International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 2
ISSN: 2088-5334

Classification Modelling of Random Forest to Identify the Important Factors in Improving the Quality of Education

Aditya Ramadhan[a,*], Budi Susetyo[a], Indahwati[a]

[a] *Department of Statistics, IPB University, Bogor 16680, Indonesia*
Corresponding author: [*]aditya.ramadhan@kemdikbud.go.id

***Abstract*— National Education Standards (SNP) is the minimum criteria that must be met by the education units and/or educational organizations to realize high-quality national education. The evaluation is implemented through accreditation, and national evaluation of graduate competencies carried out through national examination (UN). Research on the causality relationship between SNP and the UN has been done, but research using classification modelling to explain the relationship between SNP and the UN has never been done. This study employed random forest for multi-class classification to examine important variables in improving the quality of education at the high school level (SMA/MA) based on computer-based national exam (UNBK) scores and accreditation results. The highest classification accuracy and G-Mean value were obtained in multi-class random forest modelling of 88.17% and 48.95% based on the evaluation model. This model generates important factors in the classifying the quality of education by the items of accreditation instruments. Important factors are items 69, 68, 62, 71, 67, 55, 56, 83, 45, 39, 36, 33, 64, 46, and 14. Based on the indicators of important factors, SNP has an important role in classifying the quality of education, which are standards of school facilities (SSP), standards of teacher and education staff (SPT), and standards of graduate competency (SKL). The study results advise region governments and education units to collaborate in improving SSP, SPT, and SKL.**

***Keywords*— National education standards; UNBK; classification modelling; multi-class random forest.**

Manuscript received 22 May 2019; revised 1 Oct. 2020; accepted 7 Nov. 2020. Date of publication 30 Apr. 2021.

I. INTRODUCTION

The quality of education is the degree of conformity between education implementation and the National Education Standards (SNP). SNP was developed by the National Education Standards Board (BSNP) as the minimum criteria that must be met by education units and/or education providers by considering the condition and diversity of Indonesia. SNP consists of the standards of content (SI), standards of the process (SPR), standards of graduate competency (SKL), standards of teacher and education staff (SPT), standards of school facilities (SSP), standards of education management (SPL), standards of funding (SB), and standards of assessment (SPN) [1]. One way to measure SNP achievement is the assessment carried out by the National Accreditation Board (BAN) such as accreditation. BAN developed an instrument consisting of statements to assess the eight SNP based on documents, observations, and field verification. The instrument to conduct accreditation for senior high school (SMA/MA) is appointed by Regulation of the Minister of Education and Culture (Permendikbud) [2].

SNP as the basis for preparing strategies in the development of education quality based on national examinations (UN). UN is an activity to measure the achievement of graduate competencies in certain subjects nationally refer to SKL [3]. SKL is used as the primary reference for SI, SPR, SPN, SPT, SSP, SPL, and SB development. Since 2015, the implementation of the UN in Indonesia has been carried out in 2 types, the Paper and Pencil Based National Exam (UNKP) and the Computer-Based National Exam (UNBK). The implementation of UNBK aims to improve efficiency, quality, reliability, credibility, and integrity of the test.

Based on the description above, accreditation and the UN have the same reference (SNP), so the implementation of both must be in line because those are standard-based quality assurance programs. Several researchers have researched causality between SNP and UNBK. At SMA/MA, used generalized structured component analysis (GSCA), SPN and SPR have a significant effect on SKL [4]. For SMK, the PLS-path modelling (PLS-PM) method, SB and SPR have a significant effect on SKL [5]. And then, for SMP/MTs with

the GSCA method, SKL, SPN, and SPR have a significant effect on UNBK [6].

Another form of analysis that can be used to explain the relationship between eight SNP and the UN is the classification modelling of random forest. Classification is categorizing a new group of observations into a set of categories (classes). Random forests algorithm, which was introduced by Breiman [7], is a general term for ensemble methods using tree-type classifiers. Random forests are appropriate for high-dimensional data, overcome over-fitting problems, produce a high prediction accuracy, interpretable and non-parametric for various types of datasets [8].

