Journal of New Theory
ISSN: 2149-1402

40 (2022) 27-45
*Journal of New Theory*
https://dergipark.org.tr/en/pub/jnt
Open Access

# Random Ensemble MARS: Model Selection in Multivariate Adaptive Regression Splines Using Random Forest Approach

**Dilek Sabancı[1], Mehmet Ali Cengiz[2]**

**Article History**
Received: 22 Jul 2022
Accepted: 27 Sep 2022
Published: 30 Sep 2022
doi:10.53570/jnt.1147323
Research Article

**Abstract** – Multivariate Adaptive Regression Splines (MARS) is a supervised learning model in machine learning, not obtained by an ensemble learning method. Ensemble learning methods are gathered from samples comprising hundreds or thousands of learners that serve the common purpose of improving the stability and accuracy of machine learning algorithms. This study presented REMARS (Random Ensemble MARS), a new MARS model selection approach obtained using the Random Forest (RF) algorithm. 200 training and test data set generated via the Bagging method were analysed in the MARS analysis engine. At the end of the analysis, two different MARS model sets were created, one yielding the smallest Mean Square Error for the test data (Test MSE) and the other yielding the smallest Generalised Cross-Validation (GCV) value. The best model was estimated for both Test MSE and GCV criteria by examining the error of measurement criteria, variable importance averages, and frequencies of the knot values for each model. Eventually, a new model was obtained via the ensemble learning method, i.e., REMARS, that yields result as good as the MARS model obtained from the original data set. The MARS model, which works better in the larger data set, provides more reliable results with smaller data sets utilising the proposed method.

***Keywords*** – *Multivariate adaptive regression splines, random forest, model selection, machine learning, ensemble learning*

***Mathematics Subject Classification (2020)*** – 62G09, 62P99

## 1. Introduction

Machine learning is concerned with designing and analysing models learned from data and developing practical algorithms for prediction [1]. In other words, it suggests using a machine or computer to learn similarly to how the brain learns and predicts [2]. Ensemble learning, another definition of machine learning, refers to a collection of basic models assembled to create a new prediction or classification using the same learning technique. Bagging and Boosting are among the most widely used methods. These methods were designed to improve the stability and accuracy of machine learning algorithms [3]. Random Forest (RF) is an ensemble learning method for classification and regression, representing a significant advancement in machine learning. The method of RF is used to create many individual Classification and Regression Tree (CART) decision trees during the training phase. It aims to find new ways to combine the information from the individual CART trees (the class modes for classification, averaging the predictions of each regression model) [4, 5]. Contrary to the linear and non-linear regression models widely used in practice, Multivariate Adaptive

[1]dilek.kesgin@gop.edu.tr (Corresponding Author); [2]macengiz@omu.edu.tr
[1]Department of Mathematics, Faculty of Arts and Sciences, Tokat Gaziosmanpaşa University, Tokat, Türkiye
[2]Department of Statistics, Faculty of Arts and Sciences, Ondokuz Mayıs University, Samsun, Türkiye

Regression Spline (MARS) is defined as a nonparametric technique that generates different coefficients for different range values of the independent variable and reflects the actual structure better by also adding interaction terms to the model [6]. The model is obtained using the regression models' forward selection and backward elimination algorithms and the essential piecewise functions and combinations [7].

Many studies are conducted in different disciplines using the MARS algorithm. The MARS model was used by [8] in the field of medicine for disease diagnosis, [9] in the field of computer sciences for the quality assessment of web services, by [10] in the field of civil and environmental engineering to estimate pile drivability, by [11] in the field of mechanical engineering to model heat transfer properties and by [12] in the field of accounting and information systems to make financial predictions.

On the other hand, there are other studies in the literature where improvements were made in the model selection of the MARS algorithm. [13] used a genetic algorithm to interfere with knot selection. [14] suggested a new approach to the MARS as CMARS and used a penalised residual sum of squares for the MARS as a Tikhonov regularisation problem. [15] presented the new robust CMARS (RCMARS) in theory and by a robust optimisation technique. [16] proposed a new method for knot selection based on a mapping approach like self-organising maps. [17] applied information measure of complexity (ICOMP) as a powerful model selection criterion for the MARS modelling. [18] used Bootstrapping to obtain the empirical distributions of the parameters and to determine whether they were statistically significant or not in a special case of nonparametric regression.

There were many studies in which Ensemble Learning, Bagging, and Random Forest methods were used together in the literature. We can give the most up to date of these as follows. [19] showed the working mechanism of stacking and bagging on spoof fingerprint detection, which were widely used ensemble learning approaches. [20] offered a quantified way for the standard case when classifiers were aggregated by the majority ("algorithmic variance", i.e., the prediction error variance due only to the randomised training algorithm). A new method was proposed as a multi-objective optimisation approach with the two objectives of accuracy and diversity based on two main challenges in the bagging method provided in [21]. [22] proposed a new classification method for sparse functional data based on functional principal component analysis (FPCA) and bootstrap aggregating.

The present study aims to convert the MARS technique into ensemble learning using the RF algorithm. Moreover, we suggested a new method for selecting the model with superior performance and generalisation from the ensemble of the MARS models. Although there are studies in the literature where the MARS model was used as an ensemble [23-25], there is no method for selecting the best MARS model by creating the MARS models over a random ensemble. Correcting this deficiency in the literature will be attempted with the suggested new method. This study is derived from the first author's PhD dissertation conducted under the supervision of the second author.

## 2. Method

### 2.1. Random Forest Algorithm

The article on the RF model published by Leo Breiman in 2001 took its final shape by referring to the studies conducted by [4, 5, 26-31]. The algorithm of the RF method is built by following the steps presented below [8, 32-36]:

*i.* $k$ sampling is created by the Bagging method so that the number of observations in the original data set has the same number of observations as $n$. Also, each sampling represents a CART decision tree.

*ii.* While $2/3$ of the observations in the original data set are included in the sample as InBag data, $1/3$ of them are excluded from the sample as OOB data to test the established internal error rate of the model.

*iii.* The widest CART decision tree is created with the InBag data. While creating this tree, instead of selecting the best variables out of all existing prediction variables in each node, $p$ out of a total of $m$ independent variables are randomly selected in a split of each node ($p < m$). That is because the tree is not expected to demonstrate excessive growth and overfitting.

*iv.* The previous steps are repeated until the $k$ number of trees that will form the forest is obtained. Afterwards, a new prediction is made by combining the separate class predictions with $k$ trees. It counts how often an examined observation is classified and in which categories. Each observation is assigned a class with a majority of votes determined through the tree sets.

*v.* The predictions made with the OOB observations that are not used in individual trees are used to estimate the internal error rate of the forest. The OOB error rate of each decision tree making up the forest is calculated. The percentage of misclassification is determined as the classification error rate of RF.

## 2.2. Multivariate Adaptive Regression Splines

The MARS technique was developed by the physician and statistician Jerome Harold Friedman in 1991 [37]. According to [38], the MARS is a good innovation in finding suitable conversions to convert the non-linear relationships between dependent and independent variables into a linear structure and determine the interactions between independent variables.

