



Unifying Variable Importance Scores from Different Machine Learning Models Using Simulated Annealing

Asep Rusyana^{1,2}, Aji Hamim Wigena¹, I Made Sumertajaya¹, Bagus Sartono^{1*}

¹ Department of Statistics, IPB University, Bogor 16680, Indonesia

² Department of Statistics, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

Corresponding Author Email: bagusco@apps.ipb.ac.id

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290226>

ABSTRACT

Received: 18 September 2023

Revised: 24 November 2023

Accepted: 23 January 2024

Available online: 25 April 2024

Keywords:

machine learning, permutation variable importance, simulated annealing, simulation data, variable importance

Each machine learning algorithm might generate different variable importance even though the identical loss function is used. The difference in the predictor rank order makes it difficult to interpret, so a single predictor rating is required. This paper proposes a method that combines predictor rating from machine learning using simulated annealing algorithms. Simulation and empirical data are used to apply the method and its evaluation. The simulation data contain as many as 24 predictors, 1000 observations, and 100 iterations. Four machine learning algorithms are used: random forest, XGBoost, neural network, and support vector machine. Then, four permutation importance variables were produced with 100 repetitions. Next, a simulated annealing algorithm generates a combined variable importance. This proposed method will be optimal if predictors are independent, and the number of predictors is more than ten. Then, the proposed method was applied to empirical data. Using the proposed method, the machine learning model needs only 14 predictors to reach the accuracy of 74.4% which is similar to the result if the algorithm involves all predictors. The proposed method is possible to be further developed and modified. The change could be employed in the objective value and the solution strategy within the simulated annealing procedure.

1. INTRODUCTION

Whatever the model, whether a statistical model or a machine learning model, researchers must be able to explain or interpret the output obtained so that problems can be solved, and goals can be achieved. Building a model includes predicting outcomes, identifying significant predictors, and identifying the model's accuracy [1].

Important predictor identification on the value of the response variable is usually one of the purposes of building models so that the model is more straightforward because predictors that cannot explain response variables need no more attention or, in other words, are excluded from the model. If a linear model is used in a study, the importance of the variable can be obtained from the value of the variable coefficient. A coefficient value that differs significantly from zero indicates that the variable is essential. In contrast, a variable coefficient value close to or equal to zero means that the variable is unimportant [2]. The machine learning model has a unique approach to obtaining information on the importance of predictor variables. This approach is variable importance analysis.

There are several ways to perform variable importance analysis, including permutation feature importance [3], Shapley Additive Explanations Feature Importance [4], and others. The analysis will produce a score and rank the importance of the predictor variables. The order of scores or

rankings of the importance of the variables depends on the machine learning model applied. Different machine learning models will produce different variable importance measurements even though the variable importance analysis method is the same. Several papers demonstrate this.

Darmawan works to identify significant variables characterizing the incidence of family food insecurity in Indonesia. They applied four machine learning modeling techniques followed by importance analysis using SHAP. These machine learning models are extreme gradient boost (XGB), random forest (RF), neural network (NN), and support vector machine (SVM). The dataset consisted of 24 predictors from a sample of 24,769 families [5].

Figure 1 summarizes the sequence of importance of predictor variables from various implemented machine learning models. It could be seen that the order of the importance ranks differs from one model to another. For example, house size is the most crucial predictor using the SVM algorithm. Meanwhile, that predictor is ranked ninth by XGB and even fifteenth by NN.

Each machine learning model's different variable importance measurements will make determining important and less essential predictors difficult. Therefore, variable importance measures need to be combined to facilitate interpretation [6]. The users of machine learning methods will interpret importance order of a single variable importance easily. The merging methods existing are the average score of

variable importance measures, variable importance with a specific weight on its rank, and ranking mode from variable importance measures. The weakness of these existing methods is that there is no objective function. The proposed variable importance will have an objective value. The objective function can measure whether the solution is still far away or near the goal. Close or achieving a goal means combining several important variables is optimal.

The proposed method utilizes a simulated annealing algorithm to find a single ranking that maximizes the minimum value of Spearman correlation between the solution

and the original variable importance measures. This paper demonstrates that this approach can generate excellent results.

The proposed method can produce a variable importance which can assess contribution of predictors in predictive models.

The measures score helps to identify which predictors are influential to prediction. The joined variable importance can provide a more comprehensive and robust understanding of feature relevance, which can aid in building more reliable and interpretable machine learning models.

