International Journal of Innovative
Computing, Information and Control
Volume **15**, Number **6**, December **2019**

 ISSN 1349-4198
pp. **2027–2037**

RANDOM FOREST AND SUPPORT VECTOR MACHINE ON FEATURES SELECTION FOR REGRESSION ANALYSIS

Christine Dewi and Rung-Ching Chen*

Department of Information Management
Chaoyang University of Technology
No. 168, Jifeng East Road, Wufeng District, Taichung 41349, Taiwan
s10714904@cyut.edu.tw; *Corresponding author: crching@cyut.edu.tw

Received March 2019; revised July 2019

Abstract. *Feature selection becomes predominant and quite prominent in the case of datasets that are contained with a higher number of variables. RF (Random Forest) has emerged as a robust algorithm that can handle a feature selection problem with a higher number of variables. It is also very much efficient while dealing with regression problems. In this work, we proposed the combination of RF, SVM (Support Vector Machine) and tune SVM regression to improve the model performance. We use four outstanding regression datasets from the UCI (University of California Irvine) machine learning repository. In addition, the ranking of important features by RF for affection factors is given out. We prove that it is essential to select the best features to improve the performance of the model. The experimental results show that our proposed model has a better effect compared to other methods in each dataset. The trend of RMSE (Root Mean Squared Error) value is decreased, and the r-value is increased in every experiment for all datasets. Furthermore, it is indicated that the regression predictions perfectly fit the data.*

Keywords: Random forest, Features selection, SVM, Regression

1. **Introduction.** In the area of data processing and analysis, a dataset may have large numbers of variables or attributes which determine the applicability and usability of the data [1]. Furthermore, it is essential that we select the best set of attributes which improves the performance of the model, increases the computational efficiency, and decreases the storage requirements. It is clear that every feature may not be contributing substantially. For different applications, a subset of variables can provide us an equivalent and effective attention. Finding the related features may be termed as feature selection in the area of machine learning. This is also known as variable selection, attribute selection, variable or attribute subset selection. This approach may reduce the data training time and effort. Most of the time the data set includes a lot of features with different qualities that can influence the performance of the classifiers. For instance, noisy features can affect the performance of the algorithm. The reduction of the original feature set to a smaller one preserving the relevant information while discarding the redundant one is referred to as FS (Feature Selection) [2]. In order to tackle this issue and use a smaller number of training samples, the use of feature selection and extraction techniques would be of importance. The concept of feature selection came into the picture after 1995 around. Blum and Langley focus on two problems: the issue of selecting relevant features and the issue of choosing relevant examples and produce a general framework to compare different algorithms [2]. Many researchers have been done on the ranking of variables for the feature selection, for example in [3,4]. Furthermore, two popular methods, Boosting [5]

DOI: 10.24507/ijicic.15.06.2027

and Bagging [6] were proposed to generate many classifiers and aggregate their results for the classification tree.

In this work, we will compare the differences using different combinations of features. Next, we will see if it can make better performance in selecting features that have good accuracy with the data to be predicted. Machine learning needs lots of data and features to make predictions more accuracy, but feature selection is more important than designing the prediction model. Furthermore, using the dataset without pre-processing will make the prediction result worse. In this paper, we will show how important the features selection processed. The main contributions of this work can be summarized as follows. First, this work will conduct an analysis of variable importance to find out which variables are more relevant especially for regression data. The study has been carried out with Random Forest, and some discussion is provided in order to get some insight into the selection of the adequate importance metric. Second, the system will compare different machine learning models, such as SVM, RF, and combined SVM and RF together. Different models will have different strengths in predicting data; we tried to combine RF, SVM and tune SVM regression to make the accuracy better. The *tune()* function tunes hyper parameters of statistical methods using a grid search. This function is a large list. It has a lot of output but at this point, we are interested in knowing which parameter values for gamma and cost are the best. Moreover, this function will improve accuracy. The whole work has been done in R [7], a free software programming language that is specially developed for statistical computing and graphics. The remainder of the paper is organized as follows. Section 2 provides a review of the material and methods. Section 3 presents our results and discussion. Finally, conclusions are drawn, and future research directions are indicated in Section 4.

2. Material and Methods.

2.1. **Random forest.** RF consists of a combination of decision-trees. It improves the classification performance of a single tree classifier by combining the bootstrap aggregating, also called bagging method and randomization in the selection of partitioning data nodes in the construction of a decision tree [8]. A decision tree with M leaves splits the feature space into M regions R_m, $1 \leq m \leq M$. For each tree, the prediction function $f(x)$ is defined as Formulas (1) and (2):

$$f(x) = \sum_{m=1}^{M} c_m \prod(x, R_m) \tag{1}$$

where M is the number of regions in the feature space, R_m is a region corresponding to m, c_m is a constant corresponding to m:

$$\prod(x, R_m) = \begin{cases} 1, & \text{if } x \in R_m \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The final classification decision is made from the majority a vote of all trees.