The MARS uses a pair of functions in the form of $[\pm(x-t)]_+$One is the mirror reflection of the other, as the basis function in linear and non-linear expansions that predict the relationship between dependent and independent variables. These function pairs are also called mirror basis functions. The sign $[.]_+$ states that only the positive results of the related functions will be considered. Otherwise, the related functions are considered zero. The basis function pair representing the variable $x$ and the knot value $t$ is defined as follows [3,6,39]:

$$(x-t)_+ = \begin{cases} x-t, x > t \\ 0, otherwise \end{cases} and\ (t-x)_+ = \begin{cases} t-x, x < t \\ 0, otherwise \end{cases} \quad (1)$$

To find the desired model, the MARS uses a two-step process. Forward selection, the first phase of creating the MARS model, resembles forward stepwise regression. Unlike this method, the MARS uses basis function pairs instead of original inputs [6]. Each step finds the main basis function pair (meaning that both functions of a knot are included in the model) that leads to the most decrease in the SSE (Sum of Squares Error) value using the Greddy algorithm. In forward selection, the process of adding terms continues until the maximum number of terms included in the model is reached. The over fitted MARS model that is established by adding certain basis functions under the conditions and is much bigger than the optimum model is formulated as follows [37]:

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m B_m(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{m=1}^{K_m} \left[ s_{km} \left( x_{v(k,m)} - t_{km} \right) \right]_+ \quad (2)$$

In Equation (2), $M$ is the number of basis functions defined as $m = 1,2,\dots,M$. While the quantity $K_m$ represents the number of interactions, the quantity $s_{km}$ takes the value $\pm 1$. The constant term in the model is denoted by $a_0$ while the regression coefficients are denoted by $a_m$. $B_m(x)$ is the *m-th* basis function. The $v(k,m)$ label the independent variables and the $t_{km}$ present knot value on the corresponding variables [37].

Backward elimination constitutes the second phase of the creation of the MARS model. The main reason this stage is required is the inability to compare the models with the SSE value. That is because when SSE is used to compare the models, backward elimination always selects the biggest model. However, the biggest model does not possess the best generalisation performance; on the contrary, it is over fitted and does not produce good results on new data [6, 40]. The model that became over fitted with forwarding selection is subjected to the elimination process to turn it into a model capable of generalising. The GCV (Generalized Cross Validation) criterion, a goodness of fit index, is used to compare the performance of model subsets and

choose the best subset. GCV considers both the error of residuals and model complexity [3,6,41]. For this reason, lower GCV values are accepted much better. The GCV formula, introduced by [42], is calculated as follows [37]:

$$GCV = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{f}_M(x_i)\right]^2 \Big/ \left[1 - \frac{C(M)}{n}\right]^2 \tag{3}$$

In Equation (3), the value $n$ gives the number of observations in the data set, $y_i$ gives the observed value of the dependent variable and $\hat{f}(x_i)$ provides the estimated value of the dependent variable. When the numerator of the formula is examined, it can be observed that the mean of the SSE value, i.e., MSE (Mean Square Error), was used. Therefore, the denominator of the formula renders the GCV criterion both different and essential compared to the SSE value. Cost complexity function $C(M)$ is calculated with the formula as follows [37]:

$$C(M) = trace(B(B^TB)^{-1}B^T) + 1 \tag{4}$$

In Equation (4), $B$ represents the $Mxn$ dimensional data matrix of unconstant $M$ basis functions. The cost complexity function was readapted for the MARS model by [43, 44] and took the form as follows [37]:

$$\tilde{C}(M) = C(M) + dM \tag{5}$$

Here, $d$ denotes the penalty value used to determine the best knot value. $d$ is also known as Degrees of Freedom (DOF). The most suitable penalty value is in the $2 \leq d \leq 4$ range [37].

## 2.3. REMARS: Random Ensemble MARS

### 2.3.1. Difference Between MARS and REMARS

Contrary to linear regression models, in the MARS models, parameter confidence intervals and other controls in the model cannot be directly calculated, as in any nonparametric regression. Techniques related to GCV are used to validate the model. In the model selection of these techniques, the model that gives the lowest value related to the criterion is always selected. Generating hundreds or thousands of the MARS models from an original dataset, selecting the model with the lowest GCV criterion among these models and making predictions with the selected model may appear to be a highly reliable course of action. However, it cannot be deduced that the model with the lowest GCV value, etc., selected from a collection of models will always be the model with the best generalisation ability and performance. This is due to the presence of error measurement criteria such as SSE, MSE, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) and $R^2$ (Coefficient of Determination), which affect the generalisation ability and performance of the model. Although SSE forms the basis of all error measurement criteria and serves as a subset, the main differences in the formulas make the criteria different from each other. Selecting a model where these differences are not in effect impacts the model's generalisation ability. When choosing the best model, preferring the model with the lowest error measurement criteria, such as RMSE, MSE and SSE, generally increase the $R^2$ value. Reaching the highest $R^2$ value with this method causes the error variance of the model to increase in new data. Therefore, when selecting the best model from the MARS models ensemble, a better course of action is observed progressing by considering the confidence interval of the means of different error measurement criteria. This route is not enough on its own. That is because the mean of each error measurement criterion does not correspond to the same model in the MARS models ensemble. Therefore, choosing a model in which knot values and variable importance are not enabled will also affect the validity of the selected model.

### 2.3.2. REMARS Algorithm

REMARS, a new MARS model selection approach that is obtained using the Random Forest algorithm, 95% confidence interval for the mean of error measurement criteria, knot values and contribution percentages of variables to the model, is created with the following steps.

*i.* *Sampling: N* samples (SAMPLE 1, SAMPLE 2, …, SAMPLE N) are created by the Bagging method so that the number of observations in the original data set has the same number of observations as *n*. Due to the bagging method, each sample differs in terms of the observations they contain, although they have the same number of observations.

*ii.* *InBag - Train:* Each sample contains approximately $2n/3$ of the observations in the original data set, and this part is named InBag (InBag 1, InBag 2, …, InBag N). With $n/3$ observation repeated from $2n/3$ observation in InBag, the number of observations of each training dataset (Train 1, Train 2, …, Train N) will be $n(n_1 = n_2 = \cdots = n_N = n)$.

*iii.* *OOB - Test:* The number of observations in the original data set but non-existent in the samples is approximately $n/3$ and is named OOB (OOB 1, OOB 2, …, OOB N). The remaining $2n/3$ of the test data (Test 1, Test 2, …, Test N) with the number of observations $n$ is made up of repeated observations from the data separated as OOB.

*iv.* *Analysis:* After the training and test data sets are created, each training data is analysed with its test data in the MARS program. N models ($Y_1$ = Model 1, $Y_2$ = Model 2, …, $Y_N$ = Model N) are obtained at the end of the analysis. The results of these models were examined using two different selection criteria (Test MSE and GCV).

*v.* *Results:* After obtaining a total of *N* models according to Test MSE and GCV criteria, mean – standard deviation – 95% confidence interval for the mean of error measurement criteria (RMSE, MSE, GCV, MAE, MAPE, SSE and $R^2$) are firstly calculated. Secondly, the number of times each variable's knot is repeated in the models, i.e., their frequency, is determined. Thirdly, each variable's minimum, mean and maximum contribution percentages to the model are obtained.