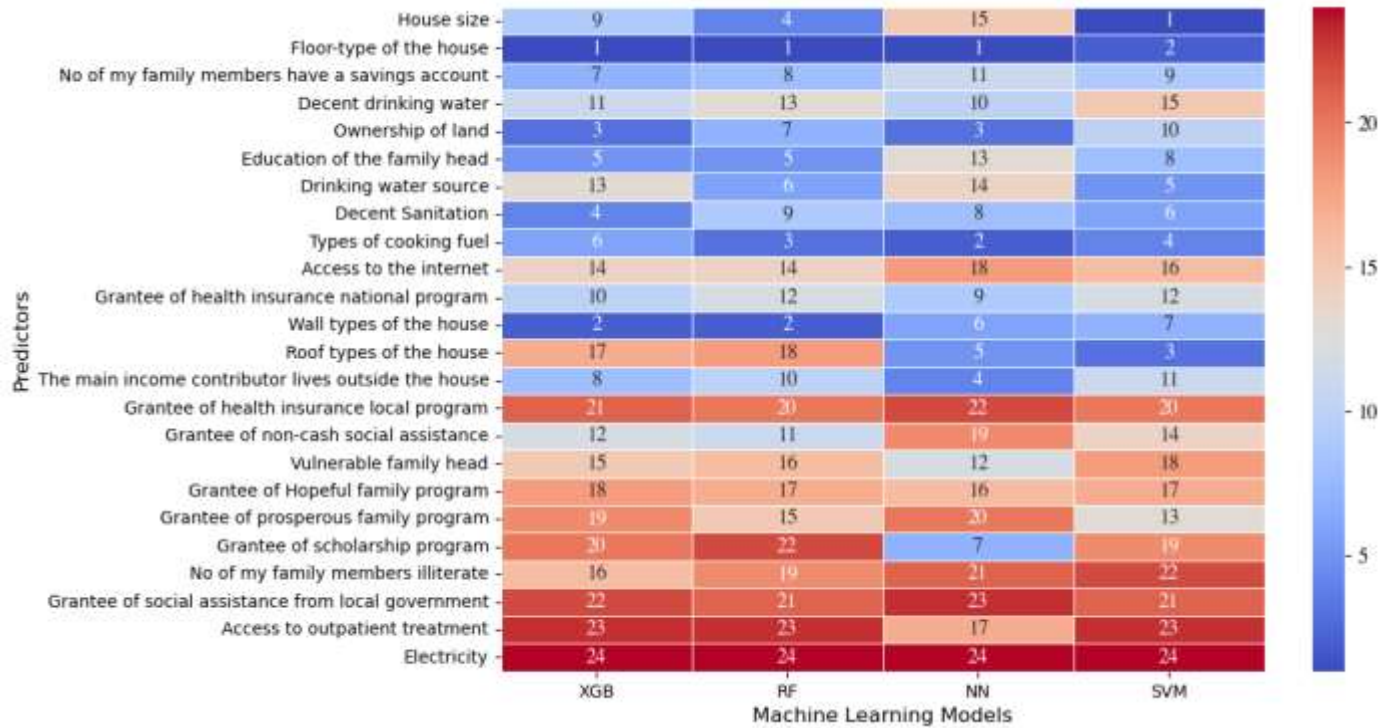


Figure 1. Variable importance of XGB, RF, NN, and SVM

2. PROPOSED METHOD TO UNIFY THE VARIABLE IMPORTANCE

2.1 Variable importance analysis

It is widely known that a predictive model resulting from a machine learning algorithm is more difficult to interpret than a classical statistical model [7]. The complication of the model comes from the fact that the model does not have an explicit mathematical function mapping a set of predictor variables to a single value of the response variable. The model tends to have a non-linear relationship among variables.

Due to that circumstance, a follow-up analysis is needed to reveal the information from the model [8]. One of the analyses useful to reveal more information from the machine learning models is the variable importance analysis (VIA) [7]. The VIA tries to reveal each predictor variable's contribution in determining the model's performance to predict the output of new observations [9]. The predictor variable with a higher contribution would have a higher ranking among the predictors [8].

Several VIA methodologies can be found in the literature. In general, the VIA procedures produce a set of scores from which the analyst can identify the relative importance of each

predictor. Some existing methods are permutation variable importance, Shapley Value, and SHAP. The permutation variable importance [10] has been a methodology that is widely implemented in many situations. This approach could be applied in models generated by any machine learning algorithm such as NN, RF, XGB, and SVM. The other advantage of this methodology is that it requires a reasonably fast computational time.

Suppose we have a predictive model F , which utilizes p predictor variables X_1, X_2, \dots, X_p to predict a variable response Y . Further, suppose that the matrix of the predictor variable values is X and the vector of variable response is y . The performance of the model F could be measured by a loss function L , which is obtained by comparing the actual response vector y and the prediction response vector $\hat{y} = F(X)$. The variable importance score of a predictor variable X_j is obtained by the following steps [9]:

- (1) Calculate $L^0 = L(\hat{y}, X, y)$; this is the value of the loss function in the original data.
- (2) Create a matrix X^{*j} by permutating the j^{th} column of X .
- (3) Calculate the predictive model of \hat{y}^{*j} based on X^{*j} , which is $\hat{y}^{*j} = F(X^{*j})$.
- (4) Calculate the second loss function value by

comparing y and \hat{y}^{*j} : $L^{*j} = \mathcal{L}(\hat{y}^{*j}, \mathbf{X}, \mathbf{y})$.

(5) Calculate the importance score of X^j by $s_j = L^{*j} - L^0$ or $s_j = \frac{L^{*j}}{L^0}$.

In a classification problem, the loss function might be $1 - \text{AUC}$ or misclassification rate. A variable with a high value of s_j indicates that X_j is more important because the exclusion of this variable from the model will increase the loss function so that the model performance is low.

Unfortunately, the result of s_j score may vary when we change the machine learning algorithm to produce model F . Even if the dataset used is identical, the importance rankings might differ between those obtained from the random forest methodology and those obtained from other machine learning algorithms. This circumstance could raise an issue in deciding which subset of variables should be considered significant. This paper will discuss the proposed approach to ease analysts' decision-making.

2.2 Notations

Suppose we ran k different machine learning algorithms to generate k predictive models. For each model, we implemented the permutation variable importance procedure to have k sets of variable importance rankings R_i , $i = 1, \dots, k$. The set R_i consists of p values or $R_i = \{s_{i1}, s_{i2}, \dots, s_{ip}\}$ where s_{ij} is the rank of the variable importance for predictor variable X_j , $j = 1, \dots, p$. Note that s_{ij} should be a whole number between 1 to p or $s_{ij} \in \{1, 2, \dots, p\}$ and $s_{ij} \neq s_{ik}$ for all $j \neq k$.