2.2. **Importance features study.** Variable importance analysis with Random Forest has received a lot of attention for many researchers, but there remain some open issues that need a satisfactory answer. An overview could be found in [9-12]. The important procedure implemented in R for RF provides two important reliable measures for each explanatory variable.

The first measure, *%IncMSE*, accounts for the mean decrease in accuracy or how the prediction gets worse when that variable changes its value. It is computed from permuting test data: For each tree, the prediction error on the test is recorded MSE (Mean Squared

Error). Then the same is done after permuting each predictor variable. The difference is the average over all trees and normalized by the standard deviation of the differences. If the standard deviation of the differences is equal to 0 for the variable, the division is not done. The average is almost always equal to 0 in that case. The higher the difference is, the more important the variable. It uses the out OOB (Out of Bagging) concept: A group of regression trees. The OOB subset, which has been kept out for the construction of each tree, is used to calculate a mean squared error as Formula (3) [13].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

where y_i is the actual hourly price, $\hat{y}_i$ the predicted one and n the number of data in the OOB set. Each tree b and variable j, which has been used to create the tree is randomly permuted in the OOB set. A new MSE (Mean Squared Error) is calculated and the value of the importance of the variable may be computed from the expression of Formula (4).

$$\bar{\delta}_j = \frac{1}{B}\sum_{b=1}^{B}\left(MSE - MSE_{permuted_j}\right) = \frac{1}{B}\sum_{b=1}^{B}\delta_{bj} \tag{4}$$

which is an average over all trees (B) of the forest where variable j has been used. The final value of the importance is obtained by normalizing with the standard error as Formula (5).

$$\%IncMSE = \frac{\bar{\delta}_{bj}}{\sigma_{\delta_{bj}}/\sqrt{B}} \tag{5}$$

where $\sigma_{\delta_{bj}}$ is the standard deviation of the δ_{bj}. A higher *%IncMSE* represents higher variable importance [13]. The second important measure, *IncNodePurity* relates to the loss function, which is chosen by best splits. The loss function is MSE for regression and Gini-impurity for classification. More useful variables achieve higher increases in node purities that is to find a split that has a high inter-node variance and a small intro node variance.

2.3. **Support vector machines & SVR (Support Vector Regression).** SVM is a machine learning algorithm. In recent years, there have been plenty of researches on SVM and introduced as a powerful method for classification. An overview can be found in [14-16]. The other research describes that SVM uses a high dimension space to find a hyperplane to perform binary classification where the error rate is minimal [17-19]. A basic input data format and an output data domain are given as Formula (6).

$$(x_i, y_i), \ldots, (x_n, y_n),\ x \in R_m,\ y \in \{+1, -1\} \tag{6}$$

where $(x_i, y_i), \ldots, (x_n, y_n)$ are training data, n is the number of samples, m is the input vector, and y belongs to the category of $+1$ or -1.

The boundary between classes is defined by a hyperplane computed as a linear combination of a subset of the data points, called Support Vectors (SVs). A regression problem requires the prediction of a quantity and regression can have real-valued or discrete input variables. Moreover, a problem with multiple input variables is often called a multivariate regression problem. The SVM approach was more recently extended to regression problems [20], a domain in which it was SVR. The output of an SVR is computed as Formula (7).

$$Y_{\text{svr}}(x) = \sum_{i=1}^{n}\beta_i k(x; x_i) + b \tag{7}$$

where β_i and x_i are respectively the weight and the position of each SVs. In addition, n is the number of SVs, b is the bias, and $k(x; x_i)$ is the kernel function corresponding to x_i. In the standard approach, a single kernel function is used, whose shape is characterized by a set of parameters. Like other methods based on kernels, the quality of the regression depends on the choice of the kernel function and its parameters, which must be suitable to the current data [21].

In order to avoid over-fitting, the SVR function allows us to penalize the regression through cost function. The SVR technique is flexible in terms of the maximum allowed error and penalty cost. This flexibility allows us to vary both these parameters to perform a sensitivity analysis in an attempt to come up with a better model. Now we will perform sensitivity analysis, by training a lot of models with different allowable errors and cost parameters. This process of searching for the best model is called tuning of the SVR model. Parameter tuning of function is a grid search. This generic function tunes hyper parameters of statistical methods using a grid search. In this research, we use *tune()* function and tuning of the SVR model can be performed as the technique provides flexibility with respect to maximum error and penalty cost. Tuning the model is extremely important as it optimizes the parameters for the best prediction.