Fig. 1 shows the flowchart for the REMARS algorithm, which is a new approach for model selection from the randomly generated MARS models ensemble.

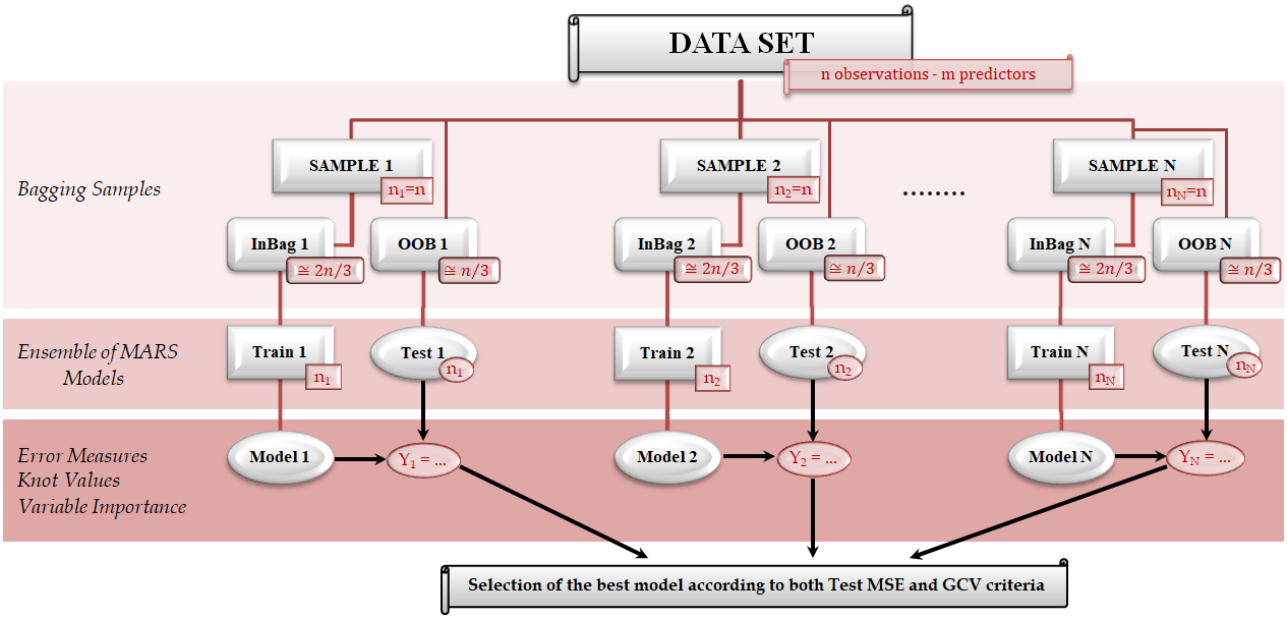


**Fig. 1.** The flowchart of the REMARS (Random Ensemble MARS) algorithm

### 2.3.3. REMARS Model Selection

The information obtained from the results is respectively used for model selection. The first model elimination is performed by selecting the models that fall within the lower and upper bound values of the 95% confidence interval for the mean of each error measurement criterion among Test MSE and GCV models, respectively. Afterwards, the models with the highest frequency in the 95% confidence interval for the mean of each error measurement criterion are determined. Determined models are examined separately by considering the frequency of the knot values obtained from all models (the ensemble) and the descriptive statistics of the contribution percentages of the variables to the model. The model that best predicts the dependent variable is selected among the N models. The model set with REMARS will be as good as the MARS model to be obtained from the original data. That is, because the MARS models obtained from randomly generated samples, combined in 95% confidence limits of the mean of all error measurement criteria of these models, the knot values obtained from all of the models in the ensemble, and the use of the contribution percentages to models of the variables will ensure that the selected model is highly consistent.

## 3. Application

In this section, the application for the selection of the most suitable MARS model using the REMARS method is presented. The MARS analysis engine in the Salford Predictive Modeller version 8.3 by the company Salford System was used in these analyses. the MARS analysis engine, instead of presenting a single MARS model at the end of the analysis, provides a tab including MARS models with several numbers of basis functions. The tab also separately indicates two different models with different selection criteria. The first of these models is the training model that gives the lowest MSE value based on the test data in the backward elimination phase, while the second model is the training model that offers the lowest GCV value in the backward elimination phase. In the application, Test MSE and GCV models were examined separately, and the results were combined.

### 3.1. Data Set

The research of the data set used in the application was conducted in 1970 in East Boston–Massachusetts. The FEV (Forced Expiratory Volume) values of children of different ages, heights and sexes in smoking and non-smoking environments were measured. [45, 46], who conducted the aforesaid research, examined the effects of parents smoking habits on the respiratory functions of children in East Boston–Massachusetts. [47] presented a section of the data from the study, conducted with Tager, for another analysis. This data set used by [48] was published on the website of [49] and used in the application section of the present study. In the FEV data set, there are measurements of 654 children in smoking/non-smoking environments, aged between 3–19 and height vary between 46–74 inches. Their age means is ten and their height mean 61.1 inches. The FEV measurement values of the children range from 0.791–5.793 litres, and the mean FEV value was calculated as 2.637 litres.

### 3.2. Analysis Results

At the end of the analysis, the Salford Predictive Modeller 8.3 MARS program presents various criteria for each model obtained by Test MSE and GCV criteria. These criteria, which will be used to determine the best model, can be listed as RMSE, MSE, GCV, SSE, MAE, MAPE, SSE and $R^2$. Each criterion was obtained separately from the 200 different MARS models created by Test MSE and GCV criteria. Table 1 shows the descriptive statistical findings of each criterion's mean values, standard deviation and 95% confidence interval for the mean (lower limit - upper limit) obtained from 200 models.

**Table 1.** Descriptive statistics of error measurements of GCV and Test MSE models

| Model Selection Criteria | Descriptive Statistics of Measurements | | Model Error Measurements | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | RMSE | MSE | GCV | MAD | MAPE | SSE | $R^2$ |
| Test MSE | Mean | | 0.388 | 0.151 | 0.157 | 0.293 | 0.116 | 98.555 | 0.798 |
| | Standard Deviation | | 0.018 | 0.014 | 0.013 | 0.012 | 0.005 | 9.017 | 0.018 |
| | 95% Confidence Interval for Mean | Lower Bound | 0.385 | 0.149 | 0.156 | 0.291 | 0.115 | 97.555 | 0.796 |
| | | Upper Bound | 0.390 | 0.153 | 0.159 | 0.294 | 0.117 | 99.812 | 0.801 |
| GCV | Mean | | 0.373 | 0.139 | 0.150 | 0.284 | 0.1128 | 90.974 | 0.814 |
| | Standard Deviation | | 0.014 | 0.010 | 0.011 | 0.010 | 0.004 | 6.624 | 0.015 |
| | 95% Confidence Interval for Mean | Lower Bound | 0.371 | 0.138 | 0.149 | 0.282 | 0.112 | 90.050 | 0.811 |
| | | Upper Bound | 0.375 | 0.141 | 0.152 | 0.285 | 0.1134 | 91.898 | 0.816 |

When Table 1 is examined, it is clearly observed that the descriptive statistics values of each error criterion of the Test MSE models are higher than the descriptive statistics values of each error criterion of the GCV models. This situation is due to the fact that the number of basis functions in the Test MSE models is generally lower than the number of basis functions in GCV models. On the other hand, while GCV is based on a formula that considers both residuals and model complexity, the formula of the Test MSE is based only on residuals. For this reason, GCV selects the models with more complex structures in the backward elimination phase. In the results, the GCV values obtained from the Test MSE models are generally higher than those in GCV models.

