Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:2
https://doi.org/10.28919/cmbn/6801
ISSN: 2052-2541

CLASSIFICATION OF FOOD MENU AND GROUPING OF FOOD POTENTIAL TO SUPPORT THE FOOD SECURITY AND NUTRITION QUALITY

M. FARIZ FADILLAH MARDIANTO*, SULIYANTO, FARIED EFFENDY, ANTONIO NIKOLAS MANUEL BONAR SIMAMORA, AYUNING DWIS CAHYASARI, CHAEROBBY FAKHRI FAUZAAN PURWOKO, NETHA ALIFFIA

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

Abstract: The Movement for Diverse, Nutritious, Balanced, and Safe Diet, in this article called by B2SA is a program from the Indonesian government to improve resilience and nutritional quality in line with one of the Sustainable Development Goals, especially during the Coronavirus Disease (COVID-19) pandemic. In this article, classification and grouping methods are carried out to determine the development of supporting the B2SA program in Indonesia, such as the classified menu arrangement and the potential for grouped foodstuffs, especially in East Java, which is one of the provinces with a high COVID-19 spread rate and contributes greatly to food security in Indonesia. The application of the classification method in this study is to compare the performance of logistic regression and random forest. In addition, the clustering method is applied by comparing the performance of Single Linkage and K-Means. The results of this study are the category of food menu recommended by the population of East Java, which turned out to be 49.3% not meeting the B2SA standard. As for the results of the grouping, there are four groups for potential food categories of staple foods and side dishes, two groups for the category of fruits and vegetables. These results are

*Corresponding author
E-mail address: m.fariz.fadillah.m@fst.unair.ac.id
Received September 19, 2021

expected to be a recommendation for the government in supporting the stability of food security to strengthen the resilience of the food industry in Indonesia because it is a region that has food potential in Indonesia.

Keywords: food security; improvement of nutritional quality; classification method; cluster analysis; computational Statistics.

2010 AMS Subject Classification: 62H30, 65C60.

1. INTRODUCTION

The Coronavirus Disease-19 (COVID-19) pandemic has significant impact on various sectors of life and disrupted the targets of most countries to achieve the Sustainable Development Goals (SDGs). Some of the SDGs goals that all countries including Indonesia want to decrease hunger, achieve food security, and ensure good welfare and prosperity. However, according to the Food Agriculture Organization (FAO) and the International Food Policy Research Institute (IFPRI), the COVID-19 pandemic can create a new food crisis that affects the food security of a country, especially the poor and developing countries [1]. One of the developing countries whose food security sector has been affected by the COVID-19 pandemic is Indonesia. In Indonesia, to improve food security during the COVID-19 pandemic, there is one government program, namely the Movement for a Diverse, Balanced, and Safe Diet, in this article called B2SA.

Indonesia with its geographical condition and abundant natural resource potential has its regional characteristics which are reviewed based on regions such as provinces. One of the provinces in Indonesia that supports national food security is East Java Province. East Java is an excellent producer of agribusiness products, several commodity products in the agriculture, plantation, and horticulture sectors, as well as animal husbandry [2].

B2SA is one of the implementations of food consumption in the family which is carried out through the selection of food ingredients and the preparation of menus. The B2SA uses a food menu structure for one meal or a day according to mealtimes in an amount that meets the rules of balanced nutrition. B2SA plays a role in maintaining body weight, increasing the body's defense against disease, as well as distributing body intake so that it can improve nutritional quality. The

B2SA can be applied as a preventive measure to prevent the spread of the COVID-19 virus. This is following the movement launched by the government recently includes wearing masks, washing hands with soap, keeping a distance, avoiding crowds, delaying travel, and maintaining a healthy diet, namely the 6M movement in Indonesia. If the B2SA is applied, the potential for a person to be exposed to the virus is reduced [3].

In the study of Statistics, many classification methods have been developed so that they can be used to identify whether the food patterns of the Indonesian population are classified as B2SA. The classification methods include logistic regression, which includes the classical classification method. Logistic regression is used because it does not require the assumption of multivariate normality and homogeneity of the variance-covariance matrix [4]. The modern classification method is also used in this study, namely random forest. The advantage of random forest is that it can handle large amounts of training data and can provide good classification results with low error [5]. Of course, the combination of food menus according to B2SA standards in East Java is a potential food whose production must be optimized.

Optimization of food potential in each region can be one way to achieve food security in Indonesia. The grouping of regions in East Java-based on their food potential is very necessary to determine the food potential is owned in each city district so that menu can be prepared according to B2SA standards in an application. Therefore, a study is conducted to classify the food potential of districts and cities in East Java to optimize production. In this study, the grouping of the East Java region is carried out using a cluster analysis approach with a hierarchical approach. For non-hierarchical approach using K-Means method. The K-Means method is a method of grouping data based on the cluster center point (centroid) closest to the data. The basis of the K-Means algorithm aims to minimize the cluster performance index, squared error, and criterion error [6]. Meanwhile, for the hierarchical approach using the Single Linkage method, the best grouping results are selected as indicated by the internal cluster dispersion rate (icd rate) so that policy recommendations formulated based on the results of mapping food potential in East Java can be accurate and on target [7]. In this study, the grouping of potential food is carried out in four categories, namely staple foods, side dishes, fruits, and vegetables.