In this study, the random forest is conducted on the response variable of UNBK, where it values were categorized into four categories ("Very Good", "Good", "Enough", and "Less"). The UNBK categorization causes data imbalance in each category/class, where the amount of data in a category is greater than the amount of data in other categories. Therefore, the handling of unbalanced data needs to be addressed to minimize misclassification. Imbalanced data in this study was handled by class weights and Synthetic Minority Oversampling Technique (SMOTE).

According to this background, this study aims to identify important factors influencing the quality of education at SMA/MA based on UNBK and accreditation results (items of accreditation instruments) by applying the classification modelling of multi-class random forest. This research is expected to provide recommendations to the government regarding policies to improve the quality of education at SMA/MA based on UNBK and accreditation results.

II. Material and Method

A. Imbalanced Dataset

The result of UNBK in 2018 schools in the category of "good" amounted to 89.1%, "enough" 8.7%, and "less" 2.2%. The amount of data in the category of "good" has a huge number of instances compared to other classes, and it is called imbalanced data. In classification modelling, algorithms used tend not to pay attention to data imbalance so that it is inadequate if there are cases of imbalanced data [9]. As a result of this condition, the minority class will experience misclassification. Therefore, handling of imbalanced data needs to be addressed to minimize misclassification. Imbalanced data in this study was handled by class weights and Synthetic Minority Oversampling Technique (SMOTE).

Class weights, in which the majority class is given less weight than minority class (minority class weights are 1) [10]. This class weight gives the effect of observing the minority class increasing and becoming balanced with the majority class. SMOTE, in which the minority class is over-sampled by creating "synthetic" examples based on k-nearest neighbours' concept rather than by over-sampling with replacement [11]. The implementation that currently used is five nearest neighbours. The Value Difference Metric (VDM) was introduced to provide an appropriate distance function for nominal attributes [12].

The VDM defines the distance between two values X and Y of an attribute *a* as [12]:

$$vdm_a(X,Y) = \sum_{i=1}^{p} \delta(x_i, y_i)^r \quad (1)$$

where

p : the number of predictor variable;
r : constant, 1 for Manhattan distance or 2 for Euclidean distance; and
$\delta(x_i, y_i)$: The distance between categories as:

$$\delta(U_1, U_2) = \sum_{j=1}^{c} \left| \frac{C_{1j}}{c_i} - \frac{C_{2j}}{c_2} \right| \quad (2)$$

with
$\delta(U_1, U_2)$: the distance between two values U_1 and U_2;
C_{1j} : the number of U_1 for attribute *j*;
C_{2j} : the number of U_2 for attribute *j*;
C_1 : the number of categories at U_1;
C_2 : the number of categories for U_2;
j : the number of classes, *j* = 1, 2, 3; and
c : the number of categories.

The procedure of synthetic samples for categorical variables as [12]:

1) Calculate the distance between observations in a minority class.
2) Determine the value of *k* (*k*=5) and the percentage of oversampling.
3) Select a random sample from minority class.
4) Determine the observation of the nearest neighbour with all observations in the minority class.
5) Create new set feature values to generate new minority class feature vectors by the majority vote of the feature vector in consideration and its k nearest neighbours.
6) Repeat steps 3 to 5 until the desired amount of oversampling is reached.

B. CART

Classification and Regression Tree (CART) is a classification method with a non-parametric statistical approach [7]. If the response variable is categorical, CART produces a classification tree. Furthermore, if the response variable is numerical, CART produces a regression tree. In this study, the predictor variable is on the ordinal scale, so there are *v*-1 possibilities for splitting. The value of impurity is used to select the best split from each predictor variable. The value of impurity measures the heterogeneity of a node. The Gini index is the value used to define the size of the impurity function. The Gini index in node *r*, as [7]:

$$i(r) = 1 - \sum_{j=1}^{3} p^2(j|r) \quad (3)$$

When:

$$p(j|r) = \frac{N_j(r)}{N(r)} \quad (4)$$

where, $p(j|r)$ is the relative proportion of class *j* cases in node *r*. $N_j(r)$ is the number of class *j* cases in node *r*, and $N_j(r)$ the total number of cases in node *r*.