Recall that instead of the variable importance score, the value of s_{ij} is the rank of the importance score descendingly. A variable with a smaller s_{ij} is considered the more important predictor since it is associated with a higher value of variable importance score.

Our proposed method aims to obtain a single set of variable importance ranks from those k sets. In other words, we would like to unify the k sets into one. Let us use the notation of $W = \{w_1, \dots, w_p\}$ as the result of the unifying process. As w_j is the importance rank of predictor variable X_j , it also has properties like what s_{ij} has. The value of w_j should satisfy the conditions

$$w_j \in \{1, 2, \dots, p\} \text{ and } w_j \neq w_k \text{ for all } j \neq k.$$

The basic idea of the unification is finding w_j ; $j = 1, \dots, p$ so that the correlation between W and R_i is as high as possible for all $i = 1, \dots, k$. In other words, we would like to find a new variable importance rank W to agree highly with all original sets of importance rank R_i 's. Suppose that z_i is the correlation between W and R_i , $z_i = \text{cor}(W, R_i)$ for $i = 1, \dots, k$. The proposed method tries to reach a situation so that z_i is high for all $i = 1, \dots, k$. To ensure that, we define

$$Z = \min \{z_i\} \quad (1)$$

and use it as the objective function of the maximization problem.

The complete formulation of the optimization problem is then could be written as follows:

Find $W = \{w_1, \dots, w_p\}$ that maximize Z

Subject to constraints:

$w_j \in \{1, 2, \dots, p\}$; for all $j = 1, \dots, p$

$w_j \neq w_k$ for all $j \neq k$

The above optimization problem could be seen as a combinatorial problem since the solution is the permutation of

p whole number from 1 to p . Therefore, we propose implementing a simulated annealing algorithm as a metaheuristic approach to handle this optimization problem. The following subsection describes the details of the algorithm.

2.3 Computational procedure

As mentioned, a simulated annealing algorithm is utilized to find the solution for the optimization problem to unify the importance ranks. The basic algorithm is described as follows.

At the initiation step, we define the values of T_{max} and T_{min} . Both values determine the number of iterations to improve the solution. Then, an initial solution is generated. In our method, a random permutation of whole numbers 1 to p will be improved during the process. Initially, we define the value $T = T_{max}$ [11].

The iteration starts by generating a new solution slightly different from w . It can be generated by exchanging the values of two entries in w and named as w' . The positions of the changed entries were randomly chosen. The process continues by replacing w by w' whenever the performance of w' is better than w [12, 13]. In this paper, the performance of the solution is Z , the minimum correlation between w and R_i . The higher the Z value, the better the solution.

The reason behind the choice of maximizing the minimum value of correlations is as follows. A good solution of feature ranking is one that has a high correlation with every single result from machine learning algorithms. If the minimum correlation value is large, then all correlation values are larger. Therefore, maximizing minimum correlation means that the algorithm would yield high values for all correlation coefficients.

The second condition to replace w by w' is a random process. Even though $Z(w')$ is lower than $Z(w)$, there is a possibility that w' replaces w with the probability is inversely proportional to the number of iterations and the difference between the performance. It means that if the iteration has gone quite long, the probability of replacing the current solution with the worse one is lower. Also, if the performance of the alternative solution is much worse, then the probability is much lower.

As the iteration goes, the value of T is decaying. The algorithm will stop if the value of T reaches T_{min} [14], and the most recent w is the final solution containing the set of the importance ranks.

The pseudocode of the algorithm can be written as follows [15]:

Initiation:

set T_{min} and T_{max}

set θ , set k

$w = w_0$ /* Generate initial solution */

Evaluate the solution $f(w)$

$T = T_{max}$ /* Initial temperature */

Repeat

Generate a new solution w'

$\Delta E = Z(w') - Z(w)$

If $\Delta E > 0$ **Then** $s = s'$ /* Accept new solution */

Else Accept s' with probability $p = \exp\left(-\frac{\Delta E}{kT}\right)$

$T = \theta T$ /* New temperature */

Until $T < T_{min}$

Output: The best solution is fulfilled

s = current solution; s_0 = initial solution; k = Boltzman constant; T = temperature; s' = new solution; $f(s)$ = objective value for solution s .

3. SIMULATION STUDY

This section discusses a study to evaluate the performance of the proposed method to obtain a single set of variable importance ranks. The evaluation was performed through a simulation with sufficient replications to draw a conclusion [16].

3.1 Design of simulation

A simulation study examined the quality of the variable importance ranks resulting from the proposed method. The general strategy of the study was comparing the variable importance ranks to the actual rank. To do that, the simulations start with a data-generating step.

The generated datasets contain p predictors and one response variable so that the contribution of each predictor differs from one to another. The second step was the modeling process using machine learning algorithms and continued by permutation variable importance analysis based on the resulting models. It was then continued with the proposed unification process using a simulated annealing methodology. In the last step, we evaluated the result by examining the level of agreement between the actual order and the order of importance obtained by the proposed approach.

