Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2022, 2022:101

https://doi.org/10.28919/cmbn/7636

ISSN: 2052-2541

A STUDY OF MACHINE LEARNING ALGORITHMS TO MEASURE THE FEATURE IMPORTANCE IN CLASS-IMBALANCE DATA OF FOOD INSECURITY CASES IN INDONESIA

H. DHARMAWAN[1,2], B. SARTONO[2,*], A. KURNIA[2], A. F. HADI[3], E. RAMADHANI[4]

[1]BPS-Statistics Aceh Province, Aceh, Indonesia

[2]Department of Statistics, IPB University, Bogor, Indonesia

[3]Department of Mathematics, University of Jember, Indonesia

[4]Department of Mathematics, Syiah Kuala University, Indonesia

Abstract: The development of various machine learning algorithms on supervised models has become one of the issues in selecting a suitable algorithm. The black box of machine learning requires a technique that can be used to interpret the feature importance using the SHAP in order to obtain predictors. The class-imbalance problem in real cases is another challenge in improving the performance of minority class predictions. This study uses a food insecurity dataset, one of the SDG's important indicators to study to achieve zero hunger. The machine learning algorithms studied consisted of Random Forest, XGBoost, SVM, and NN. Meanwhile, the study of the effect of class-imbalance used three treatments: without handling, SMOTE-N, and ADASYN-N. Twelve models are built based on a combination of four algorithms and three treatments to study the performance models and their feature importance. The SMOTE-N and ADASYN-N were able to increase the sensitivity value up to 0.48 units higher when compared

*Corresponding author

E-mail address: bagusco@apps.ipb.ac.id

Received July 29, 2022

to without handling data. The agreement level on without handling data has a low value, indicated by the 0.736 ICC value, while on SMOTE-N and ADASYN-N, it is higher, indicated by the 0.925 and 0.919 ICC values, respectively. This study dataset is more suitable for using SMOTE-N. It is based on the higher ICC and superior AUC performance. The relatively high ICC value indicates that the use of machine learning algorithms does not influence the agreement level on the feature importance score. Therefore, the choice of a machine learning algorithm can refer to a measure of its performance. Random Forest produced the best performance (AUC and sensitivity). Therefore, the Random Forest SMOTE-N is the best model in this study. It produces food insecurity household characteristics with household conditions having poor water, a small house size, low household head education, few/no savers, and cement or tile flooring.

Keywords: ADASYN-N; classification; class-imbalance; feature importance; food insecurity; ICC; machine learning; sensitivity; supervised model; SHAP; SMOTE-N.

2010 AMS Subject Classification: 92B20.

1. INTRODUCTION

Machine learning can speed up one of the analytical processes with the help of algorithms built on the model. One of its types is the supervised machine learning technique, which generates a function that maps input to the desired output, and helps produce predictive models with excellent model accuracy [1]. Its ability to capture nonlinear patterns can provide additional insight that generally fails to capture the classical linear model approach [2]. There are various machine learning algorithms, such as Random Forest (RF), Support Vector Machine (SVM), XG-Boost (XGB), Neural Network (NN), and other algorithms. Each of them has a different algorithm and has advantages and disadvantages in the accuracy and interpretation of the model.

An essential issue in supervised machine learning techniques is that interpretation is not straightforward because the model formed is a black-box. The feature importance approach is an attempt to interpret the black-box model. Several feature importance techniques include feature importance permutation, Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), information value, information gain, and various other techniques.

The SHAP method can turn black-box into white-box that can be interpreted and understood [3]. This method is proven to produce a consistent score of feature importance in modeling results across various datasets and is best in interpretation [4]. This interpretation will provide additional benefits in determining policies by the government, one of which is food insecurity, which is also a global concern.

Food insecurity is a condition that occurs when a person does not have protected access to safe and nutritious food in sufficient quantities for growth and development and active and healthy life. Food insecurity is one of the leading causes of poor nutritional status [5]. The government has compiled the 2030 Agenda for Sustainable Development Goals (SDGs), consisting of 17 goals. The second goal is to encourage governments to end hunger and ensure access to safe, nutritious, and sufficient food throughout the year. Monitoring these targets uses two indicators: the prevalence of insufficient food consumption (Prevalence of Undernourishment/PoU) and the majority of the population with moderate or severe food insecurity based on the Food Insecurity Experience Scale (FIES).

Statistical data shows that food insecurity is still a fundamental problem in Indonesia. The BPS-Statistics Indonesia released in 2019 that 8.47 percent of the population had a calorie intake below 1,400 kcal/day. The International Food Policy Research Institute (IFPRI) released the Global Hunger Index; in 2020, Indonesia was only ranked 65th out of 113 countries (not including high-income countries). Based on the National Socio-Economic Survey (SUSENAS) results, in 2020, BPS-Statistics Indonesia predicts that 7.66 percent of households experience insufficient food consumption and 5.32 percent of families experience moderate or severe food insecurity.

Based on the percentage of households experiencing FIES food insecurity (moderate and severe, not including mild), there is a class-imbalance in the proportion of food insecurity and not food insecurity household classes. Building a supervised machine learning model on class-imbalance data presents a unique challenge to the model to be made. Class-imbalance data refers to a classification problem where the number of observations per class is not evenly distributed [6]. The technique of handling class-imbalance data can be done by generating synthetic data,

some of these superior techniques include Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic (ADASYN). SMOTE is a class-imbalance handling technique that does not use oversampling principles. However, it modifies the distribution of data between majority and minority classes on the dataset to balance the quantity of data for each class [7]. SMOTE can increase the size of the Area under the Curve of ROC (Receiver Operating Characteristic) as a performance measure in machine learning [8]. ADASYN improves learning in two ways: reducing bias caused by class-imbalance and shifting the boundaries of classification decisions toward data difficulties in an adaptive manner [9]. The SMOTE technique influences the order of the feature importance in the selection of models and the handling of the class-imbalance problem [10].

Several studies of food insecurity have been conducted, among others, by [11] identified factors that determine food insecurity in 134 countries in 2014, concluding that the features that affect food insecurity are low levels of education of household head and lack of social prosperity. The study suggests adding a factor of acceptance of cash transfers. In Indonesia, it was carried out by [12], [13], and [14] using SUSENAS data. [12] concluded that the higher education of household head would increase food security. [13] found a relationship between food security of rural farmer households with access to credit, rice assistance for the poor, and unconditional cash transfers. [14] study of the household factors that characterize food insecurity concludes that the main elements are recipients of social protection programs, education level, and recipients of the poor.

Based on this background, it is necessary to have the correct model among the many machine learning algorithms that can capture the issue of class-imbalance data to produce the best accuracy model and provide a reasonable interpretation of the model using SHAP. The objectives of this research are 1) to examine the effect of choosing machine learning algorithms on the SHAP feature importance; 2) to examine the effect of handling class-imbalance on measuring the score of the SHAP feature importance; and 3) to study the interpretation of the feature importance score on the results of the best algorithm on food insecurity cases in Indonesia.

