International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

A Comparison between Random Forest and Mixed Effects Random Forest to Predict Students' Math Performance in Indonesia

Retno Mayapada[a]*, Budi Susetyo[b], Bagus Sartono[c], Rahmawati[d]

[a,b,c]*Department of Statistics, IPB University, Jl. Raya Dramaga, 16680 Bogor, Indonesia*

[d]*Center for Assessment and Learning, Ministry of Education and Culture, 10710 Jakarta, Indonesia*

[a]*Email: retno.mayapada@gmail.com,*[b]*Email: buset008@yahoo.com,*[c]*Email: bagusco@gmail.com,*[d] *Email: rahmawati@kemdikbud.go.id*

Abstract

The Center for Assessment and Learning in Indonesia developed a national assessment system called the Indonesian Student Competency Assessment (AKSI/Asesesmen Kompetensi Siswa Indonesia) to measure the competence of school graduates. The results of AKSI 2019 showed that Indonesian students had poor math performance. However, the factors that affect student scores are not only caused by individual students but can also be influenced by the quality of the school. It indicates that the data has a hierarchical structure. Thus, this study aims to compare the Mixed effects random forest (MERF) method with random forest (RF) to predict students' math performance on AKSI 2019. The data set in this study was divided into two sets, train (85% of the original data set) and test set (15% of the original data set). Furthermore, the MERF model was built using training data by estimating the fixed component (fixed part) and the random part effects on the model, while the RF ignores the random effects on the model. Then, estimation of the value of the response variable on the test data was carried out using the MERF model that has been built. These steps are repeated 10 times to give accurate results. The Model then evaluated by calculating the mean square error prediction (MSEP), root mean square error prediction (RMSEP), and mean absolute error prediction (MAEP) from 10 models. The results showed that the MERF model produced higher prediction accuracy than the RF model.

---

* Corresponding author.

This means that school variables have an effect in predicting student performance, which can increase the accuracy of predictions if the effects are included in the model.

Keywords: Hierarchical structure; Mixed effects random forest; Random forest; Prediction accuracy.

1. Introduction

It is widely believed that the process of education is responsible for the quality of the human resources of a country. Therefore, every country, including Indonesia, consistently takes effort to some extent to improve the education process especially the formal one through schools, colleges, and universities. To be able to monitor the progress, a proper assessment should be conducted in order to have appropriate information about the advancement. Indonesia participates in international assessment programs to assess students' performance, such as Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA) regularly. However, there is still a need for a national assessment system that measures student competence in a national context. Related to this, the Center for Assessment and Learning (Pusmenjar/Pusat Asesmen dan Pembelajaran) in Indonesia has developed a national assessment system called the Indonesian Student Competency Assessment (AKSI). The author in [1] reported on its official website that 79.44% of students surveyed on AKSI 2019 had less than satisfactory on mathematical literacy, while in reading and science literacy skills, the percentage of students with less than satisfactory category 55.85% and 66.11% respectively. These finding showed that students' mathematical literacy skills were the lowest of the three cognitive skills tested. Therefore, in this study, researchers will focus on predicting students' math performance.However, the factors that influence student scores are not only caused by individual students but can also be influenced by the quality of the school which indicates that there is a hierarchical data structure that is students nested in schools. The observations at the lowest level nested in the same group are not completely independent in the hierarchical data structure which indicates that the independent assumptions required in the statistical test are not fulfilled, so that multilevel modeling is needed which can address these issues of hierarchical data structure. Several studies have applied multilevel modeling in identifying factors that influence a response variable, such as in the research conducted by author in [2]. Author in [2] compared four mixed effect models (also called hierarchical linear modeling (HLM) or multilevel models), including the linear mixed effects model (LMM) [3] Bayesian LMM, Generalized LMMs with lasso (glmmLasso) [4], and mixed effects random forest (MERF) [5,6]. Author in [2] showed that MERF provided the smallest mean squared error (MSE) value and the greatest accuracy. Author in [7] also applied the MERF and mixed effects support vector regression (MESVR) methods to data on crime cases in New York. Research by author in [7] showed that MERF was better than MESVR in estimating the number of crime cases in New York. Based on this, this study aims to compare the Mixed effects random forest (MERF) method with random forest (RF) to predict students' math performance on AKSI 2019.