The working step of the simulation is as follows, repeated a hundred times:

(1) Step 1 generates datasets. The dataset contains 24,769 observations with 24 predictor variables and one binary response variable. The contribution of each predictor was designed so that X_{24} is the highest, X_{23} is the second highest, and so on until X_1 has the lowest contribution. This situation is considered the order of the variable importance ranks and would be used as the benchmark to evaluate the result. To reach those properties of the dataset, the detailed procedure of data generating is as follows:

- Generate randomly 1,000 observations of 24 predictor variables having a standard normal distribution
- For each observation, calculate q and p

$$q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{24} X_{24} \quad (2)$$

$$p = \pi(q) = \frac{1}{1+e^{-q}} \quad (3)$$

$$[\beta_0, \beta_1, \beta_2, \dots, \beta_{24}] = [0.800, 0.100, 0.135, 0.170, 0.204, 0.239, 0.274, 0.309, 0.343, 0.378, 0.413, 0.448, 0.483, 0.517, 0.552, 0.587, 0.622, 0.657, 0.691, 0.726, 0.761, 0.796, 0.830, 0.865, 0.900] \quad (4)$$

- Generate binary response variable $y \sim \text{Bernoulli}(p)$

(2) Step 2 builds modeling and variable importance analysis. For each dataset, some supervised machine learning algorithms were implemented. The set of algorithms includes RF, XGB [17], NN [18], and SVM [19]. Once the predictive model was obtained, a permutation variable importance analysis was performed. At the end of this step, we had four sets of variable importance ranks, one for each machine learning model. The classification problem in machine learning can be calculated by a scikit-learn library in Python [20].

(3) Step 3 is unification of the variable importance ranks. In this step, the proposed method was applied to generate unified ranks. A simulated annealing procedure was conducted to find

the optimal solution for each dataset.

(4) Step 4 evaluates the results. To evaluate the quality of the results, Spearman's rank correlation was calculated to examine the agreement between the unified variable importance rank and the actual order of importance. The Spearman correlation is preferred that Pearson correlation because we focus on the agreement level not the linear pattern. We compared that correlation with the correlation between the original variable importance ranks and the actual order. If the correlation of the unified ranks is higher, we could conclude that the methodology works well in obtaining better results.

Notice that in this simulation study, we varied the degree of the relationship among predictor variables. There are five different scenarios of correlation values among predictor variables ρ . The values of ρ are 0, 0.35, 0.70, 0.80, and 0.90. The scenario with $\rho=0$ represents the situation of a dataset having independent predictors, while $\rho=0.35$ represents a dataset with low correlated predictors. The other correlations represent high correlation.

3.2 Simulation results

Table 1 summarizes the general results of the simulation. Each table cell contains the median value of correlation between the ranks, which are from the machine-learning algorithms and the ground-truth ranks. It also shows the quantile (0.10) and the quantile (0.90) of the correlation values, which are presented in the parentheses.

Table 1. The median and the 80% range of correlation values between actual and predicted variable importance ranks

Methods	$r=0$	$r=0.35$	$r=0.70$	$r=0.80$	$r=0.9$
Random Forest	0.816 (0.733, 0.895)	0.612 (0.425, 0.755)	0.413 (0.220, 0.621)	0.372 (0.111, 0.599)	0.241 (-0.112, 0.483)
XGBoost	0.862 (0.792, 0.906)	0.692 (0.511, 0.811)	0.429 (0.179, 0.618)	0.416 (0.169, 0.618)	0.200 (-0.040, 0.473)
Neural Network	0.886 (0.831, 0.929)	0.723 (0.582, 0.837)	0.467 (0.272, 0.681)	0.399 (0.124, 0.614)	0.169 (-0.103, 0.499)
SVM	0.859 (0.788, 0.918)	0.653 (0.503, 0.777)	0.466 (0.247, 0.637)	0.390 (0.141, 0.569)	0.221 (-0.036, 0.403)
Proposed Method	0.862 (0.787, 0.911)	0.708 (0.558, 0.804)	0.487 (0.267, 0.670)	0.437 (0.240, 0.637)	0.270 (0.064, 0.512)

Several issues could arise from these results. First, if we compare the four machine learning algorithms, the permutation variable importance score generated from neural network algorithms is the best if the correlation between predictors is moderate. When the correlation among predictors is high, appointing which algorithm is better is difficult.

Second, we learned that as the degree of the relationship among predictor variables increases, all algorithms tend to face difficulty reaching good performance. It is seen that the median value decreases as the r increases. It means that when the predictor variables are highly correlated, the permutation variable importance methodology was unable to identify the importance order correctly.

Third, combining the importance score ranks using the simulated annealing technique could slightly improve the result. For the scenario with a very high degree of correlation

among predictors ($r = 0.70, 0.80,$ and 0.90), the proposed method could achieve higher median values than all single-alone machine learning algorithms.

A special note should be given to the simulation scenario with $r = 0.90$. The quantile (0.10) of the importance-ranking correlations is negative for all machine learning algorithms. However, the negative value was avoided when the proposed approach was implemented.

In addition to comparing the result in general, we also compare the performance of our proposed method using a one-on-one comparison for each replication. Recall that we ran as many as 100 replications. Table 2 presents the number of replications that the proposed approach outperformed the single ML methodology. If the numbers are greater than 50, we conclude that the proposed approach is better for generating the predictor importance rankings.

Table 2. The percentage of replications that the proposed method outperforms variable importance rank based on four different machine learning algorithms

Scenario	RF	XGB	NN	SVM
$r = 0$	77	60	36	50
$r = 0.35$	82	58	37	59
$r = 0.70$	66	66	51	54
$r = 0.80$	62	53	57	62
$r = 0.90$	61	66	64	63

The proposed method can perform better for all scenarios than RF, XGB, and SVM. It is also better than a NN when the relationship degrees among predictors are high (ρ equal to or greater than 0.70). Again, it emphasizes and strengthens the previous discussion that unification of variable importance using simulated annealing could provide more meaningful results to identify the importance of the predictors from machine learning models.

Based on the simulation, we conclude that (1) the proposed method is useful to unify several VIMs into a single variable importance ranking, and (2) the unified ranking performs better than any single machine learning result, especially when the predictors are highly correlated.