2.4. **Caret (classification and regression training) package.** The Caret package has several functions that attempt to streamline the model building and evaluation process. This package contains functions to streamline the model training process for complex regression and classification problems. The package utilizes some R packages but tries not to load them all at package start-up. By removing formal package dependencies, the package start-up time can be significantly decreased. The package suggests the field includes 30 packages. Caret loads packages as needed and assumes that they installed. If a modeling package is missing, there is a prompt to install it. The package contains tools for data splitting, pre-processing, feature selection, model tuning using resampling, variable importance estimation, as well as other functionality [22].

2.5. **Research workflow.** Figure 1 describes the workflow of this research. In addition, the experiment consists of several steps. First, we use the Random Forest to select essential features from each dataset. Second, the construction of the different machine learning models uses SVM, RF, and combines SVM and RF together. Different models will have different strengths in predicting data.

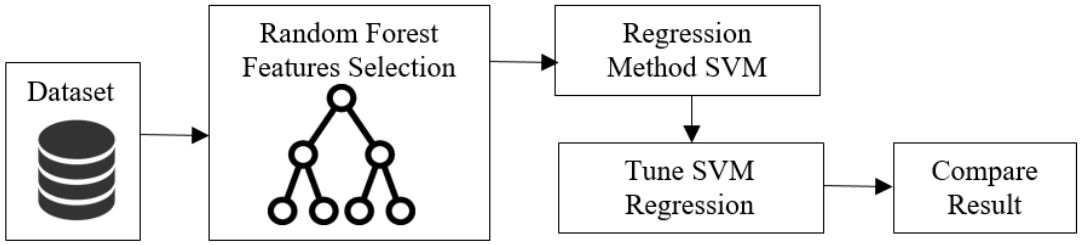

FIGURE 1. The workflow of this research

We tried to combine all advantages from each method RF, SVM and tune SVM regression to make the accuracy better and the last step is to compare the result. The important features for each dataset will be selected by RF. RF is more recent than the other techniques applied in this paper. It was developed by Breimann [23,24] as a way of obtaining more accurate predictions without overfitting the data. RF is similar to Bagging, but additionally making use of a randomized subset of predictors for each split

of each tree. This small difference in the way of building trees allows obtaining so many different trees that a better and more accurate prediction is obtained. The number of trees we have worked with in the use of RF is 500. Previous study shows that RF is able to provide accurate results both in terms of important variable assessment and prediction accuracy. They mitigate the instability problem resulting from training sample changes [23,24]. Next, we use these important features to build the SVM model. SVM is introduced as a powerful method for classification and regression analysis. An overview can be found in [14-16]. The other research describes that SVM uses a high dimension space to find a hyperplane to perform binary classification where the error rate is minimal [17-19]. Another important point is to check the SVM algorithm parameters. As many machine learning algorithms, SVM has some parameters that have to be tuned to gain better performance. This is very important because SVM is very sensitive to the choice of parameters. Even close parameter values might lead to very different classification results. To find the best solution to this problem, we will test with some different values. In addition, we use the *svm()* and *tune.svm()* function in e1071 package of R language to build SVM model. RBF (Radial Basis Function) kernel is also called the Gaussian kernel function. RBF kernel function is the most efficient one owing to its need to set very few parameters and the powerful nonlinear learning ability. Thus, the kernel function is the RBF kernel [25]. Two parameters need to be fixed: cost and gamma [26,27]. We will select the best cost and gamma using the *tune.svm()* function. Furthermore, tune SVM regression has the capacity of solving the problems of nonlinearity, small sample and high dimension [16,17,28,29]. This combination method will improve the accuracy of the regression analysis.

2.6. **Model performance evaluation.** The performance is evaluated with the statistical indicators that were selected to estimate the performance of the proposed models. Since our proposed model focuses on the regression analysis we use regression evaluation metrics. In addition, RMSE (Root Mean Squared Error) is the most common metric used to measure accuracy for continuous variables and regression analysis. In this research, we use statistical indicators. First, RMSE is just the square root of MSE. The square root makes the scale of the errors to be the same as the scale of targets. The equation is Formula (8) [30,31]. The smaller values of RMSE indicate a more satisfactory result [25].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{8}$$

Second, the Pearson correlation coefficient (r), the coefficient of determination is between 0 and 1. If the r value is 1, it is indicated that the regression predictions perfectly fit the data. The equation is Formula (9) [31,32].

$$r = \frac{n\left(\sum_{i=1}^{n} O_i \cdot P_i\right) - \left(\sum_{i=1}^{n} O_i\right) \cdot \left(\sum_{i=1}^{n} P_i\right)}{\sqrt{\left(n\sum_{i=1}^{n} O_i^2 - \left(\sum_{i=1}^{n} O_i\right)^2\right) \cdot \left(n\sum_{i=1}^{n} P_i^2 - \left(\sum_{i=1}^{n} P_i\right)^2\right)}} \tag{9}$$

where P_i and O_i are the experimental and forecast values, respectively, and n is the total number of test data.