2. Data and Methodology

2.1. Data

This study used data from the AKSI 2019 survey conducted by the Pusmenjar, Agency of Research and

Development and Book, Ministry of Education and Culture of the Republic of Indonesia, for grade 9 students in all provinces in Indonesia. The sampling method used was multistage random sampling by randomly selecting districts/cities and then selecting schools randomly from these districts/cities. Finally, from the selected schools, students were selected randomly. The number of observations used in this study was 16,651 students from 1,804 schools in Indonesia. The results of the survey consisted of students' cognitive and noncognitive assessment scores. Cognitive assessment scores are mathematics scores and students' noncognitive assessment scores used to measure students' abilities in financial literacy. In addition to the AKSI survey data conducted on students, this study also used school accreditation data obtained from the National Accreditation Agency for School/Madrasah (BAN-S/M) for variables at the school level. This study computed the effect of twelve predictors on predicting the students' mathematics scores (Y) on AKSI 2019. The predictors were gender (X1), mother's education level (X2), father's education level (X3), books about financial lessons (X4), tasks and activities about finance (X5), financial information (X6), discussion about finance (X7), attitude before buying goods (X8), ownership of financial tools (X9), financial management (X10), and financial action (X11). Besides that, the school accreditation ranking (Z1) is also entered into the model for additional information at the school level.

2.2. Methodology

2.2.1. Data Preparation Steps

The steps of data preparation are as follows.

1. Collecting data obtained from the 2019 AKSI survey and school accreditation data.
2. Combining AKSI 2019 survey data and school accreditation data and removing observations that contain missing values after the data is combined.
3. Exploring the data to get an overview of the variables to be analyzed.
4. Identify the variables that will be used in the research.
5. Dividing the research data clusters into training data and testing data with the proportion of training data and data testing were 85% and 15% which were randomly selected 10 times.

2.2.2. Data Analysis Steps

Random forest was proposed by [8] to perform classification and clustering based on an ensemble of decision trees. Random forest is a development of the classification and regression trees (CART) method by applying the bagging method and random feature selection. Each tree in the ensemble is built on the principle of recursive partitioning, the data of predictors are recursively divided so that each data set from the splitting can be observed with similar response variable values [9]. Author in [5] developed a random forest method for clustered/hierarchical data as the mixed effects random forest (MERF). This method is based on the expectation maximization (EM) algorithm which is used to estimate the random component and the random forest method to predict the value of the response variables using a fixed effect. Analysis of the data in this study using the MERF method.

The steps of data analysis are as follows.

1. In the training data, MERF modeling was carried out using 14,093 students' observations obtained from 85% of the original data clusters. The MERF model is defined as follows [5].

$$\boldsymbol{y_j} = \boldsymbol{f}(\boldsymbol{X_j}) + \boldsymbol{Z_j b_j} + \boldsymbol{\epsilon_j}, \qquad j = 1,2,\ldots,1.803$$

where $\boldsymbol{y_j} = \left[y_{\boldsymbol{j1}}, y_{\boldsymbol{j2}}, \ldots, y_{\boldsymbol{jn_j}}\right]^T$ is the math scores on the AKSI survey, $\boldsymbol{X_j}$ is the matrix of fixed-effects covariates, and $\boldsymbol{Z_j}$ is the school accreditation rank as random-effect covariate. The MERF and RF modeling steps are as follows.

a. Determine the number of iterations used in the MERF and RF algorithm. In this study, the number of iterations tried was 250, $r = 1,2,\ldots,250$.
b. Set initial values for $\widehat{\boldsymbol{b}}_{\boldsymbol{j(0)}} = \boldsymbol{0}, \hat{\sigma}_{(0)} = 1$, and $\widehat{\boldsymbol{D}}_{(\boldsymbol{0})} = \boldsymbol{I_q}$. $\boldsymbol{b_j}$ is the unknown vector of random effects for the school j with covariance matrix $\boldsymbol{D}$.
c. Calculate the value of the response variable $\boldsymbol{y}^*_{\boldsymbol{j(r)}}$ as $\boldsymbol{y_j} - \boldsymbol{Z_j}\widehat{\boldsymbol{b}}_{\boldsymbol{j(r-1)}}$
d. Estimates $\boldsymbol{f}(\boldsymbol{X_j})$ which are fixed components of the model using the random forest method. The number of trees used was 100 trees. Each tree generated by the model is constructed by taking a bootstrap example from the dataset. At this step, an RF model is generated. The next steps were taken to obtain the MERF model.
e. Estimate the random effect (($\widehat{\boldsymbol{b}}_{\boldsymbol{j}}$) using the following equation.