4. IMPLEMENTATION OF THE PROPOSED METHOD IN FOOD-INSECURITY PROBLEM

4.1 Empirical data

The data comes from the 2020 national economic survey (Susenas) in West Java, Indonesia. The number of predictor variables is 24, and one response variable has two classes: 0 indicates a food-secure family, and 1 indicates a food-insecure family. The number of observations is 24,679 families. The names of the predictors can be seen in Table 3.

4.2 Variable importance analysis for empirical data

Permutations variable importance (PVI) methods are applied to FIES data using four models and obtained variable importance measures in Figure 2. PVI is prioritized because of simplicity and ease of interpretation. The PVI is conducted through a hyperparameter tuning technique to obtain optimal results [21]. RF, XGB, NN, and SVM are applied because the models are suitable for classification data, the models can handle relationship non-linear between predictor variables and

response variable, and they are used widely because of their effectiveness. The order of the variable importance from RF machine learning is X1, which is the most important; X2, the sixth most important; and so on. For XGB variable importance (VI), X1 is in first place, X2 is in second place, and so on, which differs from the results of RF VI. For NN, the ranking of the importance of variables is the same as those of RF except for variables X6, X8, X11, X13, X14, X16, X18, and X19. Most of the VI order from SVM differs from the order of VI from other VI. There are differences in variable importance measure (VIM) in each machine learning model, as in the case of the SHAP-FI analysis in previous studies [5]. This VIM distinction makes it difficult to interpret. Therefore, a merger method is proposed in this study.

Table 3. Predictors, symbol, and measure in Food Insecure Experience Scale (FIES) data

Predictor	Symbol	Measure
House size	X1	Ratio
Drinking water source	X2	Ordinal
Education of the family head	X3	Ordinal
No of my family members have a savings account	X4	Ratio
Floor-type of the house	X5	Ordinal
Decent drinking water	X6	Nominal
Ownership of land	X7	Nominal
Access to the internet	X8	Nominal
Decent Sanitation	X9	Nominal
Grantee of health insurance national program	X10	Nominal
Types of cooking fuel	X11	Ordinal
Roof types of the house	X12	Ordinal
Wall types of house	X13	Ordinal
Grantee of non-cash social assistance	X14	Nominal
The main income contributor lives outside the house	X15	Nominal
Grantee of Hopeful Family program	X16	Nominal
Grantee of health insurance local program	X17	Nominal
Grantee of prosperous family program	X18	Nominal
Vulnerable family head	X19	Nominal
No of my family members are illiterate	X20	Ratio
Grantee of social assistance from local government	X21	Nominal
Grantee of a scholarship program	X22	Nominal
Access to outpatient treatment	X23	Nominal
Electricity	X24	Nominal

The model's performance is shown in terms of accuracy [22]. The values of accuracy and Area Under the ROC Curve (AUC) of each ML model are quite high. The accuracy values for RF, XGB, NN, and SVM are 0.7465, 0.7826, 0.7465, and 0.7488, respectively, while the AUC is 0.813, 0.858, 0.813, and 0.812. The accuracy values exceed 0.74, while the AUC values exceed 0.81 (Figure 3).

4.3 Implementation of the proposed method uses empirical data

Objective values have converged after the 300th iteration. The objective values are minimum (correlation [s'_i , RF-VI]), correlation [s'_i , XGB-VI], correlation [s'_i , NN-VI], correlation [s'_i , SVM-VI]). Correlation [s' , RF-VI] means the correlation coefficient between the proposed variable importance on iteration i and the variable importance of RF. Python software is used to carry out the simulation annealing technique. If the objective value converges, the optimal combined variable importance has been obtained (Figure 4).

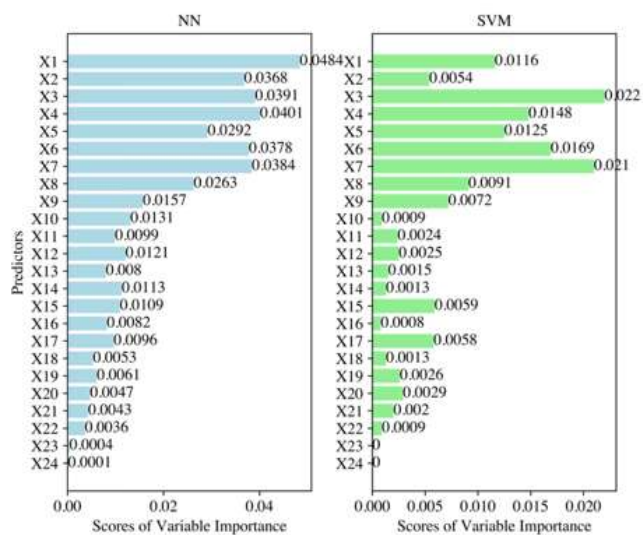
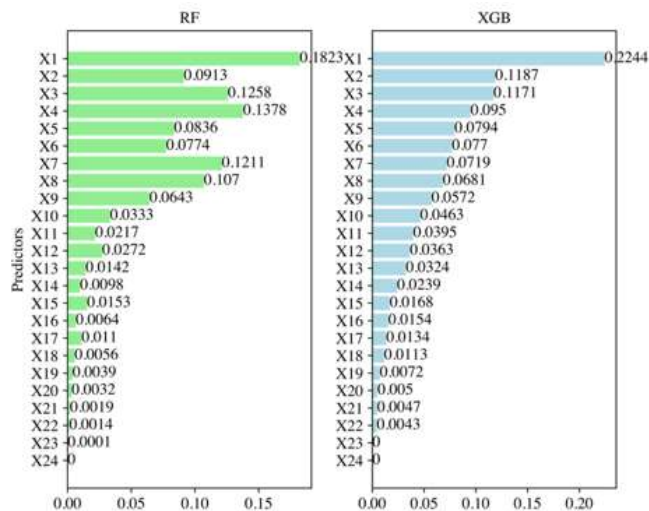