3. Results and Discussion.

3.1. Dataset descriptions.

These work simulations use four datasets publicly available from the UCI machine learning repository. All of the datasets belong to regression data and have different total instances and features. The description of each dataset could be found in Table 1.

TABLE 1. Dataset descriptions

No	Dataset	Instance	Feature	Year
1	Wisconsin Breast Cancer Database Dataset [22,33]	699	10	1991
2	Forest Fire Dataset [34]	517	13	2008
3	Wine Quality Dataset [35]	4898	12	2009
4	Bike Sharing Dataset [36]	17379	15	2013

Table 1 shows a dataset that belongs to regression data and uses in this experiment. We use the Wisconsin Breast Cancer Dataset which publishes in 1991 with 699 instances and 10 features, Forest Fire Dataset in 2008 with 517 instances and 13 features, Wine Quality Dataset in 2009 with 4898 instances and 12 features, Bike Sharing Dataset in 2013 with 17379 instances and 15 features. Furthermore, the important measure for each variable and dataset by RF could be seen in Figure 2 and Figure 3.

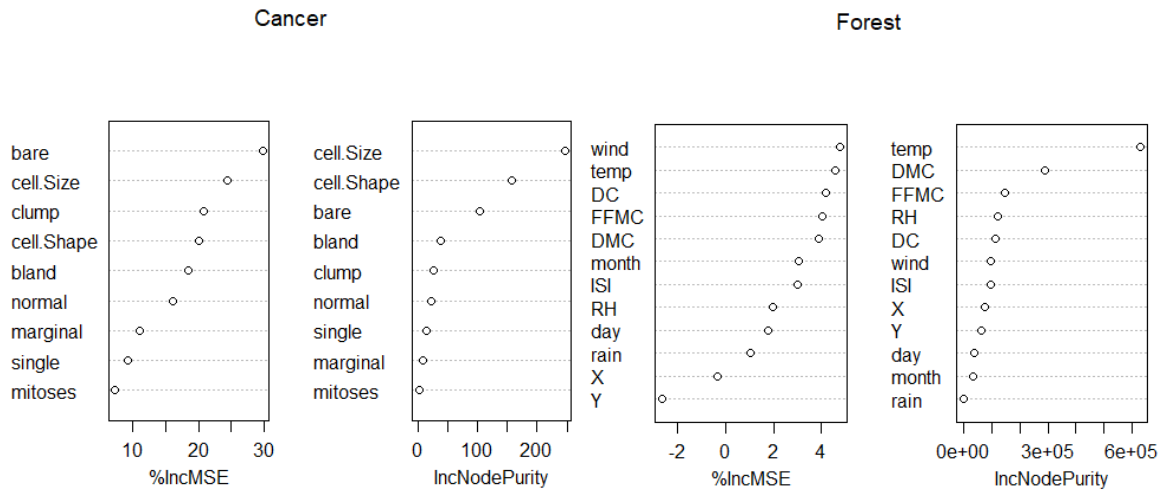

FIGURE 2. The important measure for each variable of Wisconsin Breast Cancer Database Dataset and Forest Fire Dataset according to *%IncMSE* and *IncNodePurity*

Figure 2 shows the variables sorted decreasingly by the two important measures *%IncMSE* and *IncNodePurity* for Wisconsin Breast Cancer Database Dataset and Forest Fire Dataset. Both of these measures are as assigned by the RF. The ranking for Wisconsin Breast Cancer Database Dataset according to *%IncMSE* is the following: The Bare Nuclei (bare) is the most important variable followed by Uniformity of Cell Size (cell.Size), Uniformity of Cell Shape (cell.Shape) and Clump Thickness (clump). Following Bland Chromatin (bland), Normal Nucleoli (normal), and Marginal Adhesion (marginal), parameter tuning of functions uses a grid search. We will use this rank based on *%IncMSE* to improve predictive performance. An overview could be seen in [14]. Figure 2 also describes important features for Forest Fire Dataset according to *%IncMSE* and *IncNodePurity*. The most important variable of this dataset is wind speed in km/h: 0.40 to

9.40 (wind) followed by temperature in Celsius degrees: 2.2 to 33.30 (temp), DC index from the FWI system: 7.9 to 860.6 (DC) and FFMC index from the FWI system: 18.7 to 96.20 (FFMC), next, DMC index from the FWI system: 1.1 to 291.3 (DMC) and month of the year: "Jan." to "Dec." 1 to 12 (month).