There are several previous studies regarding the classification method as a reference in this study, including research which compares the performance of discriminant analysis with logistic regression in classifying consumer behavior deviations in using prepaid electricity pulses [8]. The results obtained from this study, namely logistic regression can classify consumer behavior deviations more accurate than discriminant analysis because the addition of consumer data does not affect the performance of logistic regression. Furthermore, there was research that compared the random forest method, support vector machine, and propagation neural network in detecting beverage brands [9]. The results of this study indicated that random forest can handle unbalanced data, multiclass, a small number of samples, and data without preprocessing procedures. Thus, it was decided to compare the performance of logistic regression with random forest in categorizing food menu combinations recommended by the people in East Java.

Previous research related to grouping includes research that grouped dairy products and nuts based on the nutritional content and characteristics of each food [10]. The results of this study indicate that there are five groupings of dairy products and nuts that can be used as recommendations for dietary intake to stabilize adequate nutritional values. According to Hartigan [11], hierarchical cluster analysis using the Single Linkage method provides consistent results for large clusters. In addition, Wilkin and Huang [12] stated in their previous research related to K-Means cluster that can cluster large data even though it contains outliers. Therefore, in this study, it is decided to compare the performance of the Single Linkage and K-Means cluster methods.

The novelty of this study is that there are two main objectives, which is analyzed by comparing the performance of the two types of methods. The main objective in this study is the classification that belongs to supervised learning and the grouping that belongs to unsupervised learning, which is very popular in the development of Statistics and Data Science lately. Classification is done by comparing the performance of the logistic regression analysis method and random forest to determine whether the food menu according to the B2SA standard based on the results of food production in East Java. Meanwhile, the grouping is carried out using the hierarchical cluster method through Single Linkage and non-hierarchical cluster analysis using K-Means to map the results of certain foodstuffs in districts and cities in East Java.

The urgency of this research is to classify and map areas in East Java based on food potential for preparing menus according to B2SA standard. This research is expected to help related parties in improving national food security with the B2SA movement and optimizing local food production that can meet the food needs of the community, especially in East Java during and after the COVID-19 pandemic to be end.

2. PRELIMINARIES

A. Classification Method

1. Binary Logistics Regression

   Binary logistic regression is used to analyze the relationship between one response variable and several predictor variables, with the response variable in the form of dichotomous qualitative data, with a value of 1 to indicate the presence of a characteristic and a value of 0 to indicate the absence of a characteristic [13]. In summary, the stages of binary logistic regression analysis are as follows:

   - Simultaneous Parameter Significance Test

     This test is conducted to examine the role of predictor variables on response variables simultaneously or as a whole. The simultaneous test is also called the chi-square model test. The hypothesis of this test is as follows:

     H_0 : $\beta_1 = \beta_2 = \ldots = \beta_i = 0$

     H_1 : there is at least one parameter $\beta_i \neq 0$

     The test criteria used of H_0 is rejected if the p-value < α, thus it can be concluded that all predictor variables simultaneously have a significant effect on the model [14].

   - Partial Parameter Significance Test

     Partial testing is carried out to determine whether there is a significant effect on each parameter on the response variable [14]. The partial parameter significance test is carried out using the Wald test with the following test hypotheses:

$H_0 \ : \ \beta_j = 0$

$H_1 \ : \ \beta_j \neq 0 \ ; j = 1,2,3, \dots, p$

If the p-value < α, then it can be decided to reject H_0, thus it can be concluded that the predictor variable partially has a significant effect on the response variable.

- Model Fit Test

  Osborne [13] stated that the model can be said to be feasible or meet the goodness of fit if there is a match between the data entered in the model and the observed data. In binary logistic regression, model feasibility testing can be done using deviance test. If the p-value > α, then it can be decided to reject H_0, so it can be concluded that the model is appropriate or there is no significant difference between the observations and the possible predictions of the model.

- Interpretation of Odds Ratio

  The Odds Ratio (OR) value is used to interpret the parameter coefficients. Hosmer and Lemeshow [14] stated that OR is the average magnitude of the tendency of the response variable to have a certain value if x = 1 compared to x = 0.

2. Random Forest

The random forest method is a classification and regression-based method where there is a decision tree aggregation process. Breiman and Cutler [15] stated that the random forest algorithm can be briefly divided into the following stages:

- Bootstrap Aggregating Stage

  This stage involves taking a random number of samples from the original data set with returns.

- Decision Tree Formation Stage

  At this stage, the tree is built based on each data bagging cluster until it reaches the maximum size where there are three procedures to obtain multiple decision trees [16].

- Iteration Stage

  This stage is the stage of repeating bootstrap aggregating stage to decision tree formation stage, then a forest consisting of n decision trees (ntrees) is obtained.

- Majority Voting Stage

  This stage is the class prediction stage in the classification model based on the class that is most predicted by a set of trees.