The decrease in heterogeneity is also called Gini Gain values (reduced impurity). The variables of each split always maximize the decrease of heterogeneity in response values, and the formula of Gini Gain is as follows [7]:

$$GiniGain(r) = Gini(r) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} . Gini(r) \quad (5)$$

With *Si* is the number of observations after the partition from *S* (observations on the initial node) caused by attributes *A*. Variable *S* that have the Gini value The biggest gain is the best separator at the *i*-node.

C. Random Forest

Random forest is the CART method's development, namely by applying the bootstrap aggregating (bagging) and random feature selection methods [7]. Random forest builds multiple decision trees and merges them to get a more accurate and stable prediction. Each tree is grown as follows [7]:

1) Suppose that the number of cases in the training data is *N* and the number of predictor variables is *M*.
2) Use bootstrapping (random sampling with replacement) to generate *L* training sets and train one base-learner with each (the training grew the tree).
3) The number of *m* variables (*m*<*M*) is specified at each node, *m* variables are selected at random out of the *M,* and the best split on this *m* is used to split the node. The value of *m* is held constant during the forest growing.
4) Each decision tree was grown without pruning. The prediction results of *L* trees based on the majority vote.

In this study, the response variable has more than two classes (multi-class). The random forest algorithm is a decision tree that can be used to classify data with a multi-class response variable. Binary classifiers can solve the classifying of multi-class response variables. Breaking the multi-class response into a classification of two classes can be done using the One vs One (OVO) and One vs All (OVA) approaches [13].

The OVO approach is to divide the training dataset into several segments so that each segment has a response consisting of two classes. If the number of classes is *J*, the number of OVO segments is *J*(*J*-1)/2. The Class prediction for a case using OVO can use the majority vote method. Also, the determination of class prediction on the OVO approach can use the weighted vote method [14]. Suppose that *Aij* is a binary classification algorithm applied to a training dataset consisting of classes *i* and *j* with *i*, *j* = 1, 2, ..., *J*, then the classifier classes an observation into class *i* with probability p_{ij} and the classifier classes an observation into class *j* with probability 1-p_{ij}. The class predictions by the weighted vote for OVO is defined as [15]:

$$class = \underset{i = 1,2,..,J}{argmax} \sum_{1 \le j \ne i = \le J} p_{ij} \quad (6)$$

The OVA approach duplicates as many data groups as the classes in the response variable. The number of classes is J. The probability p_{ij} is the probability of class classifying an observation into class *i* and another class *j*. The class predictions for training data using the OVA approach are classes with the largest of probability values (p_{ij}) [16].

The variable importance in a random forest is used can be obtained by calculating the Mean Decrease Gini (MDG). MDG is the average value of reduced Gini Gain (impurity) that occurs during the sorting process in the formation of a single tree. The importance of predictor variables is used to show the order of the importance of SNP components (items of accreditation instruments) that influence classifying the quality of education. The MDG formula is as follows [17]:

$$MDG = \frac{1}{m} \sum_{i=1}^{r} [\Delta i(s,r) I(s,r)] \quad (7)$$

where,

$\Delta i(s,r)$ = Gini Gain (S)

$I(s,r)$: the indicator function that is 1 if *Xs* is used to split at node *r* and 0 otherwise.

m : the number of trees formed

D. Performance Measure

The confusion matrix is used to measure performance for evaluating the classification model by classification accuracy and G-Mean (geometry mean). The confusion matrix is a classification table obtained from the number of the accuracy of the predicted results with the actual data on each observation in the testing data. Accuracy is the ratio of the number of correctly classified instances (predictions according to actual) to the total number of tested instances. G-Mean is a measure of performance independently considering a classifier's performance on each of the classes [13]. The formula to get accuracy and G-Mean is as follows [13]:

TABLE I
CONFUSION MATRIX

Predicted (i)	Actual (j) 1	2	3	Total
1	h_{11}	h_{21}	h_{31}	$h_{.1}$
2	h_{12}	h_{22}	h_{32}	$h_{.2}$
3	h_{13}	h_{23}	h_{33}	$h_{.3}$
Total	$h_{1.}$	$h_{2.}$	$h_{3.}$	$h_{..}$

$$accuracy = \frac{h_{11}+h_{22}+h_{33}}{h_{..}} \quad (8)$$

$$G - Mean = \left(\prod_{i=1}^{3} \frac{h_{1ii}}{\sum_{j=1}^{3} h_{ji}} \right)^{1/3} \quad (9)$$