Figure 3 explains the important measure for each variable of Wine Quality Dataset and Bike Sharing Dataset according to *%IncMSE* and *IncNodePurity*. The most important feature of Wine Quality Dataset is (volatile.acidity) followed by (alcohol), (free.sulfur.dioxide), (pH), (residual.sugar) and (chlorides). Figure 3 also shows important features Bike Sharing Dataset based on *%IncMSE* ranking. The most important feature is count of registered users (registered) followed by year (0: 2011, 1: 2012) (yr), count of casual users (casual), hour (0 to 23) (hr), working day: if day is neither weekend nor holiday is 1, otherwise is 0. The feature of atemp is the normalized feeling temperature in Celsius. The values are divided to 50 (max) (atemp) and day of the week (weekday).

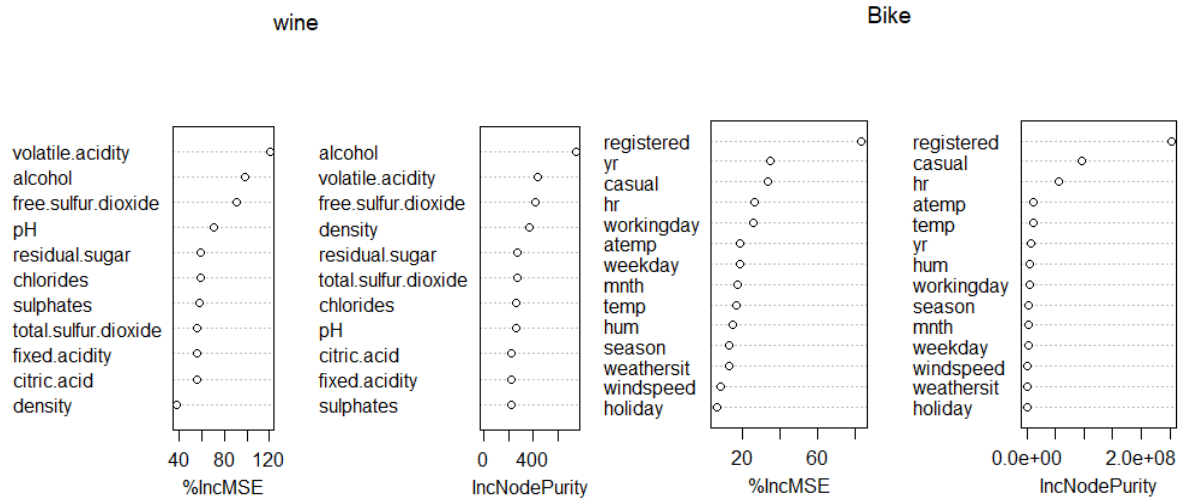

FIGURE 3. The important measure for each variable of Wine Quality Dataset and Bike Sharing Dataset according to *%IncMSE* and *IncNodePurity*

3.2. **Experiment result.** A series of experiments has been conducted. After selecting the important features with RF, we use these selected features for the SVM model. The number of trees we have worked with in the use of RF is 500. The kernel function is RBF kernel with SVM algorithm parameters type is eps-regression. We use cost value 0.001, 0.01, 0.1, 1, 5, 10, 50 and simulate results under gamma = 0.0001, 0.001, 0.01, 0.1, 1, 5, 10 for each dataset. Our model uses *tune.svm()* to know which parameter values for gamma and cost are the best and use for tuning. The tuning of the SVM model can be performed as the technique provides flexibility with respect to maximum error and penalty cost. Tuning the model is extremely important as it optimizes the parameters for the best prediction. Evaluation result for each experiment with different dataset could be seen in Tables 2, 3, 4, 5, and Figure 4.

Table 2 describes the evaluation of the regression method Wisconsin Cancer Dataset. The total number of features of this dataset is 10 and it has 699 instances. Furthermore, as a result, our proposed model has a better value of RMSE and r compared to other methods. The combination of RF, SVM, and tune SVM regression could minimize the RMSE value from 0.2785093 to 0.2310327 and the r value increases from 0.9076668 to 0.9363616 with 6 features.