The knot values accompanying the basis functions that constitute each of the 200 Test MSE and GCV models are also included in determining the best model using REMARS method. As a result, the number of each knot belonging to a variable in the 200 models was determined. Frequencies of the top ten knot values most commonly observed in the models were determined for the height and age variables. Also, frequencies of knot values for the sex and smoker variables in the models were found. All of these results for the Test MSE and GCV models are shown in Table 2. When Table 2 was examined for the height variable, it was observed that the top five most commonly observed knot values are the same for the Test MSE and GCV models; however, their existence percentages in the models differ. Knot 66 was ranked first in both model ensembles by being present with a rate of 51% in the Test MSE models and 53% in the GCV models. In the Test MSE models, knots 65; 65.5; 57.5; 69 are present with rates of 34.5%, 27%, 25.5% and 19.5%, respectively. In the GCV models, knots 69; 65; 65.5; 57.5 are present with rates of 48.5%, 39.5%, 31% and 29.5%, respectively. Knot 67.5 is present in both model ensembles in different percentages.

The age variable's most common knot value in the Test MSE and GCV models is 8. This value is presented in the Test MSE and GCV models at approximately the same rates. The presence rates of knots other than 8 in both model ensembles were lower than 25%. Knots 1 and 2 for the sex variable are observed in the Test MSE models with rates of 10% and 8%, respectively, and the GCV models with rates of 28.5% and 29%, respectively. From this, it can be concluded that while each knot value of the sex variable is present in both Test MSE and GCV models with approximately the same rates, they were more effective in the GCV models compared to the Test MSE models. When Table 2 was examined for the smoker variable, it was observed that the knot value 1 had little effect in both model ensembles. The knot value 0 was much more effective in the

GCV models with a rate of 54.5% compared to the Test MSE models (16.5%). This is because observation 0 outnumbers observation 1 in the data set. That is because the MARS models are affected by the quantitative amount of any observation of a variable in the data set to include it in the model. It is observed that the GCV values are higher than those of the Test MSE, when the knot values of a variable in the Test MSE and GCV models are the same, as shown in Table 2. Again, this is because fewer basis functions, therefore fewer knot values, are included in the Test MSE models compared to the GCV models.

**Table 2.** Frequencies of knot values of GCV and Test MSE models

| Model Selection Criteria | Variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Height | | Age | | Sex | | Smoker | |
| | Knot Values | Frequency | Knot Values | Frequency | Knot Values | Frequency | Knot Values | Frequency |
| Test MSE | 66 | 102 | 8 | 112 | 1 | 20 | 0 | 33 |
| | 65 | 69 | 9 | 34 | | | | |
| | 65.5 | 54 | 10 | 27 | | | | |
| | 57.5 | 51 | 7 | 8 | | | | |
| | 69 | 39 | 13 | 8 | | | | |
| | 58.5 | 31 | 6 | 6 | 2 | 16 | 1 | 2 |
| | 64.5 | 27 | 3 | 5 | | | | |
| | 64 | 27 | 11 | 4 | | | | |
| | 67.5 | 25 | 14 | 4 | | | | |
| | 67 | 23 | 16 | 3 | | | | |
| GCV | 66 | 106 | 8 | 113 | 1 | 57 | 0 | 109 |
| | 69 | 97 | 13 | 49 | | | | |
| | 65 | 79 | 9 | 37 | | | | |
| | 65.5 | 62 | 14 | 29 | | | | |
| | 57.5 | 59 | 10 | 26 | | | | |
| | 70 | 59 | 17 | 21 | 2 | 58 | 1 | 13 |
| | 67.5 | 58 | 15 | 19 | | | | |
| | 67 | 47 | 12 | 16 | | | | |
| | 68.5 | 44 | 16 | 15 | | | | |
| | 71 | 41 | 11 | 13 | | | | |

The MARS analysis engine includes a tab that ranks the predictors based on their contribution percentage to the model at the end of the analysis. In this tab, which is calculated at a 100% scale, variables are ranked according to their importance percentage in a way that the most important variable always scores 100%. Table 3 shows the minimum, mean and maximum value findings of the importance percentages obtained for each variable in the Test MSE and GCV models.

**Table 3**. Importance percentage of variables of GCV and Test MSE models

| Model Selection Criteria | Variables | Variable Importance | | |
|---|---|---|---|---|
| | | Minimum | Mean | Maximum |
| Test MSE | Height | 100.00 | 100.00 | 100.00 |
| | Age | 8.35 | 23.87 | 43.59 |
| | Sex | 1.25 | 9.48 | 18.50 |
| | Smoker | 3.15 | 10.20 | 22.48 |
| GCV | Height | 100.00 | 100.00 | 100.00 |
| | Age | 14.03 | 27.20 | 52.71 |
| | Sex | 1.28 | 8.48 | 22.57 |
| | Smoker | 0.50 | 8.98 | 22.53 |

Table 3 shows that height is the most important variable in the Test MSE and GCV models. It was found to have an importance percentage of 100% in all models. It is followed by the age variable as the variable with the most significant contribution to the models. However, while the age variable is present in all of the GCV models, it was absent in 8 of the Test MSE models. In comparison, the sex and smoker variables had percentages of contribution in the Test MSE models that were calculated as 15.5% and 17%, respectively, and their percentages in the GCV models are 58% and 60.5%. Therefore, while the sex and smoker variables had almost no contribution to the Test MSE models, the situation was the opposite in the GCV models. Thus, the GCV criterion considers model complexity and tends to establish models with too many variables and knots.

## 3.3. Model Selection

The results obtained from Test MSE and GCV models ensemble using the REMARS approach constitute the building block of the model selection. In addition to being an approach that converts the classical MARS method into an ensemble, REMARS also uses a different technique in model selection. Instead of selecting the models with the lowest Test MSE and GCV values among Test MSE and GCV ensembles, it considers each model's error criteria, knot values, and variable importance.