E. Data

The data used in this study is accreditation results in 2017-2018 and UNBK results in 2018 for a high school level in Indonesia. Accreditation data was obtained from BAN-S/M, and scores of UNBK were obtained from the Agency for Research and Development, Ministry of Education and Culture (Balitbang, Kemendikbud). The UNBK results data (as the response variable / Y) use in the study is the average scores of the three subjects (Indonesian, Mathematics, and English). The average scores of the UNBK per school are categorized into four categories, consists of "Very Good" category (where Y> 85), "Good" (70 <Y≤85), "Enough" (55 <Y≤70), and "Less" (with Y≤55). Simultaneously, the accreditation data is 129 items for eight SNP instruments with a Likert scale from 0 to 4 (as a predictor variable / X). The data consists of 6,771 senior high schools in Indonesia. The data is the combination of accreditation data from 8,252 schools and the scores of UNBK from 21,137 schools.

F. Method

Phase I: Data Analysis Preparation

1. Perform preprocessing data by combining accreditation data and UNBK data based on the number of the national school (NPSN).
2. Explore data to provide an overview of the data.
3. Change the response variable into four categories.
4. Divide the data into training data and testing data for various options (7: 3, 8: 2, and 9: 1).

Phase II: Classification Modeling (Random Forest)

1. Multi-class random forest

a. Weight the classes where observations in the majority class are given less weight than observations in the minority class (minority class weights = 1) so that the effect of observing minority class increases and becomes balance with the majority class.
b. Model a multi-class random forest classification on training data.
c. Evaluate the classification model by calculating the accuracy and G-Mean values on testing data.
d. Repeat steps (a) through step (c) by optimizing the hyperparameter, and folding Cross-Validation with the value k = 5.

2. Binarization random forest
a. Handle imbalanced data with SMOTE.
b. Model a multi-class binarization random forest method OVA by duplicating the training data group as much as the class in the response variable so that the *i-th* training data group has the *i-th* class and the response to the *i-th* class with $i = 1, 2, 3$.
c. Evaluate the classification model by calculating the accuracy and G-Mean values on testing data.
d. Repeat steps (a) through step (c) by doing binarization random forest method OVO by dividing the training data group into several segments (parts) so that each section has a response consisting of two classes.

Phase III: Performance Measure

Compare the random forest classification results in stage II based on the accuracy and G-Mean values to get the best classification model.

Phase IV: Variable Importance

Calculate the importance of predictor variables as essential factors in influencing the classification of education quality based on the accreditation instrument's items.

Data analyzed by software R ver. 3.5.2 used to package "mlr" (machine learning in R) with the "classify.ranger" algorithm for a random forest that can work faster for implementing random forest classifications in high-dimensional data by producing faster imputation times.

III. Results and Discussion

A. Data Exploration

The correlation between UNBK 2018 and SNP scores based in 2017-2018 accreditation result can be seen in Table 2. Table 2 shows that the correlation between SNP and UNBK shows a considerable positive correlation. This result means that the greater the value of SNP, the greater the value of UNBK. SSP has the highest correlation value when compared to other SNP for all UNBK tests.

TABLE III
CORRELATION MATRIX OF SNP AND UNBK

	SI	SPR	SKL	SPT	SSP	SPL	SB	SPN
BIN	0.38	0.39	0.43	0.41	0.51	0.40	0.32	0.37
ING	0.34	0.38	0.40	0.40	0.50	0.38	0.28	0.35
MTK	0.23	0.26	0.28	0.30	0.36	0.26	0.18	0.23