2. LITERATURE REVIEW

2.1 Food Insecurity Experience Scale (FIES)

In 2013 FAO launched the Voices of Hungry (VOH) project to develop a methodology for measuring the severity of food insecurity, namely FIES. The FIES or food insecurity experience scale measures the severity of food insecurity at the household or individual level, whose value depends on yes/no answers to eight questions regarding respondents' access to adequate food. FIES captures experiences related to access to food due to lack of money or other income over 12 months, regardless of the frequency of occurrence. [5]

2.2 Random Forest (RF)

Random forest is a classification algorithm comprising a combination of independent classification trees. The classification prediction is obtained from the classification trees formed through a majority voting process (the highest number). Random forests develop the ensemble tree method developed by [15] and improve classification accuracy. The randomization process in random forests to create a classification tree is carried out on the sample data and the taking of predictor features. This process will produce a collection of classification trees of different sizes and shapes. A small correlation will reduce the prediction error of Random Forests [15].

2.3 Xtreme Gradient Boosting (XGB)

XG-Boost or eXtreme Gradient Boosting is a tree-based algorithm [16]. Boosting is an ensemble method with the primary objective of reducing bias and variance. The goal is to create weak trees sequentially so that each new tree focuses on the previous one's weakness (misclassified data). The construction of the next tree will depend on the last tree. The first tree in XG-Boost will be weak in classifying with probability initialization determined by the researcher. Then weight updates will be carried out on each tree built to produce a robust group of classification trees. The last prediction is obtained by taking the weighted sum of all the predictions of the decision tree. The basic algorithm of XG-Boost is as follows [16]:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \quad (1)$$

where $l(y_i, \hat{y}_i^{(t)})$ is a loss function to measure prediction error and $\Omega(f_i)$ is used to control the complexity of the model.

2.4 Neural Network (NN)

The creation of a Neural Network (NN) is based on a complex learning system in the brain, consisting of closely related sets of neurons. The NN algorithm's advantages include that it does not require many assumptions, makes an excellent non-linear model, and provides a model that approximates the existing system. NN consists of three layers that are the input layer, hidden layer, and output layer. This architecture is also known as Multi-Layer Perceptron (MLP). Each input is linked to every node in the hidden layer, and each output layer has a bias and weight. The activation function to calculate weight and bias describes the relationship between inputs to output values that can be linear or non-linear. [17]

2.5 Support Vector Machine (SVM)

Vapnik first presented the Support Vector Machine (SVM) in 1992. SVM was developed with the principle of a linear classifier. For non-linear data, SVM was developed by incorporating the kernel concept. So there is a guarantee that SVM classification will produce very accurate mapping [18]. The SVM concept seeks to find the optimal hyperplane in the input space. The hyperplane function becomes the separator of the two classes in the input space. The line with the maximum hyperplane margin becomes the best dividing line. Margin is the distance between the hyperplane and the closest pattern in each class. The most comparative pattern is called a support vector. The best hyperplanes are those between the two classes.

2.6 Synthetic Minority Oversampling Technique Nominal (SMOTE-N)

SMOTE generates data from minor classes with a neighboring approach. [8] mengusulkan Synthetic Minority-Over Sampling Technique Nominal (SMOTE-N) untuk digunakan pada fitur nominal. SMOTE-N merupakan pengembangan dari SMOTE. In contrast to SMOTE, the nearest neighbor calculation in SMOTE-N is calculated using a modified version of the Value Difference Metric (VDM) proposed by [19]. VDM considers the overlapping feature values of all feature vectors and defines the distance between the feature values that is appropriate for the created

feature vector. The distance between the two corresponding feature values is formulated as follows [19]:

$$\delta(V_1, V_2) = \sum_{i=1}^{n} \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k \quad (2)$$

In Equation (2), V_1 and V_2 are the two corresponding feature values. C_1 is the total number of occurrences of the feature value F_1 and C_{1i} is the number of occurrences of the feature value F_1 for class i. The same convention applies to F_2 and C_{2i}. k is a constant with a value of 1. Equation (2) calculates the value difference matrix for each specific nominal feature in the feature vector and gives certain geometric distances, finite set values.

2.7 Adaptive Synthetic Nominal (ADASYN-N)

ADASYN was first proposed by [20]. ADASYN reproduces the training data until the proportion of each class is balanced by using the distribution weights for the data in the minority class based on the level of learning difficulty. ADASYN-N is a development of ADASYN proposed by [21] with a data approach with nominal types. Nearest neighbors in ADASYN-N are calculated using a modified version of the Value Difference Metric (VDM) as in SMOTE-N proposed by [8]. VDM looks at the overlapping feature values of all feature vectors. The matrix defines the distance between the corresponding feature values for the created feature vector.

2.8 Shapley Additive Explanations (SHAP)

SHAP is a method used by [4] to explain individual predictions based on Shapley's game scores. The purpose of SHAP is to calculate each feature's contribution to predictions to explain each individual's predictions. Shapley's value is described in equation 3) below [4]:

$$\emptyset_j = \frac{1}{M} \sum_{m=1}^{M} \hat{f}(x_{+j}^m) - \hat{f}(x_{-j}^m) \quad (3)$$

where M is the number of features used in the model, $\hat{f}(x_{+j}^m)$ is the marginal function of all the features used, and $\hat{f}(x_{-j}^m)$ is a marginal function that does not include a feature to j. While the SHAP algorithm is explained through equation 4) below [4]:

$$g(z) = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j z_j \quad (4)$$

where z is the coalition vector whose elements are 1 (if the feature is included) or 0 (if the feature is not included), $\emptyset_j$ is the Shapley Value which is the contribution of the j-th feature to the coalition. M is the size of the coalition. The value of g(x) is calculated for all observations, so the size of the Feature Importance (FI) in equation 5) is the sum of the values of all observations [4]:

$$FI_{SHAP} = I_J = \sum_{i=1}^{n} \left| \emptyset_j^{(i)} \right| \quad (5)$$

2.9 Classification Model Evaluation

Prediction results from a classification model are expected to classify all data correctly, but it cannot be denied that the performance of a model can work accurately. The performance measure of the classification algorithm is measured through a confusion matrix, as shown in Table 1, which is a cross-tabulation between the response feature data included in the prediction class and the actual [22].

Table 1. Confusion Matrix for Classification

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Classification accuracy can be evaluated by counting: the number of positive classes that can be classified correctly (TP), the number of negative classes that can be classified correctly (TN), the number of negative classes that are classified incorrectly into positive classes (FP), or the number of positive classes that are classified incorrectly into negative (FN). This research focuses on the value of sensitivity and specificity so that the Area under the Curve of ROC (Receiver Operating Characteristic) will be used as a measure of model performance.