3. Classification Accuracy Value (1-APER)

   One of the procedures to determine the accuracy of classification is through APER [4]. The APER value represents the proportion of samples that are incorrectly classified by the classification function. Classification errors can be seen based on the confusion matrix that presented on Table 1 as follows:

Actual	**Prediction**	
	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$
$\boldsymbol{y_1}$	n_{11}	n_{12}
$\boldsymbol{y_2}$	n_{21}	n_{22}

TABLE 1. Confusion Matrix

Based on the confusion matrix, the classification accuracy value can be calculated as follows:

$$1 - APER = 1 - \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \tag{1}$$

B. Clustering Method

Multivariate analysis is a statistical method that aims to analyze data consisting of many variables. Multivariate analysis is widely used to overcome various problems such as data reduction, group formation, to hypothesis testing [6, 17]. Cluster analysis is one part of multivariate analysis. Cluster analysis consists of hierarchical and non-hierarchical groupings [18, 19]. The distance measure used is the Euclidean distance.

The results of cluster formation through a hierarchical approach can be displayed through a dendrogram, including in Single Linkage [18]. While the non-hierarchical approach is very appropriate when used on large multivariate data and the number of groups to be formed has been determined. One of the popular non-hierarchical grouping methods used in K-Means.

1. Single Linkage (Nearest Neighbor)

Single Linkage performs the formation of clusters by combining the smallest distances.

$$d_{(i,j)k} = \min(d_{ik}, d_{jk}) \tag{2}$$

2. K-Means Method

The K-Means method is widely used in the study of big data and machine learning. The basic principle of the K-Means method is to partition clusters that separate data in separate areas [20]. The cluster formation algorithm using the K-Means method is as follows:

- Optimal Cluster Number Selection Size

The selection of the optimal number of clusters is done by looking at the Pseudo-F statistical value [7]. A high number of certain clusters indicates that the number of clusters has been optimal. If R^2 is the coefficient of determination, n is the number of samples, c is many clusters, then statistics of Pseudo-F formulated as follows:

$$\text{Pseudo} - \text{F} = \frac{(n-c)R^2}{(c-1)(1-R^2)} \tag{3}$$

- Criteria for Selection of the Best Cluster Method

The selection of the best cluster method is done by using the icd rate criteria. The icd rate criterion is closely related to the coefficient of determination. The coefficient of determination is a measure that shows the amount of contribution given by each variable in the cluster that is formed [7]. The coefficient of determination for the c group, p variable, n_c, and data in i-th group is defined as follows:

$$R^2 = \frac{(SST - SSW)}{SST} \tag{4}$$

with

$$SST = \sum_{i=1}^{n_c}\sum_{j=1}^{c}\sum_{k=1}^{p}(x_{ij}^{k} - \overline{x^k})^2 \tag{5}$$

$$SSW = \sum_{i=1}^{n_c}\sum_{j=1}^{c}\sum_{k=1}^{p}(x_{ij}^k - \overline{x_j^k})^2 \tag{6}$$

The results of grouping through certain cluster methods will be better if the value of the icd rate is getting smaller [7]. Mathematically, the icd rate can be written as follows:

$$icd\ rate = 1 - R^2 \tag{7}$$

- Principal Component Analysis

Principal Component Analysis (PCA) is a factor analysis technique where several factors will be formed in the form of variables that cannot be determined before the analysis is carried out. PCA is used to reduce data while maintaining the meaning of the data [21]. In reducing variables, it is necessary to determine the strength of the contribution or feasibility of each variable formed through the value of the closeness of the relationship between variables or the correlation matrix. One way to check the correlation matrix is to use the Measure of Sampling Adequacy (MSA) test. A variable is said to have a strong contribution and is feasible for factor analysis if it has an MSA value of more than or equal to 0.5 [22]. Thus, these variables can be analyzed further by eliminating variables that have an MSA value of less than 0.5.

C. Nutritional Standards based on Calories

A person's nutritional status is determined by several factors including energy consumption, physical activity, genetic conditions, and infectious diseases where nutrient-rich foods cannot be met with one meal because nutrients from various foods must be able to meet the needs of energy, protein, calcium, iron, zinc, vitamins A, B, C, D, magnesium, phosphate, potassium, and folic acid [23]. During the COVID-19 outbreak, it is necessary to increase the body's immunity which can be obtained from the intake of nutrients and food security that must be consumed and the easiest to socialize, namely the B2SA food concept [24]. In this study, the nutritional standard used is calories. The reason for using calories as a nutritional standard is related to the Prevalence of Undernourishment (PoU) as one of the SDGs indicators to combat hunger [25]. In addition, the nutritional standards have been converted from the Guidelines for Balanced Nutrition by the

Ministry of Health of the Republic of Indonesia in 2014 which are presented in Table 2 as follows:

Age group AKE	Staple food	Side dish		Vegeta-bles	Fruit	Milk	Sugar	Oil
	100 gr rice/equi-valent	Animal side dishes 45 gr	Vegetable side dish 50 gr	100 gr vegeta-bles	50 gr fruit	200 gr cow milk	10 g sugar	5 g oil
Child								
1-3 years old (1125 kcal)	3 p	1 p	1 p	1.5 p	3 p	1 p	2 p	3 p
4-6 years old (1600 kcal)	4 p	2 p	2 p	2 p	3 p	1 p	2 p	4 p
7-9 years old (1850 kcal)	4.5 p	2 p	3 p	3 p	3 p	1 p	2 p	5 p
Teenager Boy								
10-12 years old (2100 kcal)	5 p	2.5 p	3 p	3 p	4 p	1 p	2 p	5 p
13-15 years old (2475 kcal)	6.5 p	3 p	3 p	3 p	4 p	1 p	2 p	6 p
16-18 years old (2675 kcal)	8 p	3 p	3 p	3 p	4 p	-	2 p	6 p
Teenager Girl								
10-12 years old (2000 kcal)	4 p	2 p	3 p	3 p	4 p	1 p	2 p	5 p
13-15 years old (2125 kcal)	4.5 p	3 p	3 p	3 p	4 p	1 p	2 p	5 p
16-18 years old (2125 kcal)	5 p	3 p	3 p	3 p	4 p	-	2 p	5 p
Man								
19-29 years old (2725 kcal)	8 p	3 p	3 p	3 p	5 p	-	2 p	7 p
30-49 years old (2625 kcal)	7.5 p	3 p	3 p	3 p	5 p	-	2 p	6 p
50-64 years old (2325 kcal)	6.5 p	3 p	3 p	4 p	5 p	1 p	1 p	6 p
65 years old (1900 kcal)	5 p	3 p	3 p	4 p	4 p	1 p	2 p	4 p
Woman								
19-29 years old (2250 kcal)	5 p	3 p	3 p	3 p	5 p	-	2 p	5 p
30-49 years old (2125 kcal)	4.5 p	3 p	3 p	3 p	5 p	-	2 p	6 p
50-64 years old (1900 kcal)	4.5 p	3 p	3 p	4 p	5 p	1 p	2 p	4 p
65 and over (1550 kcal)	3.5 p	3 p	3 p	4 p	4 p	1 p	2 p	4 p
Pregnant (2500 kcal)	6 p	3 p	4 p	4 p	4 p	1 p	2 p	6 p
Breast Feeding (2500 kcal)	6 p	3 p	4 p	4 p	4 p	1 p	2 p	6 p

Source: Guidelines for Balanced Nutrition, Ministry of Health, 2014

TABLE 2. Food Composition to Meet the Nutritional Adequacy Number

3. METHODOLOGY

A. Data

In this study, the data used to perform the classification analysis is a combination of respondent's recommended food menu spread across East Java during the COVID-19 pandemic. The data was obtained from a survey using a questionnaire as a research instrument. The sampling technique used was quota sampling. There is a quorum limit, which is a maximum of 2 respondents per sub-district in East Java. From the survey results, 1,285 food menu recommendations were obtained by respondents spread across East Java. Then, the food menu is converted in caloric value as a nutritional standard based on the Indonesian Food Composition Table (TKPI) 2017. After the conversion, the calories in each variable are calculated so that the total calorie value can be obtained, and then it can be detected whether it meets the B2SA based on Guidelines for Balanced Nutrition by the Ministry of Health of the Republic of Indonesia in 2014. In addition, the data used to conduct cluster analysis is secondary data on food commodity yields in 29 districts and 9 cities in East Java Province taken from East Java in Figures 2020 [2].

B. Research Variable

The variables used in this study include variables for the classification method and variables for the cluster method.

Variable Type	Food menu	Scale
Independent	Staple food	Ratio
	Side dish	
	Fruits	
	Vegetables	
	Other Food	
Dependent	Food Menu Category According to B2SA Standard	Nominal

TABLE 3. Research variable

The research variables used to classify the food menu combinations of the East Java population consist of the dependent variable and the independent variable which shown in Table 3. In addition, the variables that will be grouped using the cluster method are food commodities in the East Java region in 2020 as many as 68 which are classified into four categories, namely staple foods, side dishes, fruits, and vegetables. The food commodities studied are presented in Table 4.

Type	Food Commodities
Staple food	Rice, corn, potatoes
Side dishes	Beef, the kampong chicken, laying hen, broiler, duck, skipjack, cob, tuna, shrimp, catfish (*Pangasius siluriformes)*, catfish (*Clarias siluriformes*), tilapia, carp, snapper, milkfish, goldfish, grouper
Fruits	Spring onion, shallots leaf, garlic, breadfruit, jengkol, melinjo, petai, mustard greens, spinach, beans. seaweed, red beans, long beans, kale, cauliflower, cucumber, cabbage, chayote, eggplant, carrot
Vegetables	Cantaloupe, melon, watermelon, strawberry. tomato, avocado, grapes, apple, starfruit, duku, durian, water guava, guava, big oranges, siamese oranges, mango, mangosteen, passion fruit, jackfruit, pineapple, papaya, banana, rambutan, salak, sapodilla, soursop, coconut

TABLE 4. Food Commodities in East Java in 2020

C. Analysis Procedure

The steps of data analysis of this study are as follows:

1. Perform descriptive analysis to determine the description of the data used.
2. Perform classification analysis
    a. Conducting data preprocessing which includes data cleaning from outliers, feature or variable selection, and data labeling.
    b. Conducting a classification analysis using the binary logistic regression method with the following steps:
        i. Testing the significance of the parameters simultaneously and partially.
        ii. Conducting deviance test to test the suitability of the model.
        iii. Interpreting the OR value.
        iv. Calculating the value of classification accuracy (1 - APER).
    c. Conducting a classification analysis using the random forest method with the following steps:

i. Determine n decision trees (ntrees) to be built based on the smallest OOB error value.
ii. Determine m predictor variables (mtry) which will be taken at random to be used in the classification tree sorting process.
iii. Taking samples with bootstrap aggregating technique to obtain a new dataset.
iv. Forming a decision tree based on the CART algorithm without pruning from the dataset, where for each node selects the best sorter is selected from m predictor variables taken at random.
v. Observing the class predictions on each decision tree that is formed.
vi. Repeat steps from (iii) – (v) in this random forest classification step until n decision trees are formed.
vii. Conduct majority voting on the alleged grouping of each decision tree.
viii. Calculating the value of classification accuracy (1 – APER).

d. Comparing the performance of the three classification methods based on 1 – APER, where the higher the value of 1 – APER, the better the classification accuracy.

3. Conducting cluster analysis.
   a. Performing Principal Component Analysis (PCA).
   b. Grouping districts and cities in East Java-based on food potential as follows:
      i. Conducting grouping with a hierarchical approach procedure including Single Linkage as well as with a non-hierarchical approach in the form of K-Means.
      ii. Selecting the best number of clusters for each method using Pseudo-F statistics.
      iii. Selecting the best cluster method by using the size of the icd rate and the coefficient of determination.
   c. Displaying the best cluster results in the form of a map and analyzes the characteristics of each cluster that is formed.
4. Determine conclusion based on classification and clustering method.

4. MAIN RESULTS

A. Descriptive Statistics

The description of the research data is carried out descriptively on the survey data to determine the characteristics of the community recommendation menu in East Java according to B2SA standard. The frequency of food menu combinations recommended by the people in East Java is shown in Table 5.

Staple food	**Consumer number**	**Side dish**	**Consumer number**
Rice	1,062	Tempe	138
Corn rice	135	Tofu	108
Corn	34	Salted egg	76
Bread	10	Chicken	53
Cassava	8	Catfish	53
Fruits	**Consumer number**	**Vegetable**	**Consumer number**
Not consuming	391	Not consuming	445
Orange	71	Bean	99
Watermelon	65	Spinach	87
Durian	57	Swamp cabbage	78
Melon	54	Cabbage	51
Mango	54	Mushroom	50
	Other food	Consumer number	
	No food	454	
	Cassava chips	77	
	Banana chips	74	
	Rengginang	41	
	Puli chips	33	
	Sugarcane	25	

TABLE 5. Frequency of People's Favorite Food Menu in East Java

Table 5 shows that the percentage of combinations of people's favorite food menus in East Java that are classified as non-B2SA is 49.3%. This means that there are still many people in East Java whose favorite food menu is not following B2SA standard.

Another fact shows that East Java Province is one of the largest producers of agricultural, plantation, livestock, and fishery commodities in Indonesia, especially for food crops. The summary of the highest production of each type of local food commodity in East Java in each district and city according to the B2SA standard which includes staple foods, side dishes, fruits, and vegetables, respectively, is shown by Table 6, Table 7, Table 8, and Table 9.

No	Food commodity	Maximum value	Location
1	Rice	5,019,426.1	Lamongan
2	Corn	5,069,660	Tuban
3	Potatoes	2,149,387	Pasuruan

TABLE 6. Highest Production of Staple Food Types based on Food Commodities

No	Food commodity	Maximum value	Location	No	Food commodity	Maximum value	Location
1	Beef	108,262.4	Surabaya City	9	Shrimp	215,040	Banyuwangi
2	Kampong chicken	75,147.5	Tulungagung	10	Catfish (*Pangasius siluriformes*)	104,490	Tulungagung
3	Laying hen	184,693.6	Lamongan	11	Catfish (*Clarias siluriformes*)	204,010	Sidoarjo
4	Broiler	706,528.2	Jombang	12	Tilapia	159,920	Sidoarjo
5	Duck	18,258.7	Tulungagung	13	Carp	23,290	Gresik
6	Skipjack	34,080	Pacitan	14	Snapper	3,200	Bangkalan
7	Cob	114,300	Banyuwangi	15	Milkfish	938,990	Gresik
8	Tuna	21,760	Malang	16	Grouper	10,060	Lamongan

TABLE 7. Highest Production of Types of Side Dishes based on Food Commodities

Based on Table 6 and Table 7, the highest production yields for types of staple food commodities and side dishes tend to spread in several districts or cities in East Java. However, in the Table 8 and Table 9 related to the highest production yields for the types of fruit and vegetable commodities, the majority are in Pasuruan and Malang Regencies.