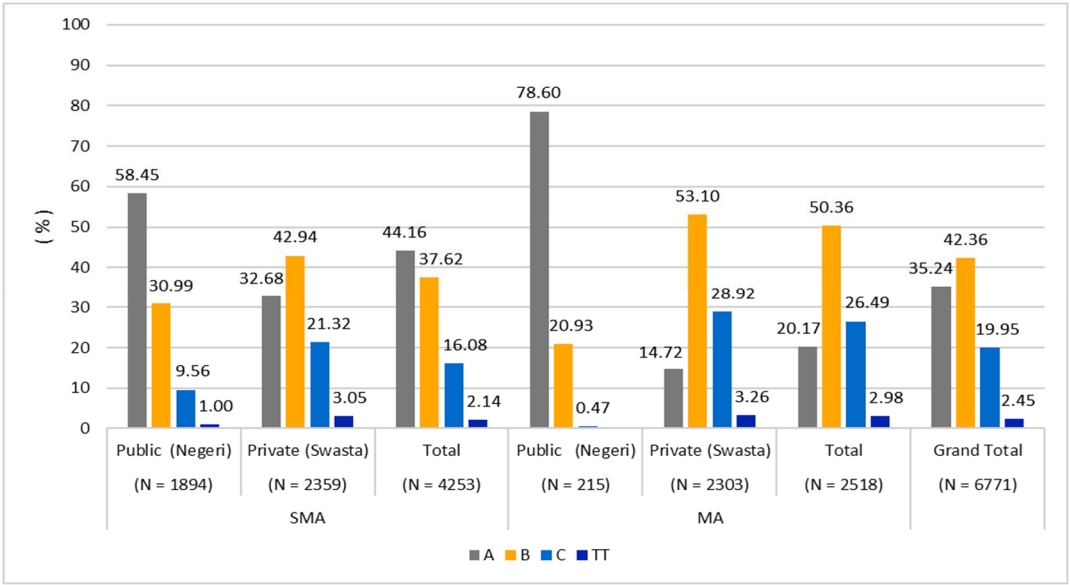
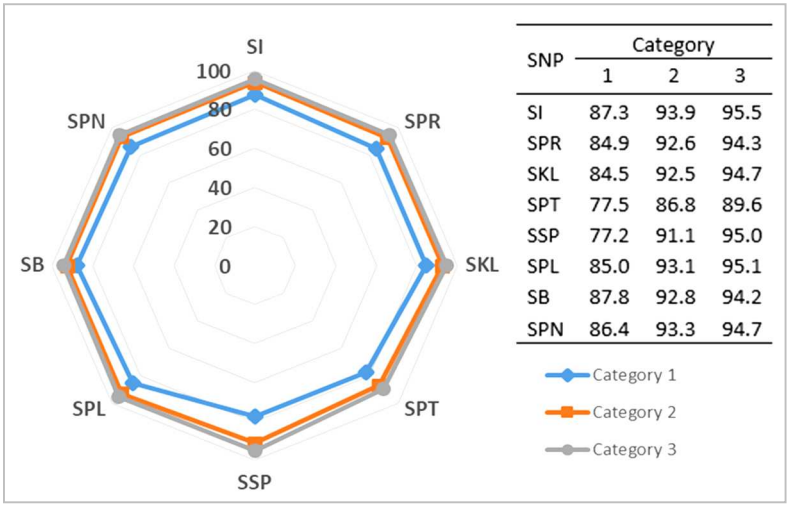


Fig. 1 School accreditation results in 2018 based on the type and status of the school

The data used in study consists of 1,894 SMAN (27.97%), 2,359 SMAS (34.84%), 215 MAN (3.18%), and 2,303 MAS (34.01%). Fig. 1 shows that public schools (SMAN) and private schools (SMAS) in 2018 that have fulfilled the criteria of 8 SNP are schools to get 'A' accredited with a percentage of 44.16% and 'B' accredited of 37.62% from 4,253 schools. Islamic public schools (MAN) and Islamic private schools (MAS) in 2018 tend to get 'B' accredited with a percentage of 50.36% and 'A' accredited of 20.17% from 2,518 schools.

The categorization of education quality based on UNBK is divided into four categories. In this study, education quality has only three categories because the 4^{th} category is no observation. Fig. 2 is the education quality categories group, poor quality education, sufficient quality education, and good

quality education. Fig. 2 also shows that the increase in SNP achievement at the school level tends to have a high level of education quality categorization based on the average score of the UNBK. So it can be concluded that SNP affects the achievement of UNBK.