2.10 Intraclass Correlation Coefficient (ICC)

Test of agreement is widely used to assess the relationship between outcomes. The degree of agreement between measurements refers to concordance between two (or more) measurements. Statistical methods are used to decide whether one technique for measuring features can replace

another.

Intra-class correlation coefficient (ICC) is one method to assess the fit between continuous feature measurements. ICC reliably reflects the degree of correlation and agreement between numerical or continuous measurements [23]. ICC is used to assess reliability between two or more measurements. ICC is the ratio between the variance between groups and the total variance. The total variance came from three sources: 1) subject, 2) measurement, and 3) residual error. If the measurement variation is assumed to be random, then the ICC formula follows [22]:

$$\mathrm{ICC} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_e^2} \tag{6}$$

where variance (σ^2) is a variation measure, subscript s = subject, o = measurement, and e = residual error. The ICC score ranges from 0 to 1. The closer to 1, the higher the agreement will be, and the number 0 indicates disagreement. The poor agreement is indicated by a score of less than 0.5; a value between 0.5 and 0.75 indicates moderate agreement, a value between 0.75 and 0.9 indicates good agreement, and very good agreement is indicated by a score greater than 0.90.

3. RESEARCH METHODOLOGY

3.1 Data Sources

This study uses SUSENAS 2020 (March) West Java Province data. These data covers 24679 sample households. The level of food insecurity is a target feature in this study consisting of Y=0 (Not Food Insecurity) and Y=1 (Food Insecurity). The predictor features (characteristics) of food insecurity used in this study refer to the results of previous household food insecurity studies. The food insecurity features used are listed in Table 2.

Table 2. Research Predictors Features

Features Name	Scale	Features Name	Scale
House Size	Ordinal	Roof Types	Nominal
Floor Types	Nominal	Main Income From the Transferee	Nominal
Decent Drinking Water	Ordinal	Grantee of Health Insurance Local Program	Nominal
Number of Family Members Having Saving Account	Ordinal	Grantee of Non Cash Social Assistance	Ordinal
Education of Household Head	Nominal	Vulnerable Household Head	Ordinal

Features Name	Scale
Ownership of Land	Nominal
Drinking Water Source	Nominal
Internet Access	Nominal
Types of Cooking Fuel	Nominal
Decent Sanitation	Nominal
Wall Types	Nominal
Grantee of Health Insurance National Program	Nominal

Features Name	Scale
Grantee of Hopeful Family Program	Ordinal
Grantee of Prosperous Family Program	Ordinal
Grantee of Scholarship Social Program	Nominal
Number of Family Members Illiterate	Ordinal
Grantee of Social Assistance From Local Government	Ordinal
Access to Outpatient Treatment	Nominal
Electricity	Nominal

3.2 Procedure Of Analysis

Pre-Processing

1) Data Preparation
   1.1. Aggregating individual data to the household level.
   1.2. Discarding observations containing missing values, “No Answering” codes, or “Don't Know” codes for 8 (eight) FIES Susenas questions.
   1.3. Establish a food insecurity class consisting of "Not Food Insecurity" and "Food Insecurity".
2) Data exploration and presenting the prevalence of food insecurity.
3) Split the data into 70% training data and 30% testing data. Balance the training data with the SMOTE-N and the ADASYN-N technique. The imbalance of the data class on the response feature (y) is an issue that will be discussed in this study. So after dividing the data into two parts, namely the training data (to form the model) and the testing data (for model evaluation), the data class-imbalance is handled. This study uses three treatments of class-imbalance problem, namely without handling data and two synthetic data, which are made using the SMOTE-N and the ADASYN-N technique.

4) Divide the data cluster randomly into ten parts for the purpose of 10-fold cross-validation, which will be used in the search for optimal hyperparameters.

Processing (Model Building)

5) Build a classification model of Random Forest (RF), Support Vector Machine (SVM), XG-Boost (XGB), and Neural Networks (NN)

6) Searching for optimal hyperparameters to get the best model for each classification model.

6.1. RF algorithm

Perform hyperparameter tuning of the following parameters: *n_estimators* (number of trees); *max_features* (number of features to consider when looking for the best split); *max_depth* (maximum tree depth); and *min_samples_leaf* (minimum number of samples required to be in the leaf node)

6.2. XGB algorithm

Perform hyperparameter tuning of the following parameters: *eta* (learning rate is the step size reduction used in updates to prevent over fitting); *min_child_weight* (minimum number of instance weight hessian required on *child*); *subsample* (ratio of subsamples of training instances); and *colsample_bytree* (column subsampling ratio when constructing each tree, subsampling occurs once for each tree constructed).

6.3. NN algorithm

Perform hyperparameter tuning of the following parameters: *hidden_layer_sizes* (number of layers and number of nodes); *activation* (activation function for hidden layer); *solver* (for weight optimization), *alpha*, and *learning rate*.

6.4. SVM algorithm

Perform hyperparameter tuning of the following parameters: *C* (regularization parameter, regularization strength is inversely proportional to C); *Kernel* (determines the type of kernel to be used in the algorithm including linear, poly, rbf, sigmoid, and pre computed); and *Gamma*.

Score of the Feature Importance Level

7) Calculate the SHAP Feature Importance as a score of the features importance in each classification model.
8) Calculation "Measures of Agreement" to assess the agreement (assess agreement) of the feature importance between the classification machine learning algorithms on without handling, SMOTE-N, and ADASYN-N using the Intra-class Correlation Coefficient (ICC).

4. RESULTS AND DISCUSSION

4.1 Data Exploration

Based on the March 2020 SUSENAS data in West Java Province, which consisted of 25091 households, there were 322 households not included in the research dataset. Because eight types of FIES questions were answered that the household did not know (code 8) or refused the answer (code 9), so only 24679 households were analyzed in this study. The number of households in the food insecurity category in this study was 5351 households, or only 21.60 percent. Based on this value, it is clear that there is a class-imbalance problem.

Table 3 shows the number of food and not food insecurity categories by data type. The number of food insecurity in the testing and training data are 1587 and 3764 households. The percentage of food insecurity categories is 21.60 percent. The formation of SMOTE-N and ADASYN-N data uses a sampling strategy=1, which means that data synthetic for the minority class (food insecurity) is the same as the amount of data for the majority class (not food insecurity).

Table 3. Amount of data by class on target feature, testing data, and class-imbalance handling techniques on training data

Classes	Testing Data	Class-Imbalance Handling Techniques on Training Data		
		Without Handling	SMOTE-N	ADASYN-N
1: Food Insecurity	1587	3764	13572	13572
0: Not Food Insecurity	5844	13574	13613	13613
Total	7431	17338	27185	27185

Formation of models in each algorithm using hyperparameter tuning. This process uses a 10-

folds cross-validation of the training data. Hyperparameter tuning uses the concept of Bayesian Optimization with a Tree-structured Parzen Estimator (BO-TPE) to produce optimal parameters used in each algorithm and class-imbalance handling techniques, as shown in Table 4. The parameters obtained in SMOTE-N and ADASYN-N data are more similar to those obtained without handling data.