$$\widehat{\boldsymbol{b}}_{\boldsymbol{j(r)}} = \widehat{\boldsymbol{D}}_{(\boldsymbol{r-1})}\boldsymbol{Z}_{\boldsymbol{j}}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}_{\boldsymbol{j(r-1)}}\left(\boldsymbol{y_j} - \hat{f}(\boldsymbol{X_j})_{(r)}\right)$$

where $\widehat{\boldsymbol{V}}^{-\boldsymbol{1}}_{\boldsymbol{j(r-1)}} = \boldsymbol{Z_j}\widehat{\boldsymbol{D}}_{(\boldsymbol{r-1})}\boldsymbol{Z}_{\boldsymbol{j}}^{\boldsymbol{T}} + \hat{\sigma}^2_{(r-1)}\boldsymbol{I}_{\boldsymbol{n_j}}$

f. Finally, calculate the variance $\hat{\sigma}^2$ and matrix $\widehat{\boldsymbol{D}}$ using the following equation.

$$\hat{\sigma}^2_{(r)} = N^{-1}\sum_{i=1}^{\boldsymbol{n}}\left\{\hat{\boldsymbol{\epsilon}}^{\boldsymbol{T}}_{\boldsymbol{j(r)}}\hat{\boldsymbol{\epsilon}}_{j(r)} + \hat{\sigma}^2_{(r-1)}\left[\boldsymbol{n_j} - \hat{\sigma}^2_{(r-1)}trace(\widehat{\boldsymbol{V}}_{\boldsymbol{j(r-1)}})\right]\right\}$$

$$\widehat{\boldsymbol{D}} = n^{-1}\sum_{i=1}^{\boldsymbol{n}}\left\{\widehat{\boldsymbol{b}}_{j(r)}\widehat{\boldsymbol{b}}^{\boldsymbol{T}}_{\boldsymbol{i(r)}} + \widehat{\boldsymbol{D}}_{(\boldsymbol{r-1})} - \widehat{\boldsymbol{D}}_{(\boldsymbol{r-1})}\widehat{\boldsymbol{D}}_{(\boldsymbol{r-1})}\right\}$$

where $\hat{\boldsymbol{\epsilon}}_{j(r)} = \boldsymbol{y_j} - \hat{f}(\boldsymbol{X_j})_{(r)} - \boldsymbol{Z_j}\widehat{\boldsymbol{b}}_{\boldsymbol{j(r)}}$

2. Estimate the responses of testing data using the MERF and RF model. The testing data consisted of 2,488 student observations.
3. Evaluate the MERF and RF model by calculating the mean square error prediction (MSEP), root mean

square error prediction (RMSEP) and mean absolute error prediction (MAEP) value of the response estimator results on the test data.

3. Results and Discussion

3.1. Response Variable Data Exploration

Student math scores on AKSI 2019 are presented in a boxplot in Figure 1, which shows there are some outliers in the math scores data, but the data tends to be symmetrical with the median line in the middle of the box. Furthermore, based on Figure 1, it is found that 25% of students get mathematics scores lower than 35.27 and 25% of students get mathematics scores higher than 43.8. The highest score obtained by the students was 73.51 while the lowest score was 0. Overall, the students' average score of mathematics was 39.459.

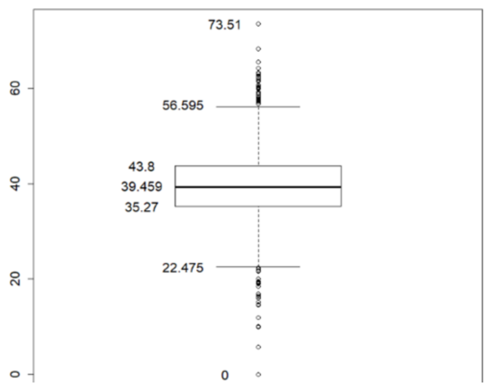

Figure 1: Mathematics Scores on AKSI 2019

The results of the AKSI on low math scores indicate that Indonesian students have not been capable to implement the application of mathematics on the various problems they encounter in real life. These results are also confirmed by the results of international assessments such as PISA. Indonesia participated in PISA for the first time in 2001. From 2001 to 2018, the achievement of Indonesian students in mathematics has a hump-shaped curve. The mathematical abilities fluctuated in the first years of Indonesia participating in PISA. Meanwhile, since 2009 it has been relatively stable. Indonesia's PISA 2018 ranking has decreased when compared to the PISA 2015 results. Indonesia is ranked 7th from the bottom (73) of the student achievement in mathematics with an average score of 379 [10,11].