SNP	Category		
	1	2	3
SI	87.3	93.9	95.5
SPR	84.9	92.6	94.3
SKL	84.5	92.5	94.7
SPT	77.5	86.8	89.6
SSP	77.2	91.1	95.0
SPL	85.0	93.1	95.1
SB	87.8	92.8	94.2
SPN	86.4	93.3	94.7

Fig. 2 The result of accreditation based on the categorization of UNBK

B. Random Forest for Multi-class Classification

Multi-class random forest algorithm (default/DF) is not set to parameters. The random forest algorithm with optimizing hyperparameter is done by several modelling parameter settings, namely *mtry* parameters, node size, and a number of trees.

1. The *mtry* parameters
   This parameter is the number of candidate variables randomly chosen at each node when growing a tree [18]. In this study, *mtry* parameters are tried from 5 to 35.
2. Minimum node size is tried from 3 to 9.
   The node size parameter specifies the minimum size of the terminal node [18]. In classification modelling, the default value is 1.
3. The number of trees built is 100, 300, 500 and 700.
   Previous studies showed that the best performance gain can be achieved when growing the first 100 trees used a large number of real datasets [19], [20].

The results of the evaluation models for the multi-class random forest in Table 3. Table 3 shows that the random forest algorithm with optimizing hyperparameter (cut-off 80:20) achieved the best in terms of G-Mean value compared to random forest default (RF-DF) algorithm with 48.95% while RF-DF took 47.22%. However, in terms of accuracy, RF-DF still performs better with 88.84% compared to the random forest by optimizing hyperparameter (RF-OH), which took 88.17%.

TABLE III
THE COMPARISON PERFORMANCE BETWEEN RF-DF AND RF-OH ALGORITHM FOR MULTI-CLASS CLASSIFICATION

	RF-DF (%)			**RF-OH (%)**		
	90:10	**80:20**	**70:30**	**90:10**	**80:20**	**70:30**
Accuracy	87.26	88.84	89.06	85.63	88.17	87.88
G-Mean	30.24	47.22	36.99	30.81	48.95	37.91

C. Binarization Random Forest

The evaluation model of binarization random forest also uses accuracy and G-Mean to measure the goodness of fit. Table 4 shows the comparative performance between OVA and OVO for binarization technique. Based on Table 4, OVA binarization has the best accuracy with 90.02% (cut off 80:20) and the best G-Mean with 25.25% (cut off 70:30). The OVO binarization does not provide a G-Mean value because no observation predicts category "3", this is contrary to the actual data in category "3". Therefore, the OVO binarization has not provided an evaluation model measured independently considering the performance results of a classification in each class.

TABEL IV
THE COMPARISON PERFORMANCE BETWEEN OVA AND OVO ALGORITHM FOR BINARIZATION TECHNIQUE

	OVA (%)			**OVO (%)**		
	90:10	**80:20**	**70:30**	**90:10**	**80:20**	**70:30**
Accuracy	89.04	90.02	89.95	89.41	89.65	89.7
G-Mean	0	25.03	25.25	0	0	0

D. Random Forest for Multi-class Classification

Random forest modelling produces variable importance obtained from MDG. Fig. 3 shows variable importance in a random forest for 15 predictor variables with the highest of MDG values generated by random forest for multi-class classification and binarization. The five items of accreditation instruments have the highest variable importance for the four classification models of random forest are indicators of the availability of chemical laboratories (x69), physics laboratories (x68), language laboratories (x71), biological laboratories (x67), and availability of installations electricity (x62) which are components of SSP.