Table 4. Optimal hyperparameters according to machine learning modeling algorithms and class-imbalance handling techniques

Algorithms	Parameter	Class-Imbalance Handling Techniques		
		Without Handling	SMOTE-N	ADASYN-N
XGB	eta	0.013	0.217	0.575
	subsample	0.742	0.684	0.910
	colsample_bytree	0.587	0.741	0.854
	max_depth	16	16	19
	min child weight	14	6	5
RF	n_estimators	325	392	429
	max_depth	40	49	32
	max_features	3	9	11
	min samples split	6	11	3
	min_samples_leaf	9	2	2
	criterion	*entropy*	*entropy*	*entropy*
NN	hidden_layer_sizes	90	110	150
	activation	*relu*	*tanh*	*relu*
	solver	*adam*	*adam*	*adam*
	alpha	0.277	0.004	0.264
	learning_rate	*constant*	*adaptive*	*constant*
SVM	C	16.358	113.954	58.052
	Kernel	*rbf*	*rbf*	*rbf*
	Gamma	*auto*	*auto*	*auto*

After obtaining the optimal hyperparameters, the model is evaluated on testing data to produce several measures of model performance, as shown in Figure 1. The sensitivity value of the algorithm on without handling data tends to be low, which means that the model does not produce many positive class predictions (food insecurity). The technique for handling class-imbalance, using the SMOTE-N and ADASYN-N techniques, can increase the sensitivity value so that the model is not biased in predicting the positive class and negative class (not food insecurity). XGB

and RF algorithms are the best for overall model performance in dealing with class-imbalance data with SMOTE-N and ADASYN-N techniques.

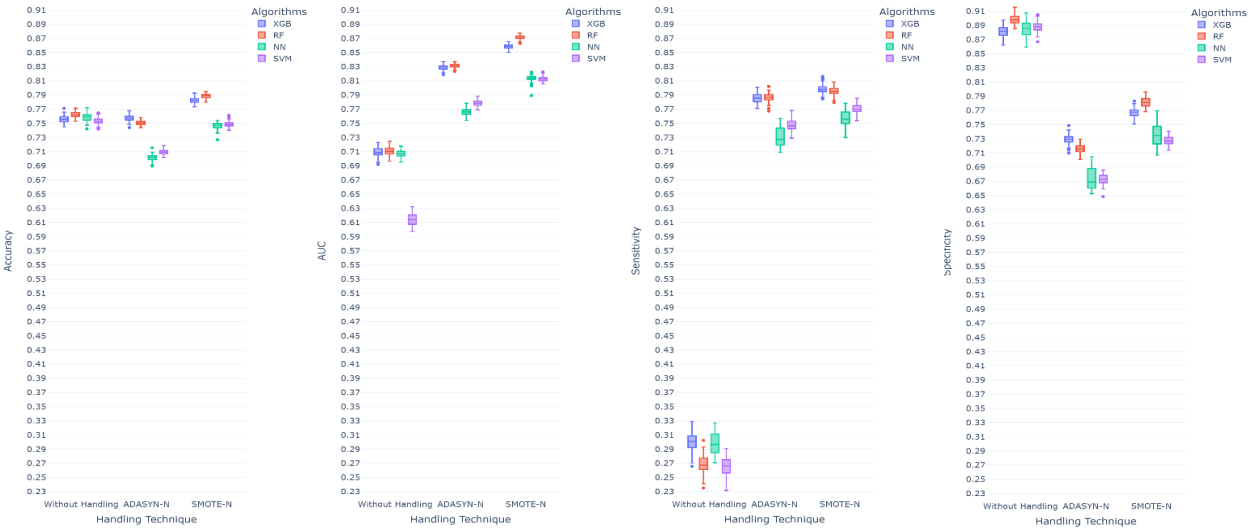


Figure 1. Boxplot of model performance measures according to machine learning algorithms and class-imbalance handling techniques (100 replicates validation)

4.2 Level of Features Importance using SHAP Feature Importance (SHAP FI)

This study's feature importance level uses the SHAP Feature Importance (SHAP FI) value approach. Figure 2 shows a heatmap ranking of SHAP FI according to algorithms and techniques for handling class-imbalance. The SHAP FI value compared is the value that has been scaled using the standard minimum-maximum method. The predictor features on the y-axis have been ordered based on the highest SHAP FI average value (SHAP FI ordered). The yellow to blue legend contains the color bar showing the SHAP FI value moving from the lowest to the highest. In contrast, the value in the heatmap cell shows the ranking of each predictor feature in each algorithm and class-imbalance handling techniques. If the SHAP FI ranking on the heatmap is similar to the color bar, it can be said that there is an agreement on the level of feature importance.

Overall, the predictor features that often appear in the top ten are the house size, floor types, decent drinking water, number of family members having saving account, education of household head, ownership of land, drinking water source, types of cooking fuel, and decent sanitation. SHAP

FI on without handling data tends to be different compared to SMOTE-N and ADASYN-N, which tend to be similar.