### 3.3.1. Test MSE Criteria

To determine the MARS model with the best performance among the 200 models determined by the Test MSE, the analysis results obtained for the Test MSE in Table 1 are used. The first model elimination is performed by selecting the models that fall within the lower and upper bound values of the 95% confidence interval for the mean of each error measurement criterion among the Test MSE models. That is because progressing by determining a value according to the mean of each error criterion prevents us from selecting over fitted or under fitted models. According to their model number order, the models selected under these conditions are shown in Table 4 for the Test MSE error measurement criteria. When Table 4 is examined, it is seen that some error criteria have common models, and some do not. At this stage, it is important that a particular model number is commonly entering in several error measurement criteria. For this reason, in order to determine the best model, the second model elimination was performed by determining which model number is found in the error measurement criteria in Table 4, and how many times at the most. When Table 4 is examined, it is observed that Model 185 is commonly present in six error measurement criteria other than the MAD error measurement criterion, and Model 3, Model 24, Model 90 and Model 140 are commonly present in five different error measurement criteria. Other Test MSE models are present in fewer numbers in different error measurement criteria.

**Table 4.** Model numbers entering 95% confidence interval for error criterion mean from Test MSE models

| Model Number | Error Measurements of Test MSE Models | | | | | | |
|---|---|---|---|---|---|---|---|
| | **RMSE** | **MSE** | **GCV** | **MAE** | **MAPE** | **SSE** | **$R^2$** |
| | 2 | 2 | 3 | 1 | 1 | 3 | 1 |
| | 3 | 3 | 7 | 3 | 16 | 6 | 3 |
| | 6 | 6 | 17 | 21 | 46 | 24 | 17 |
| | 21 | 24 | 21 | 22 | 53 | 36 | 65 |
| | 24 | 36 | 24 | 24 | 58 | 56 | 70 |
| | 25 | 56 | 25 | 25 | 90 | 59 | 103 |
| | 36 | 59 | 36 | 48 | 92 | 64 | 110 |
| | 56 | 64 | 39 | 59 | 98 | 83 | 135 |
| | 59 | 83 | 51 | 83 | 122 | 90 | 147 |
| | 64 | 90 | 62 | 91 | 125 | 97 | 166 |
| | 90 | 97 | 83 | 103 | 134 | 127 | 180 |
| | 97 | 103 | 90 | 106 | 137 | 133 | 181 |
| | 103 | 108 | 92 | 108 | 140 | 140 | 185 |
| | 108 | 127 | 97 | 117 | 149 | 150 | 191 |
| | 127 | 133 | 108 | 127 | 153 | 152 | 194 |
| | 130 | 140 | 109 | 129 | 157 | 185 | |
| | 133 | 150 | 136 | 130 | 166 | 187 | |
| | 138 | 152 | 140 | 137 | 178 | 194 | |
| | 140 | 174 | 144 | 144 | 185 | | |
| | 150 | 185 | 152 | 177 | | | |
| | 152 | 187 | 171 | 180 | | | |
| | 174 | 190 | 176 | 193 | | | |
| | 185 | 194 | 180 | | | | |
| | 187 | | 183 | | | | |
| | 190 | | 185 | | | | |
| | 194 | | 187 | | | | |
| | | | 188 | | | | |

According to these results, the five most common models in the error measurement criteria among the 200 Test MSE models were determined. Table 5 shows the knot values and variable importance percentages of these models.

In the third model elimination to determine the best out of the five models selected from the Test MSE models, the results in Table 2 are used. When Table 2 is examined, it is observed that in the Test MSE models, knot 66 is essential for the height variable while knot 8 is essential for the age variable. These values are not present in Model 3, Model 24, Model 90 and Model 185. The absence of these knots may cause significant problems in the performance of the error measurement criteria of the models for new data. According to Table 3, the importance percentages of the height and age variables in the models indicate that these variables are required to be present in the models while also serving as moderator variables. It is also observed that the smoker variable contributes to the models. However, the smoker variable is not present in any of the models. That is because knot 0 of the smoker variable was found in 33 models while knot 1 was found in 2 models.

Their percentages of contribution to the models they are included in were found to be high. This shows that the knots contribute to the models, however, this contribution is not sufficient to be included in the selected models. The same situation applies to the sex variable. Model 24 includes the sex variable; however, it cannot be selected as the significant knots of the height variable are not included in this model. When all of these results are combined, Model 140 represents the best model among the Test MSE models. That is because Model 140 has produced the most consistent results in terms of the error measurement criteria, knot values and contribution percentages of variables to the model.

**Table 5.** Knot values and variable importance percentages of selected Test MSE models

| Model Number | Particulars | Variables | | |
|---|---|---|---|---|
| | | Height | Age | Sex |
| 3 | Knot Values | 52 | 6 | |
| | | 65.5 | | |
| | | 66 | | |
| | | 69 | | |
| | Variable Importance (%) | 100 | 25.81 | |
| 24 | Knot Values | 57.5 | 8 | 1 |
| | | 58.5 | | |
| | | 59.5 | | |
| | | 65 | | |
| | Variable Importance (%) | 100 | 25.05 | 7.51 |
| 90 | Knot Values | **65** | 11 | |
| | | | 13 | |
| | Variable Importance (%) | 100 | 20.10 | |
| 140 | Knot Values | 65.5 | 8 | |
| | | 66 | | |
| | | 69 | | |
| | Variable Importance (%) | 100 | 18.65 | |
| 185 | Knot Values | 65 | 7 | |
| | | 66 | | |
| | | 67 | | |
| | Variable Importance (%) | 100 | 33.88 | |

**65**: Knot value entering the model with the mirror basis function

### 3.3.2. GCV Criteria

To determine the MARS model with the best performance among the 200 models determined in accordance with GCV firstly the analysis results obtained for GCV in Table 1 are used. The first model elimination is performed by selecting the models that fall within the lower and upper bound values of the 95% confidence interval for the mean of each error measurement criterion among the GCV models. That is because progressing by determining a value according to the mean of each error criterion prevents us from selecting over fitted or under fitted models. The models selected under these conditions are, according to their model number order, shown in Table 6 for the GCV error measurement criteria. According to Table 6, Model 59, Model 116 and Model 130 are commonly present in five different error measurement criteria. Other GCV models are present in fewer numbers in different error measurement criteria.

**Table 6.** Model numbers entering 95% confidence interval for error criterion mean from GCV models

| Model Number | Error Measurements of GCV Models | | | | | | |
|---|---|---|---|---|---|---|---|
| | RMSE | MSE | GCV | MAD | MAPE | SSE | $R^2$ |
| | 23 | 23 | 9 | 7 | 5 | 23 | 6 |
| | 34 | 34 | 21 | 21 | 6 | 34 | 8 |
| | 46 | 59 | 34 | 52 | 9 | 59 | 12 |
| | 59 | 82 | 46 | 61 | 14 | 82 | 14 |
| | 79 | 104 | 52 | 79 | 28 | 104 | 18 |
| | 82 | 105 | 91 | 98 | 30 | 105 | 25 |
| | 104 | 106 | 92 | 108 | 43 | 106 | 27 |
| | 105 | 108 | 103 | 115 | 46 | 108 | 28 |
| | 106 | 116 | 105 | 127 | 53 | 116 | 30 |
| | 108 | 130 | 116 | 130 | 58 | 130 | 35 |
| | 116 | 146 | 117 | 144 | 59 | 146 | 48 |
| | 130 | 174 | 118 | 151 | 61 | 174 | 57 |
| | 146 | 182 | 130 | 171 | 62 | 182 | 59 |
| | 174 | 190 | 144 | 174 | 70 | 190 | 63 |
| | 190 | 198 | 146 | 183 | 75 | 198 | 69 |
| | 198 | | 182 | 191 | 79 | | 82 |
| | | | 190 | 196 | 91 | | 83 |
| | | | 191 | 198 | 94 | | 94 |
| | | | 196 | | 101 | | 99 |
| | | | | | 119 | | 103 |
| | | | | | 122 | | 107 |
| | | | | | 123 | | 116 |
| | | | | | 128 | | 117 |
| | | | | | 138 | | 136 |
| | | | | | 139 | | 138 |
| | | | | | 160 | | 141 |
| | | | | | 166 | | 151 |
| | | | | | 168 | | 191 |
| | | | | | 172 | | |
| | | | | | 188 | | |
| | | | | | 189 | | |
| | | | | | 193 | | |
| | | | | | 194 | | |
| | | | | | 196 | | |