Table 1: Students' math performance by their school accreditation rating

School Accreditation Rating	Average score of mathematics on AKSI 2019
A	40.6
B	37.6
C	36.2
Non-accredited	33.8

Table 1 shows the average math score of students in AKSI 2019 according to the school accreditation ranking. Students studying in schools with an "A" accreditation rating have significantly higher math scores than students in schools with a lower accreditation rating. Students who attend schools that are not accredited have an average score of 6.8 points lower than students who attend schools with the highest quality. These findings indicate that there is an effect of school accreditation on students' AKSI Math Score.

3.2. Comparison of the Predicted Results

This study uses the *merf* package in Python Software. The algorithm implemented in this package was developed by the author in [5]. The comparison of the prediction results on the first simulation train and test data using the MERF and RF models is shown in Figure 2 and Figure 3.

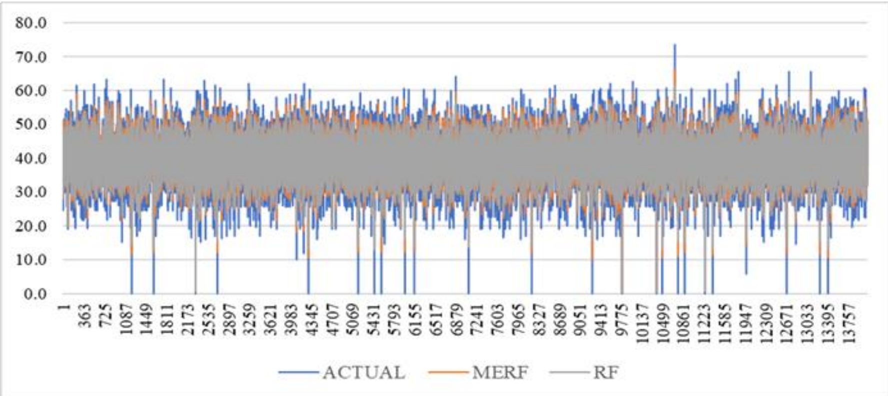

Figure 2: Comparing prediction results of MERF and RF model on train data

The prediction results by the MERF model look similar to the actual data compared to the prediction results by the RF model. However, for the outliers, both the MERF and RF models could not predict the outliers well. In train data, the highest math score of 73.51 is predicted by MERF of 66.14, while RF predicts that value of 61.6. While the value of 0 is predicted by MERF from 1.02 to 15.30 while RF predicts a value of 0 is 0.59 to 16.50.

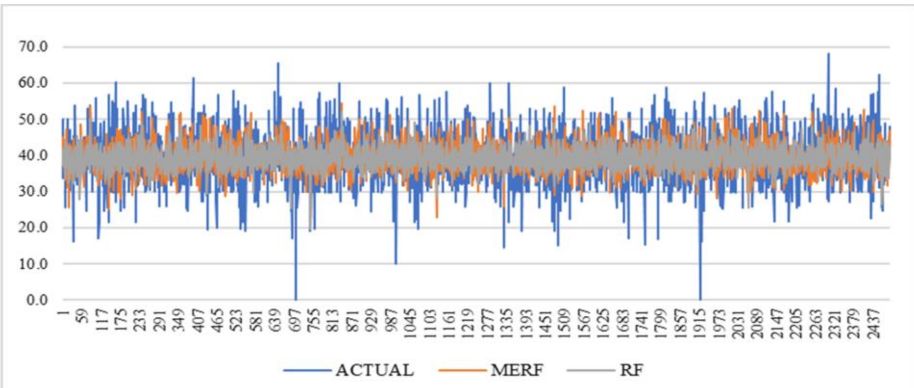

Figure 3: Comparing prediction results of MERF and RF model on test data

3.3. Evaluation Model

The evaluation of the model in this study was carried out by calculating the average of MSEP, RMSEP, and MAEP values based on the test data of 10 models. The performance of the MERF model is then compared to the RF model which is a random forest model without any random effects on it.

