2622-E-030-001-CC3).

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: <https://www.basketball-reference.com>, accessed on 15 March 2021.

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

## References

- Morgulev, E.; Azar, O.H.; Lidor, R. Sports Analytics and the Big-Data Era. *Int. J. Data Sci. Anal.* **2018**, *5*, 213–222. [CrossRef]
- Musa, R.M.; Majeed, A.P.A.; Taha, Z.; Chang, S.W.; Nasir, A.F.A.; Abdullah, M.R. A Machine Learning Approach of Predicting High Potential Archers by Means of Physical Fitness Indicators. *PLoS ONE* **2019**, *14*, e0209638. [CrossRef] [PubMed]
- Zuccolotto, P.; Manisera, M.; Sandri, M. Big Data Analytics for Modeling Scoring Probability in Basketball: The Effect of Shooting under High-Pressure Conditions. *Int. J. Sports Sci. Coach.* **2018**, *13*, 569–589. [CrossRef]
- Lam, M.W. One-Match-Ahead Forecasting in Two-Team Sports with Stacked Bayesian Regressions. *J. Artif. Intell. Soft Comput. Res.* **2018**, *8*, 159–171. [CrossRef]
- Baboota, R.; Kaur, H. Predictive Analysis and Modelling Football Results using Machine Learning Approach for English Premier League. *Int. J. Forecast.* **2019**, *35*, 741–755. [CrossRef]
- Horvat, T.; Havaš, L.; Srpak, D. The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes. *Symmetry* **2020**, *12*, 431. [CrossRef]
- Dubbs, A. Statistics-Free Sports Prediction. *Model Assist. Stat. Appl.* **2018**, *13*, 173–181. [CrossRef]
- Loeffelholz, B.; Bednar, E.; Bauer, K.W. Predicting NBA Games using Neural Networks. *J. Quant. Anal. Sports* **2009**, *5*. [CrossRef]
- Cheng, G.; Zhang, Z.; Kyebambe, M.N.; Kimbugwe, N. Predicting the Outcome of NBA Playoffs Based on the Maximum Entropy Principle. *Entropy* **2016**, *18*, 450. [CrossRef]
- Pai, P.F.; ChangLiao, L.H.; Lin, K.P. Analyzing Basketball Games by A Support Vector Machines with Decision Tree Model. *Neural Comput. Appl.* **2017**, *28*, 4159–4167. [CrossRef]
- Song, K.; Zou, Q.; Shi, J. Modelling the Scores and Performance Statistics of NBA Basketball Games. *Commun. Stat. Simul. Comput.* **2018**, *49*, 2604–2616. [CrossRef]
- Thabtah, F.; Zhang, L.; Abdelhamid, N. NBA Game Result Prediction Using Feature Analysis and Machine Learning. *Ann. Data Sci.* **2019**, *6*, 103–116. [CrossRef]
- Huang, M.L.; Lin, Y.J. Regression Tree Model for Predicting Game Scores for the Golden State Warriors in the National Basketball Association. *Symmetry* **2020**, *12*, 835. [CrossRef]
- Song, K.; Gao, Y.; Shi, J. Making Real-Time Predictions for NBA Basketball Games by Combining the Historical Data and Bookmaker's Betting Line. *Phys. A Stat. Mech. Its Appl.* **2020**, *547*, 124411. [CrossRef]
- Li, Y.; Wang, L.; Li, F. A Data-Driven Prediction Approach for Sports Team Performance and Its Application to National Basketball Association. *Omega* **2021**, *98*, 102123. [CrossRef]
- Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson Education Inc.: Taipei, Taiwan, 2006; pp. 1–3.