Figure 2. Score variable importance RF, XGB, NN, and SVM

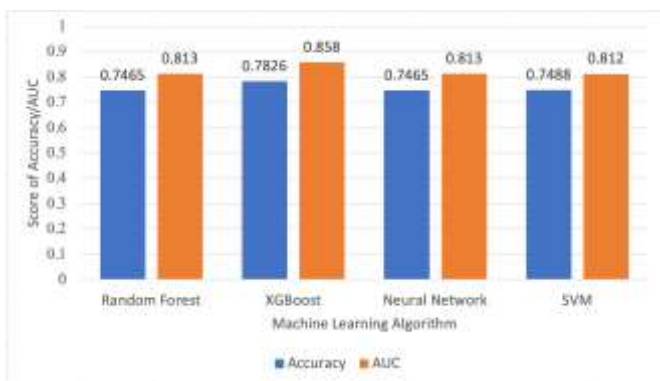


Figure 3. Accuracy and AUC score of RF, XGB, NN, SVM

The order of the variable importance for each machine learning method produces the joint VIM. The joint variable importance showed a strong Spearman correlation with the machine learning models RF, XGB, NN, and SVM, scoring 0.953; 0.923; 0.936; and 0.926, respectively. The five most important predictor variables in influencing food insecure families are X3 (education of the family head), X1 (house size), X4 (No of my family members have a savings account), X5 (Floor-type of the house), and X7 (Ownership of land). (Table 4).

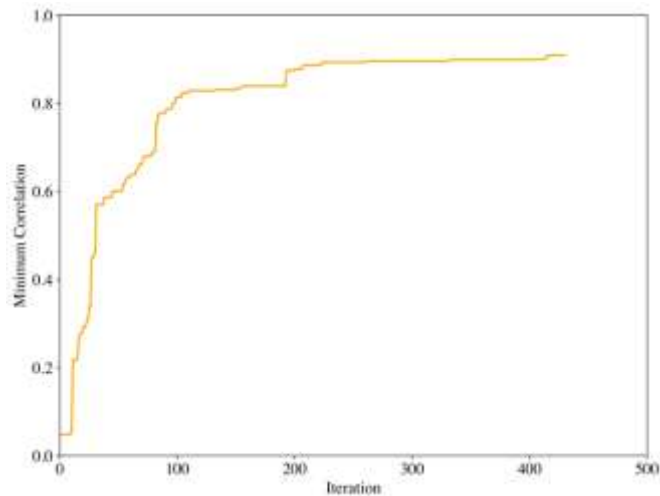


Figure 4. Iteration to obtain the proposed optimal combined variable importance measure

Table 4. Machine-learning and simulated annealing (SA) variable importance

Independent Variable	VIP of Machine Learning Algorithm				VI SA
	RF	XGB	NN	SVM	
X1	1	1	1	6	2
X2	6	2	6	11	8
X3	3	3	3	1	1
X4	2	4	2	4	3
X5	7	5	7	5	4
X6	8	6	5	3	6
X7	4	7	4	2	5
X8	5	8	8	7	7
X9	9	9	9	8	9
X10	10	10	10	20	14
X11	12	11	14	15	11
X12	11	12	11	14	12
X13	14	13	17	17	15
X14	16	14	12	18	18
X15	13	15	13	9	10
X16	17	16	16	22	21
X17	15	17	15	10	13
X18	18	18	19	18	20
X19	19	19	18	13	17
X20	20	20	20	12	16
X21	21	21	21	16	19
X22	22	22	22	20	22
X23	23	23	23	23	23
X24	24	24	24	24	24
Correlation of VIP (MLA, SA)	0.953	0.923	0.936	0.926	

4.4 Evaluation of the proposed method with empirical data

The joint variable importance scores can be seen to describe the order of influence of variables. The joint variable importance score is calculated by averaging the scores of the four PVI (Figure 5). Based on the average, X1, X2, X3, X4, X5, X6, X7, X8, and X9 have scores greater than 0.03. A score of 0.03 means the distance between the loss function score and the original data loss function score is 0.03. In other words, the predictors significantly influence the response variable. It works like the elbow method in cluster analysis [23]. If predictors. Next, characteristics of VI-SA are identified through the number of predictors.

A boxplot assesses the SA algorithm's stability in creating

VI (Figure 6). The object value (minimum correlation) increased and became more stable when more variables were included. This study looks at a variety of variables, ranging in number from few to numerous. The variables included 5, 10, 15, 20, and 24. In conclusion, various factors can improve the accuracy of the joint variable's important measure. This study examines several different variables.

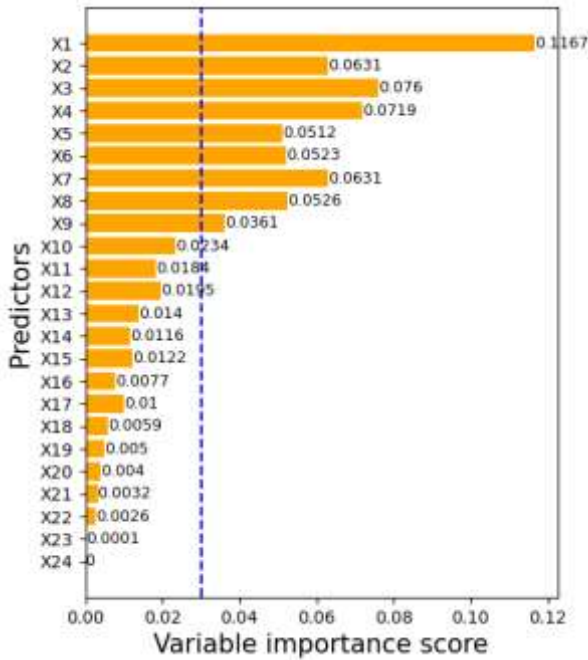


Figure 5. Scores of average variable importance of random forest, XGBoost, Neural Network, Support Vector Machine