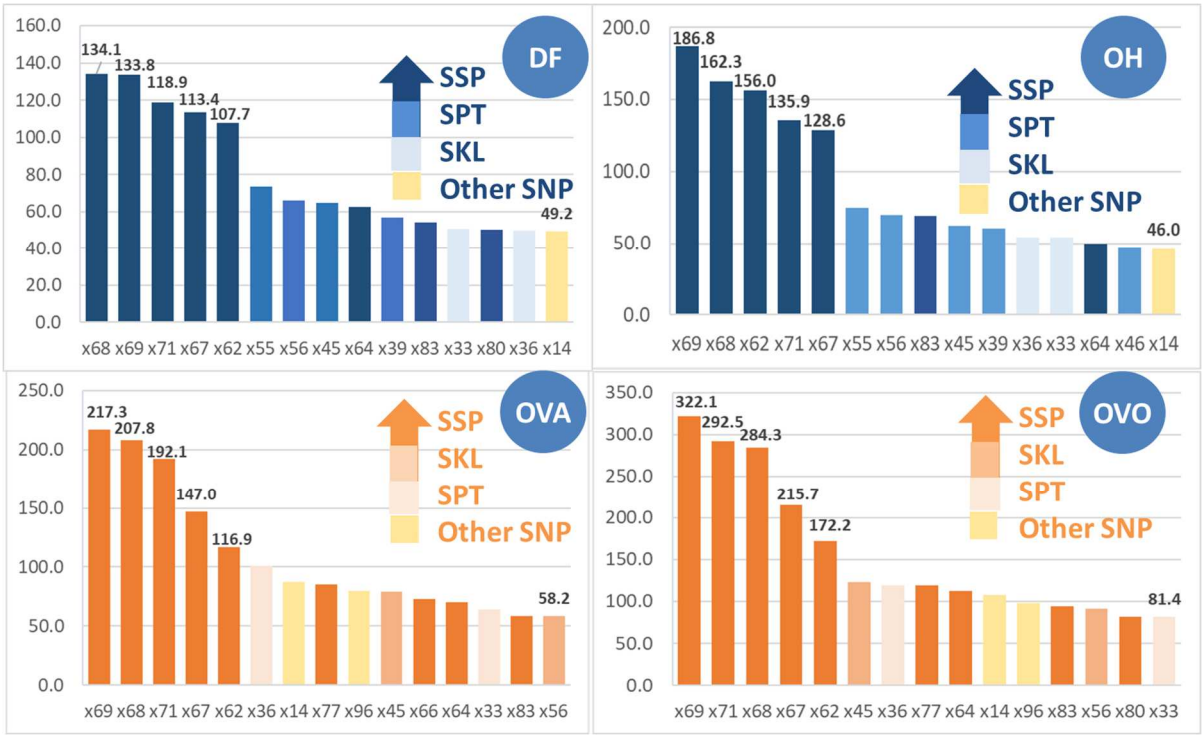


Fig. 3 Variable Importance in *Random Forest* for 15 predictor variables

The variable importance can be described as the importance of eight SNP because the predictor variables in the random forest are the items of eight SNP assessments. Table 4 shows the variable importance of eight SNP in classifying the quality of education as a result of random forest classification modelling. Based on Fig. 3 and Table 5, the three SNP with the highest variable importance average in classifying the quality of education is standards of school facilities (SSP), standards of teacher and education staff (SPT), and standards of graduate competency (SKL).

TABEL V
THE COMPARISON PERFORMANCE BETWEEN RF-DF AND RF-OH ALGORITHM FOR MULTI-CLASS CLASSIFICATION

SNP	Standard Components	RF-DF	RF-OH	OVA	OVO
SSP	X57-X84	44.82[1]	48.62[1]	81.91[1]	56.00[1]
SPT	X38-X56	34.05[2]	33.11[2]	36.33[3]	23.36[3]
SKL	X31-X37	25.68[3]	26.25[3]	42.64[2]	34.84[2]
SPR	X10-X30	21.97	19.64	24.82	15.87
SPL	X85-X100	20.98	18.56	27.53	19.31
SB	X101-X116	19.53	19.05	16.58	10.32
SPN	X117-X129	18.70	17.22	16.50	10.52
SI	X1-X9	18.35	16.18	18.36	10.66

IV. CONCLUSION

This study concludes that the multi-class random forest model was batter than the other models on SMA/MA data in 2018. This model produces important variables that make those as important factors in the classification of education quality based on the accreditation instruments' items. Fifteen important factors sequentially are items 69, 68, 62, 71, 67, 55, 56, 83, 45, 39, 36, 33, 64, 46, and 14. Based on indicators of important factors generated in a multi-class random forest, SNP has an important role for classifying the quality of education are standards of school facilities (SSP), standards of teacher and education staff (SPT) and standards of graduate competency (SKL).