Table 2: Model evaluation results

Simulation	MSEP		RMSEP		MAEP	
	MERF	RF	MERF	RF	MERF	RF
Model 1	40.841	43.880	6.391	6.624	4.881	5.080
Model 2	40.864	46.300	6.393	6.804	4.879	5.212
Model 3	40.847	44.000	6.391	6.633	4.881	5.081
Model 4	40.843	43.585	6.391	6.602	4.884	5.144
Model 5	40.783	48.783	6.386	6.984	4.877	5.289
Model 6	40.893	47.662	6.395	6.904	4.886	5.234
Model 7	40.786	44.444	6.386	6.667	4.879	5.067
Model 8	40.960	45.996	6.400	6.782	4.885	5.246
Model 9	40.786	43.914	6.386	6.627	4.880	5.097
Model 10	40.862	45.578	6.392	6.751	4.885	5.177
Average	40.846	45.414	6.391	6.738	4.882	5.163

Table 2 shows that the MERF model produces the smallest MSEP, RMSEP, and MAEP values for every simulation. Based on these results, the MERF method gives the best performance in estimating students' math scores compared to the RF method which ignores hierarchical data structures. A previous study by the author in [2] found similar results. The author in [2] compared MERF and RF for predicting children’s weight gain and made a plot of true versus predicted values in the RF and MERF models. The plots showed that MERF performs slightly better than RF where MAE is higher for RF with unscaled weight/height as the response. The author in [5] had previously introduced the MERF method by conducting a simulation study and comparing its performance to that of RF. The author's main finding in [5] is that MERF appears to be more suitable for clustered data than a RF, especially when random effects cannot be ignored.

4. Conclusions

This current study discusses the comparison of RF and MERF in predicting the student cognitive ability in mathematics. In this study, we used school as the random effect and the financial literacy and the individual student background as the fixed effects in the MERF model. As compared to RF, the MERF algorithm consistently has lower error. This shows MERF is able to predict students' math scores better than RF with the condition that the data has a hierarchical structure or is collected in multilevel way. This study also shows that school has a positive influence on the accuracy of the predictions of students' math performance.

5. Limitation and Recommendation

This study has potential limitations. This study used the *merf* package in *Python Software*. The algorithm implemented in this package was developed by the author in [5]. The existing random forest framework in this package, however, is constrained by the random forest algorithm's limitation, which allows only one parameter to be set by default. The number of iterations is the parameter. The number of iterations in this study is limited to 250. We did not compare models using different numbers of iterations so that further study could attempt to build a model using a different number of iterations. An in-depth study of the outliers that exist in the data also has not been carried out in this study. The combination of the MERF algorithm and robust methods can be carried out in further research.

References

[1]. Puspendik. "AKSI". Internet: pusmenjar.kemdikbud.go.id/aksi-2/, 2019 [Sep. 07, 2020].

[2]. E. Sheen. "An Exploration of Mixed Effects Models for Analysis of Infant Weight Gain Trajectories." M.S. thesis, Pennsylvania State University, Pennsylvania, 2019.

[3]. D. Bates, M. Mächler, B. Bolker, S. Walker. (2015, Oct). "Fitting Linear Mixed-Effects Models using lme4." *Journal of Statistical Software*. [On-line]. 67(1). Available: www.jstatsoft.org/v67/i01/ [Oct. 18, 2020].

[4]. A. Groll, G. Tutz. (2014, Mar). "Variable selection for generalized linear mixed models by L 1-penalized estimation." *Statistics and Computing*. [On-line]. 24(2), pp. 137-154. Available: link.springer.com/article/10.1007/s11222-012-9359-z [Oct. 18, 2020].

[5]. A. Hajjem. "Mixed Effects Trees and Forests for Clustered Data." Ph.D. dissertation, University of Montreal, Montreal, 2010.

[6]. A. Hajjem, F. Bellavanve, D. Larocque. (Nov, 2012). "Mixed-effects random forest for clustered data." Journal of Statistical Computation and Simulation, [On-line]. 84(6), pp. 1313-1328. Available: https://doi.org/10.1080/00949655.2012.741599.

[7]. A. Fakhrurrozi. "On the Use of Mixed Effects Machine Learning Regression Models to Capture Spatial Patterns: A Case Study on Crime." M.S. thesis, University of Twente, Enschede, 2019.

[8]. L Breiman. (Aug, 2001). "Statistical modeling: The two cultures". Statistical Science, [On-line]. 16(3), pp. 199–231. Available: https://doi.org/10.1214/ss/1009213726.

[9]. C. Strobl, J. Malley, G. Tutz. (Dec, 2009). "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests". Psychol Methods, [On-line]. 14(p), pp. 323–348. Available: https://doi.org/10.1037/a0016973.

[10]. OECD. "Education GPS". Internet: http://gpseducation.oecd.org, n.d. [Mar. 13, 2021].

[11]. F. Avvisati, A. Echazarra, P. Givord, M. Schwabe. "Indonesia - Country Note - PISA 2018 Result". Internet: https://www.oecd.org/pisa/publications/PISA2018_CN_IDN.pdf, 2019 [Mar. 13, 2021].