No	Food commodity	Maximum value	Location	No	Food commodity	Maximum value	Location
1	Cantaloupe	58,256	Jombang	15	Siamese Orange	1,901,466	Banyuwangi
2	Melon	96,340	Ngawi	16	Mango	2,470,106	Pasuruan
3	Watermelon	441,788	Jember	17	Mangosteen	78,766	Ponorogo
4	Strawberry	2,666	Batu	18	Passion Fruit	87,593	Pasuruan
5	Tomato	333,262	Malang	19	Jackfruit	298,813	Pasuruan
6	Avocado	1,088,938	Pasuruan	20	Pineapple	1,589,305	Kediri
7	Grape	3,433	Pasuruan	21	Papaya	515,733	Malang
8	Apple	2,896,662	Pasuruan	22	Banana	9,922,545	Malang
9	Starfruit	82,334	Blitar	23	Rambutan	265,405	Pasuruan
10	Duku	62,342	Malang	24	Salak	568,204	Malang
11	Durian	1,073,483	Pasuruan	25	Sapodilla	40,648	Pasuruan
12	Water Guava	110,744	Pasuruan	26	Soursop	163,286	Pasuruan
13	Guava	255,976	Lamongan	27	Coconut	449,860	Sumenep
14	Big Orange	149,380	Magetan				

TABLE 8. Highest Production of Types of Fruits based on Food Commodities

No	Food commodity	Maximum value	Location	No	Food commodity	Maximum value	Location
1	Spring Onion	988,537	Batu	12	Red Bean	1,991	Tuban
2	Shallots Leaf	1,730,608	Nganjuk	13	Long Bean	108,014	Malang
3	Garlic	12,643	Malang	14	Kale	114,705	Malang
4	Breadfruit	30,848	Malang	15	Cauliflower	209,700	Malang
5	Jengkol	6,594	Pasuruan	16	Cucumber	119,468	Malang
6	Melinjo	95,718	Pacitan	17	Cabbage	724,911	Malang
7	Petai	436,674	Pasuruan	18	Chayote	164,052	Malang
8	Mustard Greens	358,045	Malang	19	Eggplant	254,971	Malang
9	Spinach	58,281	Malang	20	Carrot	226,508	Malang
10	Beans	104,113	Malang	21	Soybean	446,360	Banyuwangi
11	Seaweed	6,704,530	Sumenep	22	Mushroom	49,224.52	Malang

TABLE 9. Highest Production of Types of Vegetables based on Food Commodities

B. Classification Results

1. Binary Logistics Regression

- Simultaneous Parameter Significance Test

The parameter significance test is simultaneously carried out to determine whether the predictor variables had a significant effect on the model. The following are the results of the simultaneous parameter significance test for the factors that are thought to categorize the combination of food menu recommended by the community in East Java.

χ^2	df	$\chi^2_{(0.05;4)}$	p-value	Decision	Information
821.113	4	9.4877	0.000	Reject the H_0	Significance

TABLE 10. Simultaneous Parameter Significance Test Results

Table 10 shows that the value of $\chi^2_{count} = 821.113$ is bigger than $\chi^2_{(0.05;4)} = 9.4877$ or p-value $= 0.000$ is less than $\alpha = 0.05$. Thus, decided to reject the H_0 which means that there is at least one predictor variable that has a significant effect on the model.

- Partial Significance Parameter Test

A partial test is conducted to determine the predictor variables that have a significant effect on the response variable.

Variable	Coef	Wald	df	p-value	Decision	Information
Side dish	0.004	17.357	1	0.000	Reject the H_0	Significance
Fruit	0.036	73.618				
Vegetable	0.143	136.047				
Other food	0.008	104.735				
Constant	- 6.932	150.945				

TABLE 11. Partial Significance Parameter Test Results

Based on Table 11, at a significance level of 5%, all variables partially have a significant effect on the categorization of food combinations according to the B2SA standard. This is evidenced by the value of p-value less than α, or the value of Wald for each variable more than $Z_{0.025}$(1.96).

- Model Fit Test

The suitability test of this model using Deviance Test is carried out to test whether the resulting model is feasible.

Step	χ^2	df	$\chi^2_{(0,05;df)}$	p-value	Decision	Information
1	361.88	8	1,363.31	1.000	Accept the H_0	Model fitted

TABLE 12. Model Conformity Test Results (Deviance Test)

Based on Table 12 shows that the value of $\chi^2_{count} = 361.88$ is less than $\chi^2_{(0.05;8)} = 1{,}363.31$ or p-value $= 1.000$ is more than $\alpha = 0.05$ thus decided to accept the H_0 which means that the model fits or there is no significant difference between the observed results and the possible predictions of the model. The goodness of the model obtained in the analysis is used to determine the extent to which the predictor variables can explain the model formed by looking at the Negalkerke R-square (R-Sq) value. The following is the R-sq value of the formed model. In this study, the Negalkerke R-square value was 82.2%, which means the model can be explained by predictor variables of 82.2%, while 17.8% is explained by other predictor variables outside of this study.