Table 2. Evaluation of regression method in Wisconsin Cancer Dataset

Method	RMSE	r	Features
SVM	0.2785093	0.9076668	10
RF+SVM	0.2613756	0.9179706	6
RF+SVM+tune SVM Regression	**0.2310327**	**0.9363616**	**6**

Table 3. Evaluation of regression method in Forest Fire Dataset

Method	RMSE	r	Features
SVM	30.0832	0.01147437	12
RF+SVM	29.87215	0.017500154	6
RF+SVM+tune SVM Regression	**20.42034**	**0.6706681**	**6**

Table 4. Evaluation of regression method in Wine Quality Dataset

Method	RMSE	r	Features
SVM	0.6104656	0.5343693	11
RF+SVM	0.6049379	0.5403586	6
RF+SVM+tune SVM Regression	**0.4909776**	**0.6974942**	**6**

Table 5. Evaluation of regression method in Bike Sharing Dataset

Method	RMSE	r	Features
SVM	10.84292	0.9967097	15
RF+SVM	8.74268	0.997732	7
RF+SVM+tune SVM Regression	**4.813466**	**0.999314**	**7**

The next experiment uses the Forest Fire Dataset, and the result could be seen in Table 3. Moreover, our proposed model has a better effect than the other methods. RMSE value decreases and r value increases with 6 features.

Table 4 describes the evaluation of the regression method of the Wine Quality Dataset. The best RMSE value is 0.4909776 and r value is 0.6974942 with 6 features.

Evaluation of the regression method Bike Sharing Dataset could be found in Table 5. As a result, our proposed model has the highest r value 0.999314 and the smallest RMSE value 4.813466 with 7 features compared than other methods.

In general, a lower RMSE is better than a higher one. In the other hand for r value, the higher is better. If the r value 1, it is indicated that the regression predictions perfectly fit the data. The finding of this study clearly shows that the trend of RMSE value is decreasing and r value increases in every experiment for all datasets. We could see this result in Figure 4. Based on our evaluation result, our proposed model has a better result compared to other methods in each dataset. We conclude that the RF method is robust to select the important features and the performance of SVM method will be powerful in small size of data. We could see in all experiments that when we use limited features, we could minimize RMSE value and maximize r value.

4. **Conclusions.** Based on the evaluation of regression method on Tables 2, 3, 4, and 5, our proposed model, the combination of RF, SVM, and tune SVM regression can reduce the RMSE value and increase the r value. For instance, in Table 4 we could see RMSE value for the Wine Quality Dataset decreases from 0.6104656 to 0.4909776 and r value