According to these results, the three most common models in the error measurement criteria among the 200 GCV models were determined. Table 7 shows the knot values and variable importance percentages of these models. The results in Table 2 and Table 3 are used to determine the best out of the three models selected from the GCV models. According to Table 2, knots 66 and 69 for the height variable and knots 8 and 13 for

the age variable repeat a lot in the GCV models. Knot value 69 from the height variable was not included in Model 130 while knot value 8 from the age variable was not included in Model 116. According to the importance percentages of the height and age variables in Table 3, they should be included in the model to be selected. The sex and smoker variables are seen to have similar contributions to the models. However, as knot 0 of the smoker variable has a significantly higher number of repetitions in the models compared to the knot values of the sex variable, knot 0 of the smoker variable should be included in the model. Based on these observations, Model 59 represents the best model among the GCV models.

**Table 7.** Knot values and variable importance percentages of selected GCV models

| Model Number | Particulars | Variables | | | |
|---|---|---|---|---|---|
| | | Height | Age | Sex | Smoker |
| 59 | Knot Values | **58** | 8 | 2 | 0 |
| | | 62 | 11 | | |
| | | 66 | 13 | | |
| | | 69 | 15 | | |
| | Variable Importance (%) | 100 | 34.53 | 15.62 | 6.60 |
| 116 | Knot Values | 65 | 9 | 1 | |
| | | 66 | | | |
| | | 67 | | | |
| | | 67.5 | | | |
| | | 69 | | | |
| | | 70.5 | | | |
| | Variable Importance (%) | 100 | 31.08 | 14.71 | |
| 130 | Knot Values | 56.5 | 8 | 2 | |
| | | 60 | | | |
| | | 66 | | | |
| | | 67.5 | | | |
| | | 68.5 | | | |
| | | 71 | | | |
| | Variable Importance (%) | 100 | 25.06 | 7.13 | |

**58**: Knot value entering the model with mirror basis function

## 4. Discussion

The results related to the error measurement criteria what presented incidental to Model 140 selected from the Test MSE models using the REMARS method, Model 59 selected from the GCV models using the REMARS method and the MARS model obtained from the original data set are shown in Table 8.

**Table 8.** Error measurement criteria of the MARS models obtained from original data and REMARS

| Data Sets | Error Measurement Criteria | | | | | | |
|---|---|---|---|---|---|---|---|
| | RMSE | MSE | GCV | MAD | MAPE | SSE | $R^2$ |
| Original Data Set | 0.385 | 0.149 | 0.160 | 0.291 | 0.115 | 97.173 | 0.802 |
| GCV-Model 59 | 0.372 | 0.138 | 0.152 | 0.277 | 0.113 | 90.349 | 0.813 |
| Test MSE-Model 140 | 0.389 | 0.151 | 0.157 | 0.298 | 0.116 | 98.938 | 0.788 |

When each of the values of the error measurement criteria shown in Table 8 is compared individually for both the Test MSE-Model 140 and GCV-Model 59, it is observed that GCV-Model 59 produces more consistent and reliable results compared to the Test MSE-Model 140. That is because the GCV criterion is calculated with a formula that considers both error and the number of effective parameters. For this reason, it tends to create models with a more complex structure and higher variable efficiency compared to the Test MSE criterion. This situation causes the error measurement criteria of the Test MSE models, which have simpler structures, to deteriorate. Therefore, the model created based on the GCV criterion has a better ability to generalise compared to the model created based on the Test MSE criterion. This is clearly observed when the knot values and importance percentages of the variables in the Test MSE-Model 140 and GCV-Model 59 are examined.

When the values of the error measurement criteria of the MARS model obtained from the original data set are compared with the error measurement criteria of GCV-Model 59, it is observed that GCV-Model 59 produces better results. On the other hand, the knot values in the MARS model obtained from the original data set were obtained as 58.5; 59.5; 65; 66; 69 for the height variable, 8 for the age variable, 2 for the sex variable and 0 for the smoker variable. When these knot values are compared with the knot values of GCV-Model 59, it is observed that knots 66 and 69, which are important for the height variable, are present in both models. The age variable entered the model with a higher knot value in GCV-Model 59. This can be an advantage or disadvantage for the model. However, [48] proved that the abundance in age knots is an advantage for the model. That is because it was shown diagrammatically that the FEV distribution of the children in the smoking environment changed direction in knots 11, 13 and 15. The number of smoker parents is very low in the FEV data and when the number of smoker parents is increased in a different sample, these knots in the age variable take on an important role for the model. Both models have the same knot values for the sex and smoker variables. The contribution percentages of the height, age, sex and smoker variables in the MARS model were obtained as 100%, 27.4%, 4.88% and 2.91%, respectively. When these values are compared with GCV-Model 59, it is understood that the age, sex and smoker variables contribute more to GCV-Model 59. In conclusion, Model 59, which was obtained with REMARS method based on the GCV criterion, produced better results than any other model.

All prediction curves for the actual and predicted FEV values shown in Fig. 2 appear to be overlapped. Here, it can be understood that MSE-Model 140 and GCV-Model 59, which were obtained with REMARS method, produce results that are as consistent and reliable as the ones produced by the MARS model obtained from the original data. On the other hand, the Pearson Correlation between the actual and predicted FEV values was calculated as $r = 0.902$ $(p = 0.001)$ for the GCV-Model 59, $r = 0.888$ $(p = 0.001)$ for the Test MSE-Model 140 and $r = 0.896$ $(p = 0.001)$ for the MARS model obtained from the original data set. The fact that the best correlation between the actual and predicted FEV values are produced by GCV-Model 59 is understood from both the Pearson Correlation Coefficient Value and the scattering of the observation values in Fig. 2.