- Odds Ratio (OR)

OR is the value of the tendency between one category and another on the qualitative explanatory variable. The trend ratio value can be seen in Table 13 as follows:

Variable	Odds Ratio
Calorie of side dish	1.004
Calorie of fruits	1.037
Calorie of vegetable	1.154
Calorie of other food	1.008

TABLE 13. Odds Ratio Value

The interpretation of the OR value is as follows:

a. The side dish calorie variable has an OR value = 1.004, meaning that for every 1 unit calorie increase in side dishes, the probability of the combination of food menus classified as B2SA increases by 0.4%.

b. The fruit calorie variable has an OR value of 1.037, meaning that for every increase in fruit calories by 1 calorie, the probability of a combination of food menus classified as B2SA increases by 3.7%.

c. The vegetable calorie variable has an OR value of 1.154, which means that for every increase in vegetable calories by 1 calorie, the probability of a combination of food menus classified as B2SA increases by 15.4%.

d. Other food calorie variables have an OR value = 1.008, meaning that for every increase in calories from side dishes of 1 calorie, the probability of combining food menus classified as B2SA increases by 0.8%.

- Model Classification Accuracy Test

The percentage of classification accuracy is the ratio between the number of observations correctly classified by the model and the total number of observations.

Actual	**Prediction**	
	Non- B2SA	**B2SA**
Non-B2SA	474	32
B2SA	34	330

TABLE 14. Model Classification Accuracy Test Results

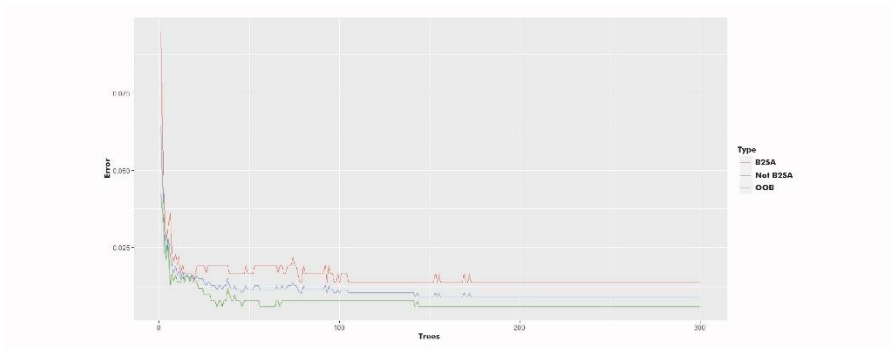
Table 14 shows that 474 data are classified correctly in the Non-B2SA category, while 32 data are incorrectly classified into the B2SA category. In addition, 330 data are properly classified in the B2SA category, while there are 34 data that are not classified as inappropriate in the Non B2SA category. Based on the calculation of the accuracy level or the accuracy of the model classification carried out using 1-APER, the classification accuracy value is 92.414%.

3. Random Forest

- Determination of Number of Trees (ntree) and Number of Variables (mtry) for Modeling based on Random Forest

Obtaining ntree and mtry values in random forest is based on the smallest OOB error value. Fig. 1 shows a graph of the OOB error value considering 300 trees and Table 15 shows the OOB error value of the three possible mtry values.

FIGURE 1. Graph of OOB Error Value to 300 Tree

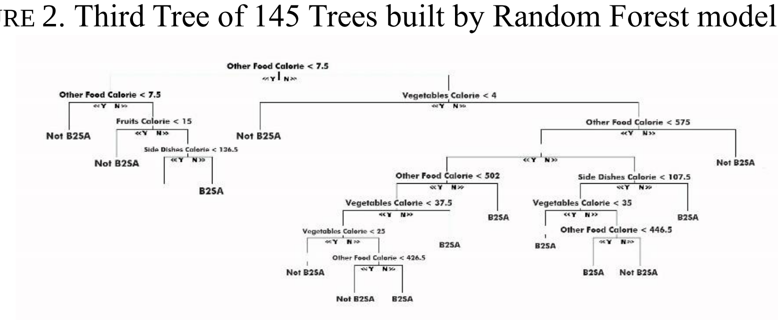


i	mtry$_i$	Error OOB value
1	1	0.0103
2	2	0.0091
3	4	0.0138

TABLE 15. Mtry Value and OOB Error Value

The blue line in Fig. 1 shows that the OOB error value begins to converge when using 145 trees so that the ntree value of 145 can then be used. While Table 15 shows that the smallest OOB error value is obtained when mtry = 2. Thus, it can be concluded that the modeling of the classification of food menu combinations can be built based on the random forest model with an ntree value of 145 and an mtry of 2. Fig. 2 shows one of the tree forms from the 145 trees that have been built by the random forest model with mtry = 2.