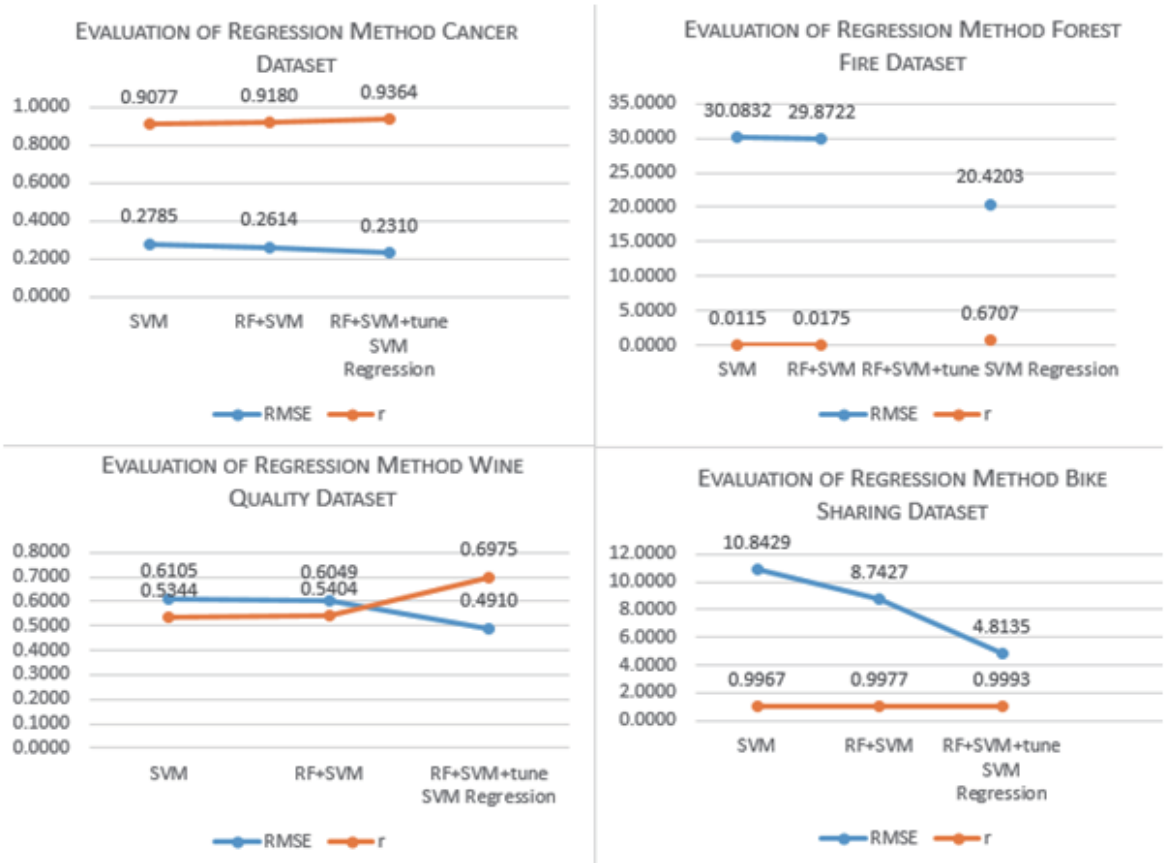

FIGURE 4. Evaluation of the regression method

rises from 0.5343693 to 0.6974942 with 6 features. From Figure 4 we could conclude that the trend of RMSE value decreases and r value increases in every experiment for all datasets. It indicates that the regression predictions perfectly fit the data. Moreover, we use the Random Forest for important feature selection. The important measure for each variable of each dataset according to *%IncMSE* and *IncNodePurity* could be seen in Figure 2 and Figure 3. In this work, we prove that it is essential to select the important features to improve the performance of the model. With the limited features, we could have a good value of RMSE and r. Furthermore, SVM has some parameters that have to be tuned to gain better performance. The use of *tune.svm()* method effectively makes the performance better than other methods and this method is useful for the small size of data. For example, this tuning method could reduce the RMSE from 8.74268 to 4.813466 in Table 5. Tuning the model is extremely important as it optimizes the parameters for the best prediction. The simulation experimental results prove the feasibility and accuracy of the proposed algorithm. The smart combination prediction algorithm is a significant issue nowadays, and this combination algorithm proposed in the paper can also be applied to other fields. The comparison of a more different model, kernel, method, dataset and affection factors can be considered in the future work.

Acknowledgment. This paper is supported by the Ministry of Science and Technology, Taiwan. The Nos are MOST-107-2221-E-324-018-MY2 and MOST-106-2218-E-324-002, Taiwan.

REFERENCES

[1] J. K. Jaiswal and R. Samikannu, Application of random forest algorithm on feature subset selection and classification and regression, *Proc. of World Congress on Computing and Communication Technologies (WCCCT)*, Tiruchirappalli, India, pp.65-68, 2017.

[2] A. L. Blum and P. Langley, Selection of relevant features and examples in machine learning, *Artificial Intelligentce*, vol.97, nos.1-2, pp.245-271, 1997.

[3] R. Bekkerman, R. El-Yaniv, N. Tishby and Y. Winter, Distributional word clusters vs. words for text categorization, *JMLR*, vol.3, pp.1183-1208, 2003.

[4] R. Caruana and V. R. de-Sa, Benefitting from the variables that variable selection discards, *JMLR*, vol.3, pp.1245-1264, 2003.

[5] R. Shapire, Y. Freund, P. Bartlett and W. Lee, Boosting the margin: A new explanation for the effectiveness of voting methods, *Annals of Statistics*, vol.26, no.5, pp.1651-1686, 1998.

[6] B. Wang and J. Pineau, Online bagging and boosting for imbalanced data streams, *IEEE Trans. Knowledge and Data Engineering*, vol.3, no.12, pp.3353-3366, 2016.

[7] R Development Core Team, R: A language and environment for statistical computing, *The R Foundation for Statistical Computing*, Vienna, Austria, http://www.R-project.org, 2008.

[8] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[9] K. B. Jyothi, K. H. Bindu and D. Suryanarayana, A comparative study of random forest & k-nearest neighbors on the HAR dataset using Caret, *International Journal of Innovative Research in Technology*, vol.3, no.9, pp.6-9, 2017.

[10] U. Grömping, Variable importance assessment in regression: Linear regression versus random forest, *Amer. Statistician*, vol.63, no.4, pp.308-319, 2009.

[11] H. Ishwaran, Variable importance in binary regression trees and forests, *Electr. J. Stats*, no.1, pp.519-537, 2007.