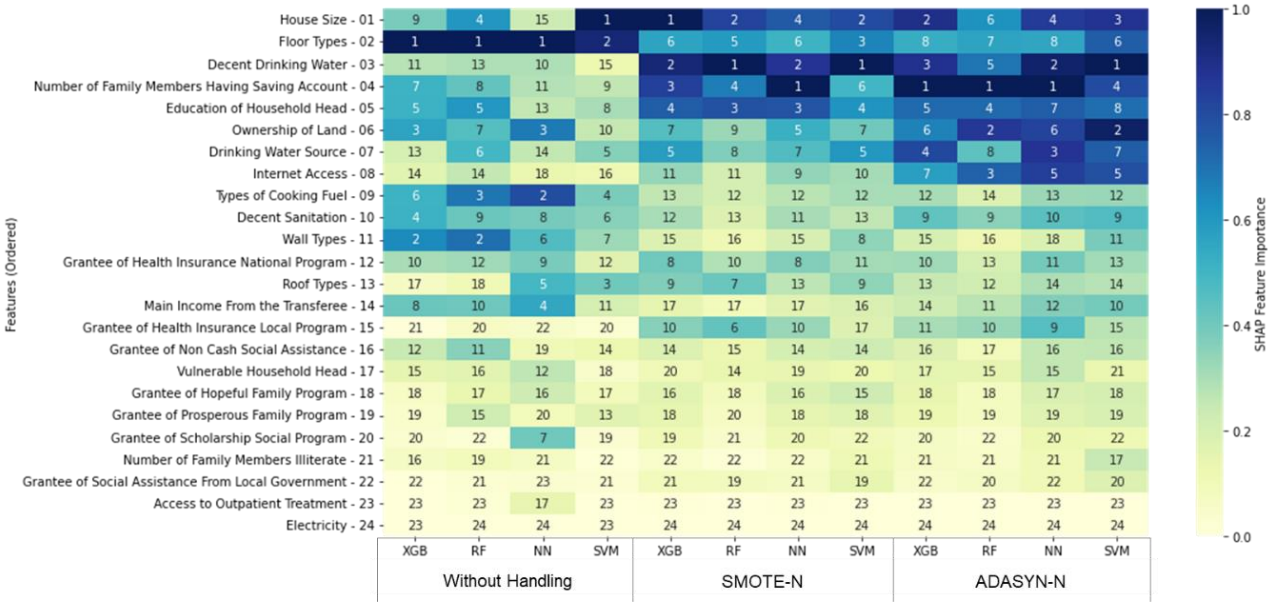


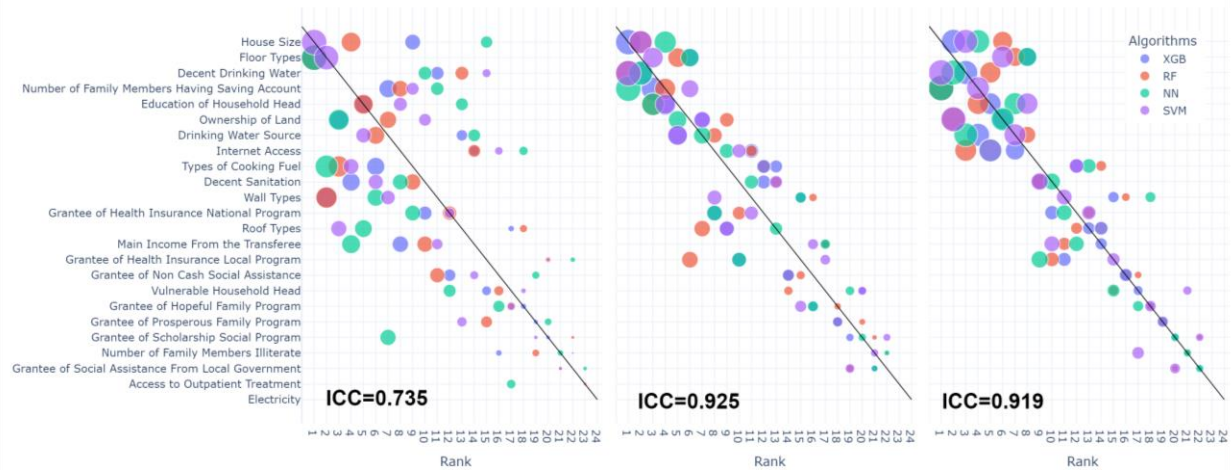
Figure 2. Heatmap of features importance scores from machine learning modeling algorithms on data with three treatments for class-imbalance problems

Several features of the SHAP FI score generated from machine learning algorithms without handling data are outside the top ten rankings. For example, the wall types feature is at the lowest rank are 2, 2, 6, and 7 are generated by the XGB, RF, NN, and SVM algorithms, respectively, but are ranked 11th in the SHAP FI order. This is supported by the heatmap of high-value FI SHAP values (which tend to be blue), which are in low ratings, including the grantee of health insurance national program, roof types, and main income from the transferee.

In the top ten rankings, five features are above the top ten SHAP FI ordered, namely the wall types in all model algorithms, followed by the grantee of health insurance national program in the XGB, SVM, and NN algorithms, the roof types feature on the SVM and NN algorithms, the feature of main income from the transferee on the XGB, RF, and NN algorithms, and grantee of scholarship social program on the NN algorithm. Meanwhile, at the top of the SHAP FI ordered ranking, there are low SHAP FI values (which tend to be yellow), such as house size, decent drinking water, and internet access.

The SHAP FI rating generated from the SMOTE-N and ADASYN-N tends to have a better level of agreement. Few features are outside the top ten SHAP FI order. In the SMOTE-N model, four features are above the top ten SHAP FI ordered, namely wall types in SVM; grantee of health insurance national program in XGB, RF, and NN; roof types on XGB, RF, and SVM; and grantee of health insurance local program in XGB, RF, and NN. In the ADASYN-N, there are only three features above the top ten SHAP FI ordered: grantee of health insurance national program in XGB; grantee of health insurance local program in RF and NN, and the main income from the transferee in SVM. At the top level of SHAP FI ordered, there is also a low-value SHAP FI (tends to be yellow) such as types of cooking fuel and decent sanitation.

Figure 3 shows the difference in the ranking of importance by type of data. The y axis is a SHAP FI ordered symbolized by feature notation. The size of the bubble chart indicates the value of SHAP FI. The importance of the same feature occurs when generating a bubble chart close to the 45-degree diagonal line.



(a) (b) (c)

Figure 3. The level of agreement between the scores of the feature importance of the results of the machine learning algorithms on the data with three treatments for the imbalanced class problem (a) Without Handling, (b) SMOTE-N, (c) ADASYN-N

Figure 3(a) shows a pattern that tends to spread in without handling data, while Figure 3(b) and Figure 3(c) show a pattern that is closer to the diagonal line; this shows the ranking of features importance generated by the SMOTE-N and ADASYN-N tend to be more similar. The Intraclass Correlation Coefficient (ICC) value shows the level of agreement on the features importance in each class-imbalance handling technique. The ICC value supports the interpretation of the resulting bubble chart, from the smallest value to 0.735 for without handling, followed by 0.925 for SMOTE-N and 0.919 for ADASYN-N.

The features importance score of the machine learning algorithm results on SMOTE-N and ADASYN-N tends to have the same level of agreement. Therefore, the user can choose one of the more appropriate techniques according to the characteristics of the data. According to [24], SMOTE can overcome data with a scale/unit inconsistency in its predictor variables. However, this technique is not very effective on high-dimensional data. In addition, the resulting synthetic example does not consider neighbors of other classes, so class overlap increases. Meanwhile, the synthetic example in ADASYN considers neighbors from other classes using a weighting technique based on density distribution. The amount of ADASYN synthetic data is dependent on the density distribution of the data. According to [25], this technique emphasized a difficult sample set to compensate for slope distribution. According to [26], the class imbalance problem depends on complexity of the data (located of minority data), level of class imbalance, size of data and classifier involved. The dataset used in this study has the characteristics of a response variable consisting of two classes and twenty-four predictor variables on a nominal/ordinal scale. Statistically, applying the SMOTE-N technique to the dataset resulted in a slightly higher level of agreement with the feature importance score compared to ADASYN-N. Therefore, the dataset in this study is more suitable for using SMOTE-N.