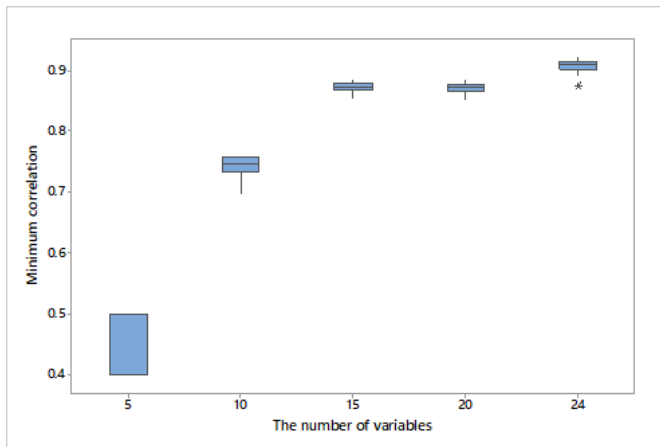


Figure 6. Comparison of objective value for the number of variables $m=5$, $m=10$, $m=15$, $m=20$, and $m=24$ with 50 replications

Evaluation of uncertainty was carried out by observing the SA process 100 times. The predictors have a rating uncertainty. The results of the SA process can produce insignificant different VI. Patterns of median ranks, ranks in percentile 5, and ranks in percentile 95. Ranks of the proposed variable importance have a level of uncertainty. The ranks can differ between line p0.05 and p0.95 (Figure 7)

The proposed variable importance measure is seen through individual classification tree algorithms. The classification tree is a method to predict output [24]. The accuracy of the joint VIM increases when the number of variables increases.

The accuracy is quite good when the number of predictors exceeds 10 (Figure 8). The accuracy shows that the proposed method is suitable.

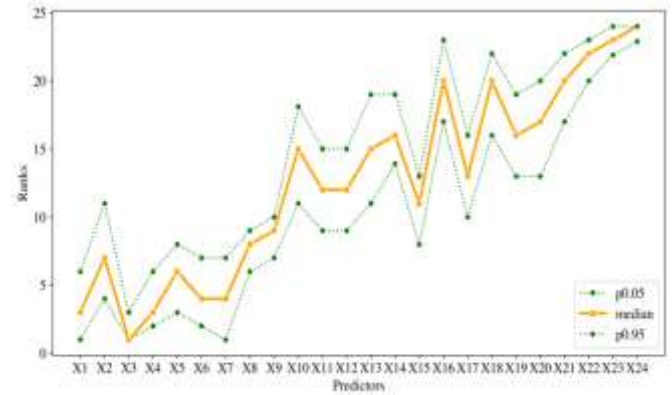


Figure 7. Patterns for ranks of each predictor with 100 replications

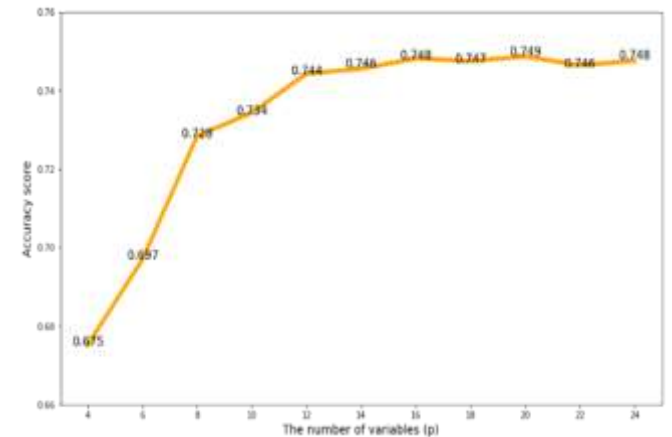


Figure 8. Accuracy scores of simulated annealing variable importance performances use a classification tree

In summary, results of its application to empirical data support that it can be applied, and results of evaluation is satisfied. The machine learning and joined variable importance is optimal when predictors are independent each others. The constituent variable importances should have high accuracies so the proposed variable importance has high accuracy. The optimal accuracy is gotten when the number of predictors exceeds 10 (ten). The proposed variable importance have uncertainty characteristics.

5. CONCLUSION

This proposed variable importance measure (VIM) is a solution if there are several variable importance measures of the machine learning model and a decision about the influence of the predictors will be determined. The method is optimal if the predictors are independent each other, the number of predictors more than 10. The method has uncertainty characteristics and high accuracy. This joint VIM makes it easy to identify rank of predictors which influence on the response variable. The proposed method should be used to dataset which have independent predictors each other. Constituent variable importance measures have optimal hyperparameter. The VIM can be better if the number of

predictors is more than ten (10). The three most influential predictors of family food insecurity in FIES data are the education of the head of the family (X3), the house size (X1), and the number of family members who have an account (X4).

Novelty of the research is a method to joint several variable importance with high accuracy. In addition, this method has a chance to be developed or modified. The modified parts include the form of the objective function and the strategy of changing the solution on the simulated annealing algorithm.

ACKNOWLEDGMENT

This work is supported by department of statistics, graduate school-IPB University.