Future research is suggested to develop a classification model by adding predictor variables were indicated to influence the classifying quality of education, such as a database for primary and secondary education (dapodikdasmen) and/or using the quality assurance of education data (PMP). The multivariate classification modelling needs to be developed to identify variables important in classifying the quality of education to improve the quality of education.

ACKNOWLEDGEMENT

The authors are grateful to Center of Assessment Education Balitbang, Kemendikbud RI and BAN S/M for support towards this research and DRPM-RISTEKDIKTI RI for support financial at sponsorship through the PKN-PTM Scheme for the year 2019 as the corresponding author in this study is Dr Budi Susteyo and Dr Indahwati.

REFERENCES

[1] Indonesian government, "National Education Standards (SNP)," 2005.
[2] Indonesian government, Permendikbud No.004/H/AK/2017, "Criteria and Instrument Accreditations for SMA/MA," 2017.
[3] Indonesian government, Permendikbud No. 3 of 2017, "Educational Assessment by The Government and Schools," 2017.
[4] D. Vita, B. Susetyo, and B. Indriyanto, "Generalized Structured Component Analysis (GSCA) for National Education Standards (NES) of Secondary School In Indonesia," *Global Journal of Pure and Applied Mathematics*, vol. 11, pp. 2441–2449, Apr 2015.
[5] M. Hijrah, B. Susetyo, and B. Sartono, "Structural Equation Modeling of National Standard Education of Vocational High School Using Partial Least Square Path Modeling," *IJSRSET*, vol. 4, pp. 1418–1422, Apr. 2018.
[6] I. A. Setiawan, B. Susetyo, and A. Fitrianto, "Application of Generalized Structural Component Analysis to Identify Relation between Accreditation and National Assessment," *IJSRSET,* vol. 4, pp 93–97, Oct 2018.

[7] L. Breiman, “Random Forest,” *Machine Learning*, vol. 45, pp. 5-32, Apr 2001.
[8] Q. Yanjun. (2017) The CMU website. [Online]. Available: www.cs.cmu.edu/~qyj/papersA08/11-rfbook.pdf.
[9] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, ”DBSMOTE: Density-based Synthetic Minority Over-Sampling Technique,” *Application Intelligence,* vol. 36, pp. 664–684. Mar. 2012.
[10] J. Brownlee. (2015) Machine Learning Process homepage on machinelearningmastery. [Online]. Available: https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/.
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “{SMOTE}: Synthetic Minority Over-Sampling Technique,” *Journal of Atrificial Intelligence Research,* vol. 9, pp. 321-357, Jun 2002.
[12] S. Cost and S. Salzberg S, “A Weighted Neighbour Algorithm for Learning with Symbolic Features,” *Machine Learning,* vol. 10, pp. 57-58, Jan. 1993.
[13] M. N. Adnan and M. Z. Islam, “One-vs-all binarization technique in the context of random forest,” *Computational. Intelligence and Machine Learning,* vol. 5*,* pp. 385-390, Apr 2015.
[14] L. Zhou, Q. Wang, and H. Fujita, “One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies,” *Information Fusion,* vol. 36, pp 80–89, Nov. 2016.
[15] E. Hullermeier and S. Vanderlooy S, “Combining predictions in pairwise classification : An optimal adaptive voting strategy and its relation to weighted voting,” *Pattern Recognit,* vol. 43 pp. 128–142, Jan. 2010.
[16] A. Sen, M. M. Islam, K. Murase, and X. Yao. (2015) IEEEtran homepage on CS.BHAM. [Online]. Available: http://www.cs.bham.ac.uk/~xin/papers/Binarization.pdf.
[17] M. Sandri and P. Zuccolotto, *Data Analysis, Classification and the Forward Search*, Zani S., cccc A., M. Riani, and M. Vichi., Ed. Berlin, Germany: Springer, 2006.
[18] P. Probst, M. Wright, and A-L. Boulesteix. (2018) The ARXIV website. [Online]. Available: https://arxiv.org/pdf/1804.03515.pdf.
[19] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?” in *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference,* 2012, paper Proceedings, vol. 7376, p. 154.
[20] P. Probst and A-L. Boulesteix. (2017) The ARXIV website. [Online]. Available: https://arxiv.org/pdf/1609.06146.pdf.