4.3 Comparison of SHAP FI on Imbalanced with SMOTE-N and ADASYN-N

Figure 4 and Figure 5 show the comparison of SHAP FI rankings according to predictor features generated from algorithms and class-imbalance handling techniques. The interpretation in the

figure is the same as in Figure 3 but combines without handling data with SMOTE-N and ADASYN-N. Figure 4 compares the SHAP FI ratings generated by the algorithm without handling data (green color gradient) and SMOTE-N (orange color gradient). Bubble charts are not exactly on a diagonal line but tend to spread out. The bubble chart of the ranking of features without handling data looks spread out, while the ranking of features in SMOTE-N tends to have a relative ranking on each feature. For example, the ranking of the house floor area ranges from 1, 2, and 4 (narrower range), while without handling data, it can be seen that there are pretty far differences in ranking, which are in the range of 1, 4, 9, and 15 (wider range). This is supported by the 0.717 ICC value, which indicates the level of agreement on the feature importance is low.

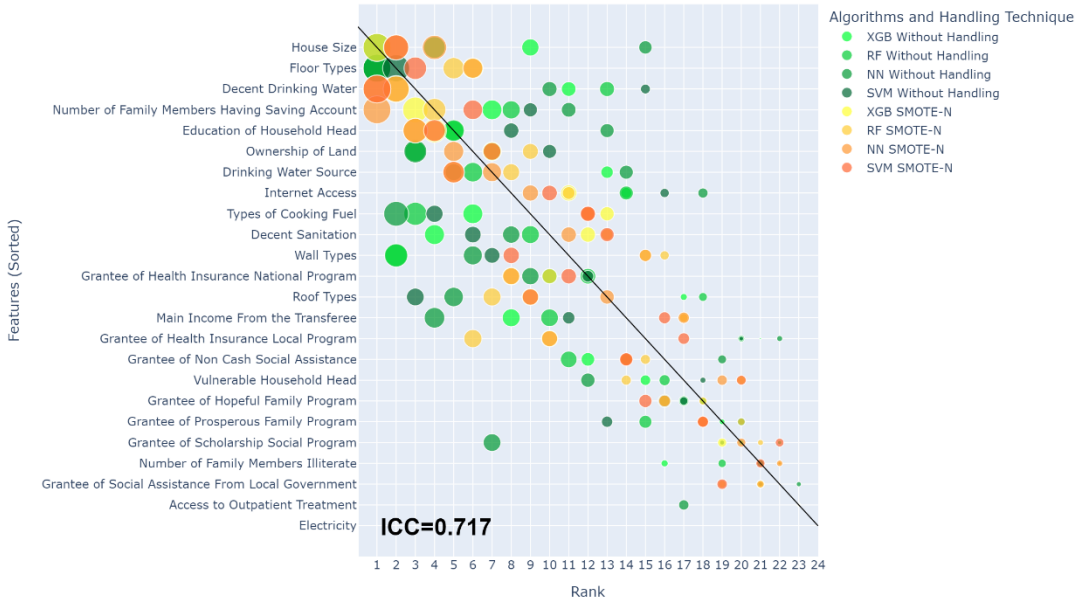


Figure 4. The level of agreement between the scores of the feature importance of the results of the machine learning modeling algorithms on without handling data and SMOTE-N

Figure 5 compares the SHAP FI ratings generated by the algorithm on without handling data (green color gradient) and ADASYN-N (blue color gradient). Similar to Figure 4, the bubble chart is not exactly on a diagonal line but tends to spread out. The bubble chart of the ranking of features without handling data appears to be spread out, while the ranking of features on ADASYN-N tends to be more closely related to each feature. This is supported by the 0.714 ICC value, which is not much different from the without handling data and SMOTE-N ICC value, which indicates a low

level of agreement in producing the level of feature importance.

Based on several feature importance generated by the SMOTE-N and ADASYN-N, those with high ratings tend to have adjacent ratings. However, they tend to be different in the medium, and those at low levels tend to be more similar. The pattern generated in without handling data does not follow the pattern found in the SMOTE-N and ADASYN-N ratings.

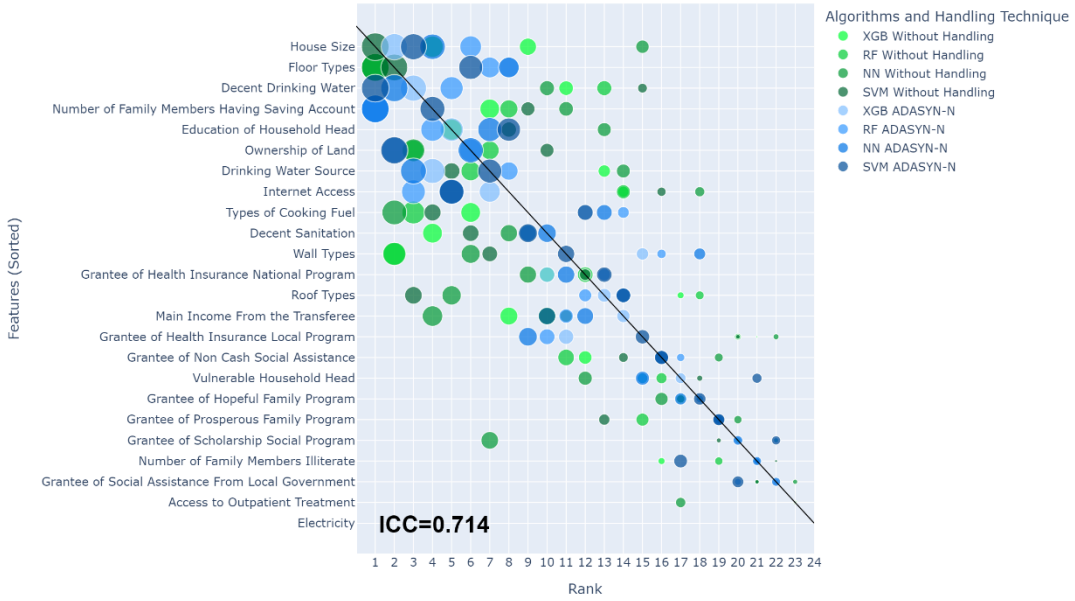


Figure 5. The level of agreement between the scores of the feature importance of the results of the machine learning modeling algorithms on without handling data and ADASYN-N

4.4 Comparison of SHAP FI on SMOTE-N and ADASYN-N

Figure 6 compares the SHAP FI rankings according to the features generated from the algorithm on SMOTE-N (orange color gradation) and ADASYN-N (blue color gradation). It can be seen that the bubble chart tends to be closer to the diagonal line because it is compared to the class-imbalance handling. For example, in the first-order feature, namely the floor area of the house, from 8 algorithms and class-imbalance handling techniques, five ranks (1, 2, 3, 4, and 6) are pretty close together. This is supported by the 0.921 ICC value, which tends to be high and indicates a better level of agreement in producing the feature importance.