[12] C. Strobl, A. Boulesteix, T. Kneib, T. Augustin and A. Zeileis, Conditional variable importance for random forests, *BMC Bioinf.*, vol.9, p.307, 2008.

[13] C. Strobl, A. Boulesteix, A. Zeileis and T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources, and a solution, *BMC Bioinf.*, vol.8, p.25, 2007.

[14] C. González, J. McWilliams and I. Juárez, Important variable assessment and electricity price forecasting based on regression tree models: Classification and regression trees, Bagging and Random Forests, *IET Generation, Transmission & Distribution*, vol.9, no.11, pp.1120-1128, 2015.

[15] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, U.K., 2000.

[16] P. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, 2002.

[17] G. Xie, S. Wang and K. Lai, Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches, *Journal of Air Transport Management*, vol.37, pp.20-26, 2014.

[18] R. C. Chen and C. H. Hsieh, Web page classification based on a support vector machine using a weighted vote schema, *Expert Systems with Applications*, vol.31, no.2, pp.427-435, 2006.

[19] R. C. Chen, K. F. Cheng and C. F. Hsieh, Using rough set and support vector machine for network intrusion detection, *International Journal of Network Security & Its Applications (IJNSA)*, vol.1, no.1, pp.1-13, 2009.

[20] A. J. Smola and B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.*, vol.14, no.3, pp.199-222, 2004.

[21] F. Bellocchio and F. Piuri, Hierarchical approach for multiscale support vector regression, *IEEE Trans. Neural Networks and Learning Systems*, vol.23, no.9, pp.1448-1460, 2012.

[22] O. L. Mangasarian and W. H. Wolberg, Cancer diagnosis via linear programming, *SIAM News*, vol.23, no.5, pp.1-18, 1990.

[23] L. Breimann, Bagging predictors, *Machine Learning*, vol.24, pp.123-140, 1996.

[24] L. Breimann, Random forests, *Machine Learning*, vol.45, pp.5-32, 2001.

[25] C. Luo, C. Huang, J. Cao, J. Lu, W. Huang, J. Guo and Y. Wei, Short-term traffic flow prediction based on least square support vector machine with hybrid optimization algorithm, *Neural Processing Letters*, pp.1-18, https://doi.org/10.1007/s11063-019-09994-8, 2019.

[26] Y. Mei, F. Hong, Z. Jia and Z. Kang, Debris flow forecasting of northwest of Yunnan province based on LR, SVM, and RF statistical models, *Proc. of the 26th International Conference on Geoinformatics*, 2018.

[27] L. Lan, Z. Wang, S. Zhe, W. Cheng, J. Wang and K. Zhang, Scaling up kernel SVM on limited resources: A low-rank linearization approach, *IEEE Trans. Neural Networks and Learning Systems*, vol.30, no.2, 2019.

[28] K. Kobayashi and F. Komaki, Information criteria for support vector machines, *IEEE Trans. Neural Networks*, vol.17, no.3, 2006.

[29] C. Li, H. Zhang, H. Zhang and Y. Liu, Short-term traffic flow prediction algorithm by support vector regression based on artificial bee colony optimization, *ICIC Express Letters*, vol.13, no.6, pp.475-482, 2019.

[30] A. Sharma, Y. Lee and W. Chung, High accuracy human activity monitoring using neural network, *Proc. of the 3rd International Conference on Convergence and Hybrid Information Technology*, pp.430-435, 2008.

[31] S. Shamshirband, D. Petkovi' and H. Javidnia, Sensor data fusion by support vector regression methodology – A comparative study, *IEEE Sensors Journal*, vol.15, no.2, pp.850-854, 2015.

[32] S. Kavitha, S. Varuna and A. Ramya, A comparative analysis on linear regression and support vector regression, *Proc. of 2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, Coimbatore, India, 2016.

[33] W. H. Wolberg and O. L. Mangasarian, Multisurface method of pattern separation for medical diagnosis applied to breast cytology, *Proc. of the National Academy of Sciences of the United States of America*, vol.87, pp.9193-9196, 1990.

[34] P. Cortez and A. Morais, A data mining approach to predict forest fires using meteorological data, *Proc. of the 13th EPIA 2007 – Portuguese Conference on Artificial Intelligence*, Guimaraes, Portugal, pp.512-523, http://www. dsi.uminho.pt/∼pcortez/fires.pdf, 2007.

[35] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis, Modeling wine preferences by data mining from physicochemical properties, *Proc. of Decision Support Systems*, vol.47, no.4, pp.547-553. 2009.

[36] H. Fanaee-T and J. Gama, Event labeling combining ensemble detectors and background knowledge, *Progress in Artificial Intelligence*, pp.113-127, 2013.