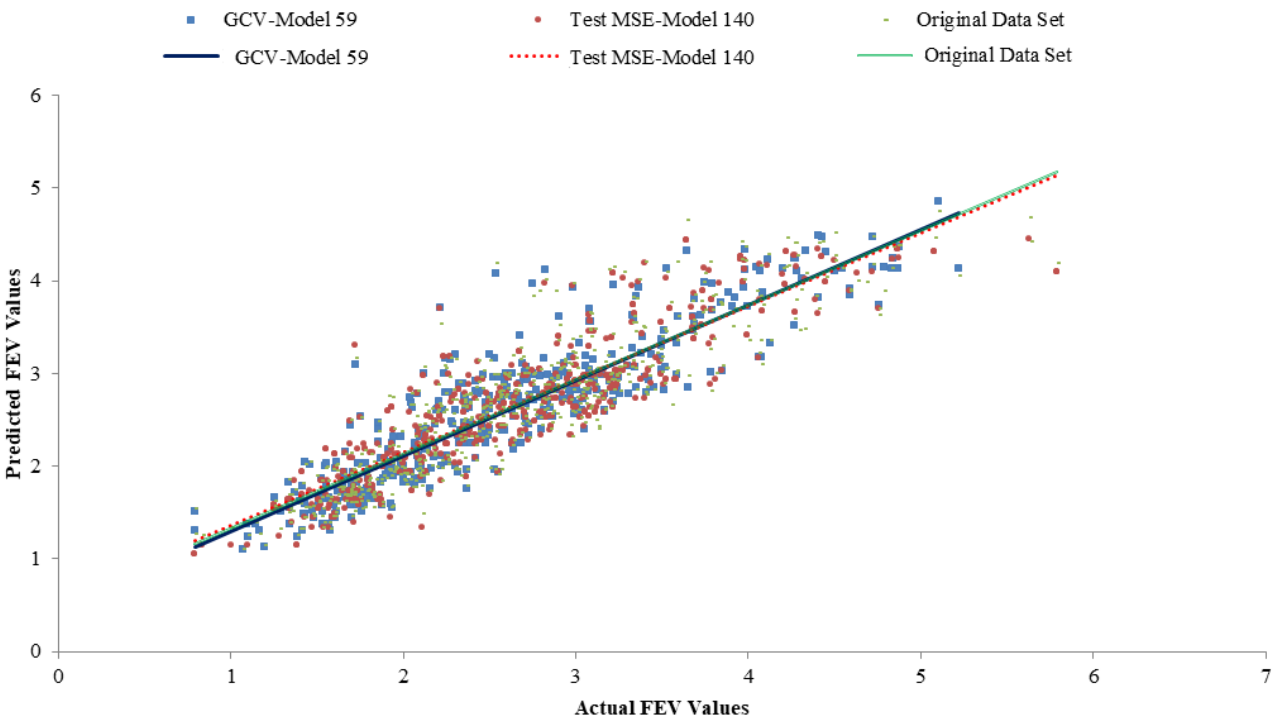


**Fig. 2.** Distribution of actual and predicted values of the MARS model obtained original data and REMARS

Based on the results obtained, Table 9 shows the basis functions and knot values for Model 59, which was selected with REMARS method based on the GCV criterion, and information on the prediction model.

**Table 9.** Basis functions and corresponding equations of the MARS model for GCV-Model 59

| Basis Functions | Equation | Model |
|---|---|---|
| BF1 | Max (0, Height-57.5) | FEV = 1.71963 +0.14058*BF1 - 0.0692608*BF2+0.0754913*BF3 +0.145284*BF5+0.260453*BF7 - 0.280703*BF9-0.106875*BF11 +0.172786*BF13-0.315717*BF23 +0.295305*BF25+0.127262*BF27 |
| BF2 | Max (0, 57.5- Height) | |
| BF3 | Max (0, Age-8) | |
| BF5 | (Sex (2)) | |
| BF7 | Max (0, Height -66) | |
| BF9 | Max (0, Height -69) | |
| BF11 | Max (0, Height -62) | |
| BF13 | (Sex (0)) | |
| BF23 | Max (0,Age-13) | |
| BF25 | Max (0, Age -15) | |
| BF27 | Max (0, Age -11) | |

## 5. Conclusion

Model 59, selected from GCV models, produces results more consistent than MARS model created by taking the original data set as training data. It is more suitable to be used for model selection particularly in data set where observations such as FEV data do not demonstrate homogeneity. The MARS model obtained from the original data set was created without being tested with separate test data. Therefore, it is not known whether it is the most useful model for new data. For this reason, the use of the model obtained through ensemble learning instead of the model obtained with a single learner produces more valid and reliable results. The MARS model obtained based on the REMARS method is suggested for this reason. The MARS model works better in big data set. The MARS model obtained using the REMARS method can produce reliable results with smaller data set due to the different samples generated with the Bagging Method. In data set with too many parameters, the procedure of independent variable selection can be carried out, as in the RF method.

## Author Contributions

All authors contributed equally to this work. They all read and approved the last version of the manuscript.

## Conflict of Interest

The authors declare no conflict of interest.

## References

[1] S. Theodoridis, Machine Learning a Bayesian and Optimisation Perspective, Academic Press of Elsevier, 125 London Wall, London, 2015.

[2] S. Suthaharan, Machine Learning Models and Algorithms for Big Data Classification, Springer International Publishing, New York, 2016.

[3] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer Series in Statistics, Stanford, California, 2001.

[4] T. K. Ho, *Random Decision Forests*, Proceedings of 3rd International Conference on Document Analysis and Recognition (IEEE), Montreal, Canada, 1995, pp. 278–282.

[5] T. K. Ho, *The Random Subspace Method for Constructing Decision Forests*, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (2) (1998) 832–844.

[6] T. Hill, P. Lewicki, Statistics: Methods and Applications, StatSoft, Tulsa OK, 2006.

[7] J. R. Leathwick, J. Elith, T. Hastie, *Comparative Performance of Generalised Additive Models and Multivariate Adaptive Regression Splines for Statistical Modelling of Species Distributions*, Ecological Modelling 199 (2) (2006) 188–196.

[8] D. Yao, J. Yang, X. Zhan, *A Novel Method for Disease Prediction: Hybrid of Random Forest and Multivariate Adaptive Regression Splines*, Journal of Computers 8 (1) (2013) 170–177.

[9] L. Kumar, S. K. Rath, *Quality Assessment of Web Services Using Multivariate Adaptive Regression Splines*, in: J. Sun, Y. R. Reddy, A. Bahulkar, A. Pasala (Eds.), 22nd Asia-Pacific Software Engineering Conference, New Delhi, India, 2015, pp. 238–245.

[10] W. Zhang, A. T. Goh, *Multivariate Adaptive Regression Splines and Neural Network Models for Prediction of Pile Drivability*, Geoscience Frontiers 7 (1) (2016) 45–52.

[11] P. Dey, A. K. Das, *Application of Multivariate Adaptive Regression Spline-Assisted Objective Function on Optimisation of Heat Transfer Rate Around a Cylinder*, Nuclear Engineering and Technology 48 (6) (2016) 1315–1320.

[12] Y. J. Chen, J. A. Lın, Y. M. Chen, J. H. Wu, *Financial Forecasting with Multivariate Adaptive Regression Splines and Queen Genetic Algorithm-Support Vector Regression*. IEEE Access 7 (2019) 112931–112938.

[13] J. Pittman, *Adaptive Splines and Genetic Algorithms*, Journal of Computational and Graphical Statistics 11 (3) (2002) 615–638.