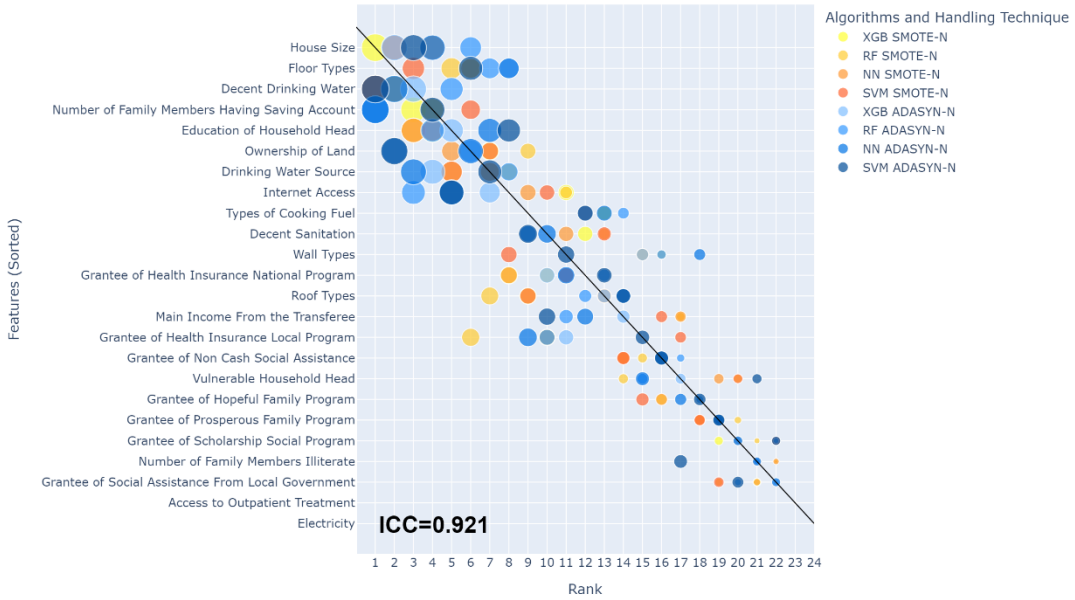


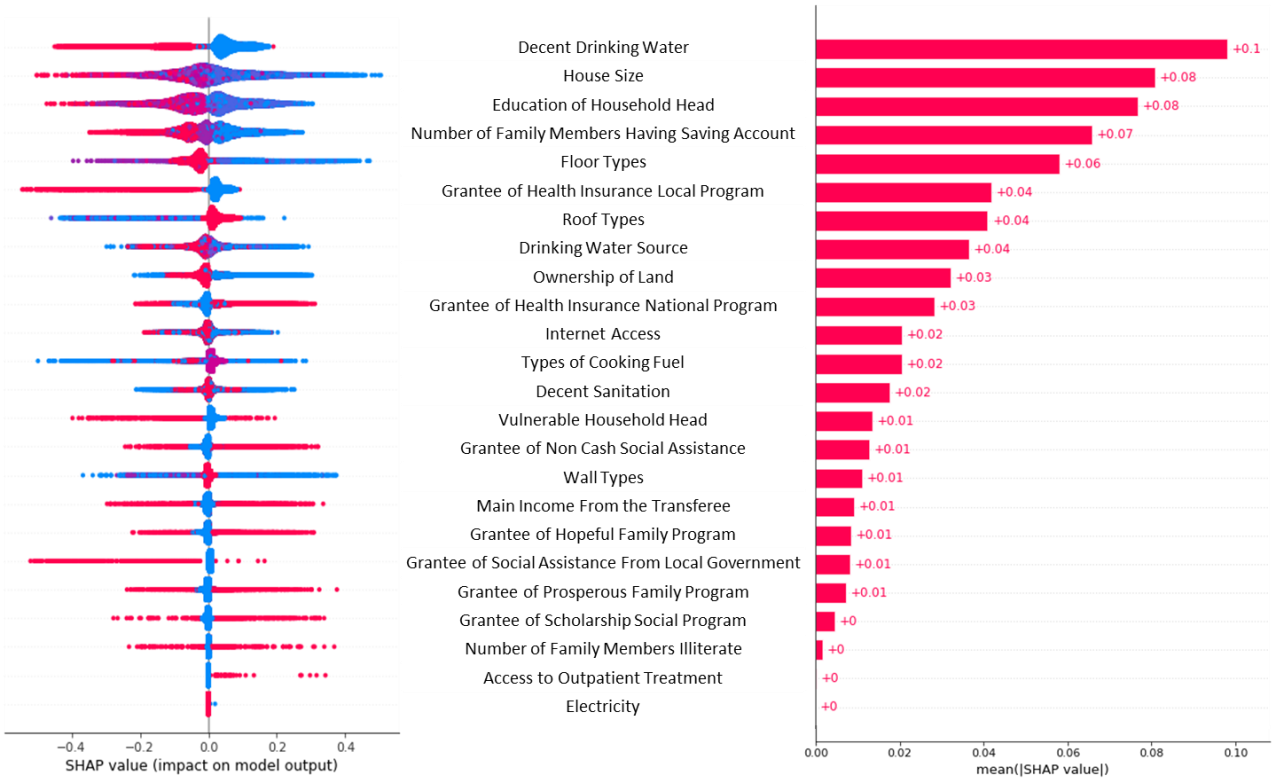
Figure 6. The level of agreement between the scores of the feature importance of the results of the machine learning modeling algorithms on SMOTE-N and ADASYN-N

4.5 Interpretation of Feature Importance from Random Forest Algorithm on SMOTE-N

Based on model performance evaluation, the RF algorithm on SMOTE-N data is the best model for identifying food insecurity cases. Therefore, the features importance is interpreted based on the SHAP Feature Importance generated from the algorithms and techniques for without class-imbalance.

Globally, the level of feature importance is reflected in the SHAP FI value, which is the absolute average value of the SHAP Value contained in Figure 7 (Global Interpretation). It shows that decent sanitation ranks first, followed by house size, education of household head, number of family members having saving account, and floor types are the top five features important. The difference in SHAP FI values between the top ranks tends to be small. For example, drinking water is decent (0.098) with a house floor area (0.081) of 0.017. The difference between ranks 2 and 3 is also smaller, only 0.004. This difference indicates that the level of importance of the food insecurity feature can change positions if using different algorithms and class-imbalance handling techniques. However, suppose the top features importance are analyzed (e.g., the top 10). In that case, the features that appear in the top group tend to be the same in various algorithms and data, including

the bottom group of features importance.



Local Interpretation Global Interpretation

Figure 7. Score of feature importance as a result of the Random Forest modeling algorithm on data that is handled by handling class-imbalance problems with the SMOTE-N technique according to the type of interpretation

Figure 7 (Local Interpretation) shows the interpretation locally. In the first rank, the low value of decent drinking water (in this case, the households using unsafe drinking water) reflected in the blue dots will be more conducive to predicting households with food insecurity status. On the other hand, as reflected in the red dots, high-value drinking water (households use proper drinking water) will encourage the prediction that households are not food insecurity. In the following few stages, it can be concluded that household conditions that are more conducive to predicting food insecurity are households with low floor area, low household head education, few savers, and low-quality floor types.