17. Jordan, S.E.; Hovet, S.E.; Fung, I.C.H.; Liang, H.; Fu, K.W.; Tse, Z.T.H. Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data* **2019**, *4*, 6. [[CrossRef](#)]
18. Ting, W.C.; Chang, H.R.; Chang, C.C.; Lu, C.J. Developing a Novel Machine Learning-Based Classification Scheme for Predicting SPCs in Colorectal Cancer Survivors. *Appl. Sci.* **2020**, *10*, 1355. [[CrossRef](#)]
19. Nobre, J.; Neves, R.F. Combining Principal Component Analysis, Discrete Wavelet Transform and Xgboost to Trade in the Financial Markets. *Expert Syst. Appl.* **2019**, *125*, 181–194. [[CrossRef](#)]
20. Yu, H.; Ming, L.J.; Sumei, R.; Shuping, Z. A Hybrid Model for Financial Time Series Forecasting—Integration of EWT, ARIMA with the Improved ABC Optimized ELM. *IEEE Access* **2020**, *8*, 84501–84518. [[CrossRef](#)]
21. Shariati, M.; Mafipour, M.S.; Mehrabi, P.; Zandi, Y.; Dehghani, D.; Bahadori, A.; Shariati, A.; Trung, N.T.; Salih, M.N.; Poi-Ngian, S. Application of Extreme Learning Machine (ELM) and Genetic Programming (GP) to Design Steel-Concrete Composite Floor Systems at Elevated Temperatures. *Steel Compos. Struct* **2019**, *33*, 319–332.
22. Zhou, J.; Li, E.; Wang, M.; Chen, X.; Shi, X.; Jiang, L. Feasibility of Stochastic Gradient Boosting Approach for Evaluating Seismic Liquefaction Potential Based on SPT and CPT Case Histories. *J. Perform. Constr. Facil.* **2019**, *33*, 4019024. [[CrossRef](#)]
23. Chau, V.H.; Vo, A.T.; Le, B.T. A Gravitational-Double Layer Extreme Learning Machine and its Application in Powerlifting Analysis. *IEEE Access* **2019**, *7*, 143990–143998. [[CrossRef](#)]
24. Kapadia, K.; Abdel-Jaber, H.; Thabtah, F.; Hadi, W. Sport Analytics for Cricket Game Results Using Machine Learning: An Experimental Study. *Appl. Comput. Inform.* **2020**. [[CrossRef](#)]
25. Rajšp, A.; Fister, I. A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training. *Appl. Sci.* **2020**, *10*, 3013. [[CrossRef](#)]
26. Sarlis, V.; Chatziilias, V.; Tjortjis, C.; Mandalidis, D. A Data Science Approach Analysing the Impact of Injuries on Basketball Player and Team Performance. *Inf. Syst.* **2021**, *99*, 101750. [[CrossRef](#)]
27. Friedman, J.H. Multivariate Adaptive Regression Splines. *Ann. Stat.* **1991**, 1–67. [[CrossRef](#)]
28. Harrington, P. *Machine Learning in Action*; Manning Publications Co.: Shelter Island, NY, USA, 2012.
29. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
30. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
31. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. *Math. Intell.* **2005**, *27*, 83–85.