REFERENCES

- [1] Speiser, J.L. (2021). A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *Journal of Biomedical Informatics*, 117: 103763. <https://doi.org/10.1016/j.jbi.2021.103763>
- [2] Bourguignon, M., Leão, J., Gallardo, D.I. (2020). Parametric modal regression with varying precision. *Biometrical Journal*, 62(1): 202-220. <https://doi.org/10.1002/bimj.201900132>
- [3] Hosseinzadeh, A., Zhou, J.L., Altaee, A., Li, D. (2022). Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process. *Bioresource Technology*, 343: 126111. <https://doi.org/10.1016/j.biortech.2021.126111>
- [4] Feng, D.C., Wang, W.J., Mangalathu, S., Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *Journal of Structural Engineering (United States)*, 147(11). [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0003115](https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115)
- [5] Dharmawan, H., Sartono, B., Kurnia, A., Hadi, A.F., Ramadhani, E. (2022). A study of machine learning algorithms to measure the feature importance in class-imbalance data of food insecurity cases in Indonesia. *Communications in Mathematical Biology and Neuroscience*, 2022: 101. <https://doi.org/10.28919/cmbn/7636>
- [6] Rusyana, A., Wigena, A.H., Sumertajaya, I.M., Sartono, B. (2023). An optimal approach to identify the importance of variables in machine learning using cuckoo search algorithm. *Mathematics and Statistics*, 11(6): 895-909. <https://doi.org/10.13189/ms.2023.110604>
- [7] Wei, P., Lu, Z., Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering and System Safety*, 142: 399-432. <https://doi.org/10.1016/j.ress.2015.05.018>
- [8] Salazar, F., Toledo, M.Á., Oñate, E., Suárez, B. (2016). Interpretation of dam deformation and leakage with boosted regression trees. *Engineering Structures*, 119: 230-251. <https://doi.org/10.1016/j.engstruct.2016.04.012>
- [9] Fisher, A., Rudin, C., Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177): 1-81.
- [10] Biecek, P., Burzykowski, T. (2021). *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*. First Ed. CRC Press, Taylor & Francis Group, Boca Raton.
- [11] Larasati, M.R., Wang, I.L. (2021). An integrated integer programming model with a simulated annealing heuristic for the carrier vehicle traveling salesman problem. *Procedia Computer Science*, 197: 301-308. <https://doi.org/10.1016/j.procs.2021.12.144>
- [12] Chauvin, J., Duran, R., Tavakolian, K., Akhbardeh, A., Mackinnon, N., Qin, J., Chan, D.E., Hwang, C., Baek, I., Kim, M.S., Isaacs, R.B., Yilmaz, A.G., ROUNGCHUN, J., Hellberg, R.S., Vasefi, F. (2021). Simulated annealing-based hyperspectral data optimization for fish species classification: Can the number of measured wavelengths be reduced? *Applied Sciences*, 11(22): 10628. <https://doi.org/10.3390/app112210628>
- [13] Cruz-Chavez, A.M., Moreno-Bernal, P., Rivera-Lovez, R., Ávila-Melgar, E.Y., Martínez-Bahena, B., Cruz-Rosales, M.H. (2020). GIS spatial optimization for corridor alignment using simulated annealing. *Applied Sciences*, 10(8): 6190. <https://doi.org/10.3390/app10186190>
- [14] Bouddou, R., Benhamida, F., Haba, M., Belgacem, M., Meziane, M.A. (2020). Simulated annealing algorithm for dynamic economic dispatch problem in the electricity market incorporating wind energy. *Ingénierie des Systèmes d'Information*, 25(6): 719-727. <https://doi.org/10.18280/isi.250602>
- [15] Grabusts, P., Musatovs, J., Golenkov, V. (2019). The application of simulated annealing method for optimal route detection between objects. *Procedia Computer Science*, 149: 95-101. <https://doi.org/10.1016/j.procs.2019.01.112>
- [16] Xu, Z., De, A. (2023). Assessing model accuracy using random data split: A simulation study. *Journal of Biopharmaceutical Statistics*, 33(2): 131-139. <https://doi.org/10.1080/10543406.2022.2089158>
- [17] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2: 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- [18] van Engelen, J.E., Hoos, H.H. (2020). A survey on semi-supervised learning. *Machine Learning*, Springer. 109: 373-440. <https://doi.org/10.1007/s10994-019-05855-6>
- [19] Varna Kumar Reddy, P.G., Meena, M. (2022). Automatic modulation classification using a support vector machine-based pattern recognition algorithm. *Ingénierie des Systèmes d'Information*, 27(6): 999-1007. <https://doi.org/10.18280/isi.270617>
- [20] Barbosa, A., Pelofske, E., Hahn, G., Djidjev, H.N. (2021). Using machine learning for quantum annealing accuracy prediction. *Algorithms*, 14(6): 1-11. <https://doi.org/10.3390/a14060187>
- [21] Arifin, M., Widowati, W., Farikhin, F. (2023). Optimization of hyperparameters in machine learning for enhancing predictions of student academic performance. *Ingénierie des Systèmes d'Information*, 28(3): 575-582. <https://doi.org/10.18280/isi.280305>
- [22] Hussain, E., Hasan, M., Rahman, M.A., Lee, I., Tamanna,

- T., Parvez, M.Z. (2021). CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons and Fractals*, 142: 110495. <https://doi.org/10.1016/j.chaos.2020.110495>
- [23] Yuan, C., Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *Multidisciplinary Scientific Journal*, 2(2): 226-235. <https://doi.org/10.3390/j2020016>
- [24] Huang, J.Z., Huang, W., Ni, J. (2019). Predicting bitcoin returns using high-dimensional technical indicators. *Journal of Finance and Data Science*, 5(3): 140-155. <https://doi.org/10.1016/j.jfds.2018.10.001>