5. CONCLUSION

Class-imbalance handling using the SMOTE-N and ADASYN-N techniques were both able to increase the accuracy in predicting the positive class/minority class (food insecure) as indicated by an increase in the sensitivity measure of 0.48 units. Treatment using both techniques produced a better level of agreement from the feature importance score. Using machine learning modeling algorithms in each technique produces different feature importance scores. However, based on the level of agreement (ICC) measures on SMOTE-N (0.925) and ADASYN-N (0.919), it shows that these values are classified as very good. The dataset in this study is more suitable for using SMOTE-N based on the feature importance score. In the interpretation of the best model (RF SMOTE-N), the features importance that characterizes food insecurity are decent drinking water, house size, education of household head, number of family members having saving accounts, and floor types. The household's characteristics of food insecurity are the condition of the household do not have proper water, small house floor area, low education of household head, few numbers family members having saving accounts, and cement or tile type floors.

ACKNOWLEDGMENT

The authors thank to Ministry of Research and Technology/National Agency of Research and Innovation - Republic of Indonesia (BRIN) for funding and supporting this research. The content is solely the responsibility of the authors and does not represent the official views of the National Agency of Research and Innovation.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] J. Boelaert, É. Ollion, The great regression: Machine learning, econometrics, and the future of quantitative social sciences, Rev. Franç. Sociol. 59 (2018), 475–506. https://doi.org/10.3917/rfs.593.0475.

[2] S. Mullainathan, J. Spiess, Machine learning: An applied econometric approach, J. Econ. Perspect. 31 (2017), 87–106. https://doi.org/10.1257/jep.31.2.87.

[3] E.F.K. Hidayati, B. Sartono, A.M. Soleh, Features analysis of the research and development industry in Indonesia. Int. J. Inform. Sci. Manage. 20 (2022), 55–65.

[4] Scott M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, In: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4768–4777.

[5] Food and Agriculture Organization, International Fund for Agricultural Development, World Food Programme, The state of food insecurity in the world 2013: The multiple dimensions of food security, Roma (IT) (2013). https://www.fao.org/3/i3434e/i3434e00.pdf.

[6] M. Tan, L. Tan, S. Dara, et al. Online Defect Prediction for Imbalanced Data, in: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, IEEE, Florence, Italy, 2015: pp. 99–108. https://doi.org/10.1109/ICSE.2015.139.

[7] S.L. Phung, A. Bouzerdoum, G.H. Nguyen, Learning pattern classification tasks with imbalanced data sets, In: P. Yin (ed). Pattern recognition, chap 10. In-Tech, Vukovar, Croatia, pp 193–208.

[8] N.V. Chawla, C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure, In: ICML Workshop on Learning from Imbalanced Data Sets II (2003).

[9] S. Rahayu, T. Bharata Adji, N. Akhmad Setiawan, Penghitungan k-NN pada Adaptive Synthetic-Nominal (ADASYN-N) dan Adaptive Synthetic-kNN (ADASYN-kNN) untuk Data Nominal-Multi Kategori, J. Oto. Ktrl. Inst (J. Auto. Ctrl. Inst.), 9 (2017), 119-129. https://doi.org/10.5614/joki.2017.9.2.5.

[10] H. Mardiansyah, R. Widia, S. Effendi, Penanganan Masalah Data Kredit untuk Kelas Tidakseimbang Menggunakan Smotexgboost, Thesis, Universitas Sumatera Utara, (2019). http://repositori.usu.ac.id/handle/123456789/15054.

[11] M.D. Smith, M.P. Rabbitt, A. Coleman- Jensen, Who are the world's food insecure? new evidence from the food and agriculture organization's food insecurity experience scale, World Develop. 93 (2017), 402–412. https://doi.org/10.1016/j.worlddev.2017.01.006.

[12] I. Sundari, N.D. Nachrowi, Analisis Raskin dan Ketahanan Pangan Rumah Tangga di Indonesia (Analisis Data Susenas 2011), Jurnal Ekonomi dan Pembangunan Indonesia. 15 (2016), 121-143.

https://doi.org/10.21002/jepi.v15i2.452.

[13] A.S. Wardani, Determinan Ketahanan Pangan Dan Gizi Rumah Tangga Petani Indonesia Di Kawasan Pedesaan, (2018). https://doi.org/10.13140/RG.2.2.14278.98881.

[14] H. Irawan, B. Sartono, Erfiani, Faktor-Faktor Rumah Tangga Yang Mencirikan Tingkat Kerawanan Pangan. Thesis, IPB University. (2019). http://repository.ipb.ac.id/handle/123456789/100745.

[15] L. Breiman, Machine learning. 45 (2001), 5–32. https://doi.org/10.1023/a:1010933404324.

[16] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, 2016: pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[17] Y. Uzun, G. Tezel, Rule learning with machine learning algorithms and artificial neural networks, J. Selçuk Univ. Nat. Appl. Sci. 1 (2012), 54-64.

[18] C.W. Hsu, C.C. Chang, C.J. Lin, A practical guide to support vector classification. Technical report. Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan; 2003. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[19] S. Cost, S. Salzberg, A weighted nearest neighbor algorithm for learning with symbolic features, Mach Learn. 10 (1993), 57–78. https://doi.org/10.1007/bf00993481.

[20] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, Hong Kong, China, 2008: pp. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

[21] Y.E. Kurniawati, Multiclass Imbalance Learning dengan Adaptive Synthetic-Nominal (ADASYN-N) dan Adaptive Synthetic-KNN (Adasyn-KNN) untuk Resampling Data pada Data Hasil Tes Pap Smear, Thesis, Universitas Gadjah Mada, (2017).

[22] M. Kuhn, K. Johnson, Applied predictive modeling. Springer, New York, (2013), pp. 247–273.

[23] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J. Chiropractic Med. 15 (2016), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

[24] D. Lv, Z. Ma, S. Yang, X. Li, Z. Ma, F. Jiang, The application of SMOTE algorithm for unbalanced data, in: Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality - AIVR 2018,

ACM Press, Nagoya, Japan, 2018: pp. 10–13. https://doi.org/10.1145/3293663.3293686.

[25] M. Galar, A. Fernandez, E. Barrenechea, et al. A review on ensembles for the class imbalance problem: Bagging-, Boosting-, and hybrid-based approaches, IEEE Trans. Syst., Man, Cybern. C. 42 (2012), 463–484. https://doi.org/10.1109/TSMCC.2011.2161285.

[26] B. Santoso, H. Wijayanto, K.A. Notodiputro, et al. Synthetic over sampling methods for handling class imbalanced problems: A review, IOP Conf. Ser.: Earth Environ. Sci. 58 (2017), 012031. https://doi.org/10.1088/1755-1315/58/1/012031.