32. Yang, C.H.; Lin, H.Y. Is There Salary Discrimination by Nationality in the NBA? Foreign Talent or Foreign Market. *J. Sports Econ.* **2012**, *13*, 53–75. [[CrossRef](#)]
33. Kopkin, N. Tax Avoidance: How Income Tax Rates Affect the Labor Migration Decisions of NBA Free Agents. *J. Sports Econ.* **2012**, *13*, 571–602. [[CrossRef](#)]
34. Price, J.; Soebbing, B.P.; Berri, D.; Humphreys, B.R. Tournament Incentives, League Policy, and NBA Team Performance Revisited. *J. Sports Econ.* **2010**, *11*, 117–135. [[CrossRef](#)]
35. Arel, B.; Tomas III, M.J. The NBA Draft: A Put Option Analogy. *J. Sports Econ.* **2012**, *13*, 223–249. [[CrossRef](#)]
36. Pollard, R. Evidence of a Reduced Home Advantage When a Team Moves to a New Stadium. *J. Sports Sci.* **2002**, *20*, 969–973. [[CrossRef](#)] [[PubMed](#)]
37. Pollard, R.; Pollard, G. Long-Term Trends in Home Advantage in Professional Team Sports in North America and England (1876–2003). *J. Sport Sci.* **2005**, *23*, 337–350. [[CrossRef](#)] [[PubMed](#)]
38. Zhang, S.; Lorenzo, A.; Gómez, M.A.; Mateus, N.; Gonçalves, B.S.V.; Sampaio, J. Clustering Performances in the NBA According to Players’ Anthropometric Attributes and Playing Experience. *J. Sports Sci.* **2018**, *36*, 2511–2520. [[CrossRef](#)]
39. Morgulev, E.; Azar, O.H.; Bar-Eli, M. Searching for Momentum in NBA Triplets of Free Throws. *J. Sports Sci.* **2020**, *38*, 390–398. [[CrossRef](#)]
40. Chou, J.S.; Pham, A.D. Smart Artificial Firefly Colony Algorithm-Based Support Vector Regression for Enhanced Forecasting in Civil Engineering. *Comput. -Aided Civ. Infrastruct. Eng.* **2015**, *30*, 715–732. [[CrossRef](#)]
41. Guerard, J.B.; Xu, G.; Markowitz, H. A Further Analysis of Robust Regression Modeling and Data Mining Corrections Testing in Global Stocks. *Ann. Oper. Res.* **2020**, 1–21. [[CrossRef](#)]
42. Javed, M.F.; Amin, M.N.; Shah, M.I.; Khan, K.; Iftikhar, B.; Farooq, F.; Aslam, F.; Alyousef, R.; Alabduljabbar, H. Applications of Gene Expression Programming and Regression Techniques for Estimating Compressive Strength of Bagasse Ash Based Concrete. *Crystals* **2020**, *10*, 737. [[CrossRef](#)]
43. Xie, K.; Liang, B.; Dulebenets, M.A.; Mei, Y. The Impact of Risk Perception on Social Distancing During the COVID-19 Pandemic in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6256. [[CrossRef](#)] [[PubMed](#)]
44. Chancellor, S.; De Choudhury, M. Methods in Predictive Techniques for Mental Health Status on Social Media: A Critical Review. *Npj Digit. Med.* **2020**, *3*, 1–11. [[CrossRef](#)] [[PubMed](#)]
45. Abioye, O.F.; Dulebenets, M.A.; Ozguven, E.E.; Moses, R.; Boot, W.R.; Sando, T. Assessing Perceived Driving Difficulties under Emergency Evacuation for Vulnerable Population Groups. *Socio Econ. Plan. Sci.* **2020**, *72*, 100878. [[CrossRef](#)]
46. Kao, L.J.; Chiu, C.C.; Lu, C.J.; Chang, C.H. A Hybrid Approach by Integrating Wavelet-Based Feature Extraction with MARS and SVR for Stock Index Forecasting. *Decis. Support Syst.* **2013**, *54*, 1228–1244. [[CrossRef](#)]

47. Lu, C.J.; Kao, L.J. A Clustering-Based Sales Forecasting Scheme by Using Extreme Learning Machine and Ensembling Linkage Methods with Applications to Computer Server. *Eng. Appl. Artif. Intell.* **2016**, *55*, 231–238. [[CrossRef](#)]
48. Zhang, W.; Goh, A.T. Multivariate Adaptive Regression Splines and Neural Network Models for Prediction of Pile Drivability. *Geosci. Front.* **2016**, *7*, 45–52. [[CrossRef](#)]
49. Natekin, A.; Knoll, A. Gradient Boosting Machines, a Tutorial. *Front. Neurobotics* **2013**, *7*, 21. [[CrossRef](#)] [[PubMed](#)]
50. Torlay, L.; Perrone-Bertolotti, M.; Thomas, E.; Baciú, M. Machine Learning–XGBoost Analysis of Language Networks to Classify Patients with Epilepsy. *Brain Inform.* **2017**, *4*, 159–169. [[CrossRef](#)]
51. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
52. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient kNN Classification with Different Numbers of Nearest Neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1774–1785. [[CrossRef](#)]
53. Friedman, J. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
54. Friedman, J. Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
55. Lawrence, R.; Bunn, A.; Powell, S.; Zambon, M. Classification of Remotely Sensed Imagery Using Stochastic Gradient Boosting as a Refinement of Classification Tree Analysis. *Remote Sens. Environ.* **2004**, *90*, 331–336. [[CrossRef](#)]
56. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
57. Moisen, G.G.; Freeman, E.A.; Blackard, J.A.; Frescino, T.S.; Zimmermann, N.E.; Edwards Jr, T.C. Predicting Tree Species Presence and Basal Area in Utah: A Comparison of Stochastic Gradient Boosting, Generalized Additive Models, and Tree-Based Methods. *Ecol. Model.* **2006**, *199*, 176–187. [[CrossRef](#)]
58. R Core Team. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available online: <http://www.R-project.org> (accessed on 1 February 2021).