[14] G. W. Weber, I. Batmaz, G. Köksal, P. Taylan, F. Y. Özkurt, *CMARS: A New Contribution to Nonparametric Regression with Multivariate Adaptive Regression Splines Supported by Continuous Optimisation*, Inverse Problems in Science and Engineering 20 (3) (2012) 371–400.

[15] A. Özmen, G. W. Weber, I. Batmaz, E. Kropat, *RCMARS: Robustification of CMARS with Different Scenarios Under Polyhedral Uncertainty Set*, Communications in Nonlinear Science and Numerical Simulation 16 (12) (2011) 4780–4787.

[16] E. K. Koc, C. Iyigun, *Restructuring Forward Step of MARS Algorithm Using a New Knot Selection Procedure Based on a Mapping Approach*, Journal of Global Optimization 60 (2014) 79–102.

[17] E. K. Koc, H. Bozdogan, *Model Selection in Multivariate Adaptive Regression Splines (MARS) Using Information Complexity as the Fitness Function*, Machine Learning 101 (2015) 35–58.

[18] C. Yazıcı, F. Y. Özkurt, I. Batmaz, *A Computational Approach to Nonparametric Regression: Bootstrapping CMARS Method*, Machine Learnig 101 (2015) 211–230.

[19] S. Agarwal, C. R. Chowdary, C. R., *A-Stacking and A-Bagging: Adaptive Versions of Ensemble Learning Algorithms for Spoof Fingerprint Detection*, Expert Systems with Applications Article ID 113160 (2020) 10 pages.

[20] M. E. Lopes, *Estimating the Algorithmic Variance of Randomised Ensembles via the Bootstrap*, The Annals of Statistics 47 (2) (2019) 1088–1112.

[21] S. E. Roshan, S. Asadi, *Improvement of Bagging Performance for Classification of Imbalanced Datasets Using Evolutionary Multi-Objective Optimization*, Engineering Applications of Artificial Intelligence Article ID 103319 (2020) 19 pages.

[22] H. Kim, Y. Lim, *Bootstrap Aggregated Classification for Sparse Functional Data*, Journal of Applied Statistics 49 (8) (2022) 2052–2063.

[23] W. Pintowati, B. W. Otok, *Pemodelan Kemiskinan di Propinsi Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Splines Ensemble*, Jurnal Sains dan Seni ITS 1 (1) (2012) 283–288.

[24] K. D. Roy, B. Datta, *Multivariate Adaptive Regression Spline Ensembles for Management of Multilayered Coastal* Aquifers, Journal of Hydrologic Engineering 22 (9) (2017) 04017031.

[25] R. Zheng, M. Li, X. Chen, S. Zhao, F. Wu, Y. Pan, J. Wang, *An Ensemble Method to Reconstruct Gene Regulatory Networks Based on Multivariate Adaptive Regression Splines*, IEEE/ACM Transactions on Computational Biology and Bioinformatics 18 (1) (2019) 347–354.

[26] L. Breiman, J. Friedman, C. J. Stone, R. Olshen, Classification and Regression Trees. Belmont: Taylor & Francis, New York, 1984.

[27] E. M. Kleinberg, *Stochastic Discrimination*, Annals of Mathematics and Artificial Intelligence 1 (1990) 207–239.

[28] E. M. Kleinberg, *An Overtraining-Resistant Stochastic Modelling Method for Pattern Recognition*, The Annals of Statistics 24 (6) (1996) 2319–2349.

[29] E. M. Kleinberg, *On the Algorithmic Implementation of Stochastic Discrimination*, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (5) (2000) 473–490.

[30] L. Breiman, Bagging Predictors (Report No. 421). Department of Statistics University of California. Berkeley, California, 1994.

[31] Y. Amit, D. Geman, *Shape Quantization and Recognition with Randomised Trees*, Neural Computation 9 (7) (1997) 1545–1588.

[32] L. Breiman, *Random Forest*, Machine Learning 45 (1) (2001) 5–32.

[33] M. Akman, Y. Genç, H. Ankaralı, *Random Forests Methods and an Application in Health Science*, Turkiye Klinikleri Journal of Biostatistics 3 (1) (2011) 36–48.

[34] J. Abellán, C. J. Mantas, J. G. Castellano, *A Random Forest Approach Using Imprecise Probabilities*, Knowledge-Based Systems 134 (2017) 72–84.

[35] A. Liaw, M. Wiener, R Project. The R Project for Statistical Computing: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf. Accessed on April 9, 2019.

[36] Minitab, Minitab: http://www.minitab.com/uploadedFiles/Content/Products/SPM/IntroRF_v_8_2.pdf. Accessed on April 9, 2019.

[37] J. H. Friedman, *Multivariate Adaptive Regression Splines*, The Annals of Statistics 19 (1) (1991) 1–67.

[38] J. Deichmann, A. Eshghi, D. Haughton, S. Sayek, N. Teebagy, *Application of Multiple Adaptive Regression Splines (MARS) in Direct Response Modeling*, Journal of Interactive Marketing 16 (4) (2002) 15–27.

[39] G. O. Temel, H. Ankaralı, A. C. Yazıcı, *An Alternative Approach to Regression Models: MARS*, Turkiye Klinikleri Journal of Biostatistics 2 (2) (2010) 58–66.

[40] J. Strickland, Predictive Analytics Using R. Lulu Press (Lulu.com), Morrisville, North Carolina, USA, 2015.

[41] L. C. Briand, B. Freimut, F. Vollei, Using Multiple Adaptive Regression Splines to Understand Trends in Inspection Data and Identify Optimal Inspection Rates (Report No. 062.00/E). Fraunhofer IESE, Kaiserslautern, 2001.

[42] P. Craven, G. Wahba, *Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalised Cross-Validation*, Numerische Mathematik 31 (4) (1978) 377–403.

[43] J. H. Friedman, Fitting Functions to Noisy Data in High Dimensions (Technical Report No. LCS 101). Stanford University, Department of Statistics, Stanford, CA, 1988.

[44] J. H. Friedman, B. W. Silverman, *Flexible Parsimonious Smoothing and Additive Modelling*, Technometrics 31 (1) (1989) 3–21.

[45] I. B. Tager, S. T. Weiss, B. Rosner, F. E. Speizer, *Effect of Parental Cigarette Smoking on the Pulmonary Function of Children*, American Journal of Epidemiology 110 (1) (1979) 15–26.

[46] I. B. Tager, S. T. Weiss, A. Munoz, B. Rosner, F. E. Speizer, *Longitudinal Study of the Effects of Maternal Smoking on Pulmonary Function in Children*, New England Journal of Medicine 309 (12) (1983) 699–703.

[47] B. Rosner, Fundamentals of Biostatistics. Duxbury Press, Pacific Grove, CA, 1999.

[48] M. Kahn, *An Exhalent Problem for Teaching Statistics*, Journal of Statistics Education 13 (2) (2005) 1–11.

[49] Journal of Statistics Education, JSE Data Archive. http://jse.amstat.org/datasets /fev.dat.txt. Accessed on October 10, 2017.