59. Milborrow, S.; Hastie, T.; Tibshirani, R. Earth: Multivariate Adaptive Regression Splines. R Package Version 5.3.0. Available online: <https://www.rdocumentation.org/packages/earth> (accessed on 1 February 2021).
60. Chen, T.; He, T.; Benesty, M. XGBoost: Extreme Gradient Boosting. R Package Version 1.3.2.1. Available online: <https://www.rdocumentation.org/packages/XGBoost> (accessed on 1 February 2021).
61. Kuhn, M.; Wing, J.; Weston, S. Caret: Classification and Regression Training. R Package Version, 6.0-86. Available online: <https://www.rdocumentation.org/packages/caret> (accessed on 1 February 2021).
62. Greenwell, B.; Boehmke, B.; Cunningham, J. Gbm: Generalized Boosted Regression Models. R Package Version, 2.1.8. Available online: <https://www.rdocumentation.org/packages/gbm> (accessed on 1 February 2021).
63. Gosso, A. ElmNN: Implementation of ELM (Extreme Learning Machine) Algorithm for SLFN (Single Hidden Layer Feedforward Neural Networks). R Package Version, 1.0. Available online: <https://www.rdocumentation.org/packages/elmNN> (accessed on 1 February 2021).
64. Schliep, K.; Hechenbichler, K.; Lizée, A. Kknn: Weighted k-Nearest Neighbors. R Package Version, 1.3.1. Available online: <https://www.rdocumentation.org/packages/kknn> (accessed on 1 February 2021).
65. Basketball Reference. Available online: <https://www.basketball-reference.com> (accessed on 15 December 2020).
66. Tseng, C.J.; Lu, C.J.; Chang, C.C.; Chen, G.D.; Cheewakriangkrai, C. Integration of Data Mining Classification Techniques and Ensemble Learning to Identify Risk Factors and Diagnose Ovarian Cancer Recurrence. *Artif. Intell. Med.* **2017**, *78*, 47–54. [[CrossRef](#)] [[PubMed](#)]
67. Bolón-Canedo, V.; Sánchez-Maróño, N.; Alonso-Betanzos, A. A Review of Feature Selection Methods on Synthetic Data. *Knowl. Inf. Syst.* **2013**, *34*, 483–519. [[CrossRef](#)]
68. Lewis Colin, D. *Industrial and Business Forecasting Methods*; Butterworth & Co Publishers Ltd.: London, UK, 1982.
69. Juang, W.C.; Huang, S.J.; Huang, F.D.; Cheng, P.W.; Wann, S.R. Application of Time Series Analysis in Modelling and Forecasting Emergency Department Visits in A Medical Centre in Southern Taiwan. *BMJ Open* **2017**, *7*, e018628. [[CrossRef](#)] [[PubMed](#)]
70. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *20*, 134–144. [[CrossRef](#)]