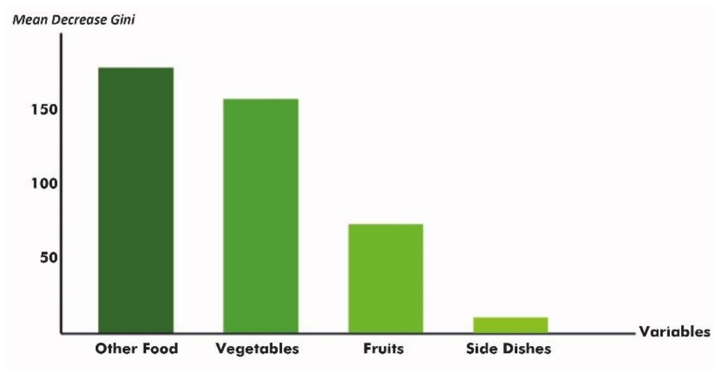
FIGURE 2. Third Tree of 145 Trees built by Random Forest model



- Variable Importance

The importance of the independent variable on the classification of food menu combinations can be detected from the Mean Decrease Gini value shown in Fig. 3.

FIGURE 3. Mean Decrease Gini



Based on Fig. 3, it can be concluded that other food variables are the most important variables in the classification of food menu combinations. This can be seen from the largest Mean Decrease Gini value and is evidenced in Fig. 3, where another food variable becomes the root node in the tree. Another food variable is a variable that contains various types of snacks consumed by people in East Java. The types of snacks consumed are very diverse, from high nutrition to low nutrition. Indonesia is a country with a high number of snack consumers, and only 2% of them choose healthy snacks [26]. In fact, the type of snack consumed has a significant effect on nutritional status in the body [23].

On the other hand, based on Fig. 3, vegetables and fruit variables have a more important role than side dishes in classifying food menu combinations. Many studies show that adequate intake of vegetables and fruit is the most important part of a healthy life [27]. Vegetables and fruit can also improve the nutritional status of the body and reduce the risk of chronic diseases [28].

- Model Evaluation

The analysis step then proceeds to the evaluation of the model. The evaluation of the model in the random forest model is carried out based on the confusion matrix on Table 16.

Actual	**Prediction**	
	B2SA	**Non-B2SA**
B2SA	359	5
Non-B2SA	3	503

TABLE 16. Confusion Matrix in Random Forest Classification Model

Based on the confusion matrix and equation (1), the classification accuracy value of the random forest model can be calculated as follows:

$$1 - APER = 1 - \frac{3+5}{359+5+3+503} = 0.990805 = 99.0805\% \quad (8)$$

From the calculation in the equation, it can be concluded that the accuracy value of the classification of B2SA food potential based on the random forest model is 99.0805%. This classification accuracy value is then compared with other classification models.

4. Classification Performance Comparison

Classification performance is generally measured by the value of classification accuracy (1 – APER), where the greater the value of classification accuracy, the better the classification accuracy. Table 17 shows the comparison of the classification accuracy values of the two classification methods that have been analyzed.

Classification Model	**1 - APER**
Logistic Regression	92.414%
Random Forest	99.081%

TABLE 17. Performance Comparison of the Two Classification Methods

Based on Table 17, it can be seen that the classification results using random forest have a classification accuracy value that is closest to 100%. Thus, it can be concluded that among the three classification methods, random forest provides the most accurate results in categorizing food menu combinations in East Java according to B2SA standard.

C. Clustering Results

1. Data Reduction with Principal Component Analysis (PCA)

The overall variables which are food commodities consist of 68 variables with details of 3 main food commodities, 16 side dishes consisting of 5 livestock products and 11 marine and aquaculture commodities, 27 fruit commodities, and 22 vegetable commodities. However, data with high dimensions can result in decreased classification accuracy and cluster quality [29]. Thus, it is necessary to reduce data through PCA which involves a correlation matrix between variables through the magnitude of Measures of Sampling Adequacy (MSA). In this case, PCA was performed on the category of fruits and vegetables that had the most variables. The MSA results for the categories of fruits and vegetables are shown in Table 18 and Table 19, respectively:

No	Variable	Correlation	No	Variable	Correlation	No	Variable	Correlation
1	Cantaloupe	0.134	10	Duku	0.616	19	Jackfruit	0.734
2	Melon	0.085	11	Durian	0.68	20	Pineapple	0.137
3	Watermelon	0.171	12	Water Guava	0.77	21	Papaya	0.535
4	Strawberry	0.345	13	Guava	0.463	22	Banana	0.625
5	Tomato	0.435	14	Big Orange	0.015	23	Rambutan	0.734
6	Avocado	0.676	15	Siamese Orange	0.324	24	Salak	0.627
7	Grape	0.883	16	Mango	0.683	25	Sapodilla	0.623
8	Apple	0.639	17	Mangosteen	0.375	26	Soursop	0.605
9	Starfruit	0.563	18	Passion Fruit	0.596	27	Coconut	0.117

TABLE 18. MSA Results for Fruits Category

No	Variable	Correlation	No	Variable	Correlation	No	Variable	Correlation
1	Spring Onion	0.339	9	Spinach	0.673	17	Cabbage	0.646
2	Shallots Leaf	0.53	10	Bean	0.626	18	Chayote	0.731
3	Garlic	0.729	11	Seaweed	0.172	19	Eggplant	0.878
4	Breadfruit	0.809	12	Red Bean	0.121	20	Carrot	0.545
5	Jengkol	0.496	13	Long Bean	0.644	21	Soybean	0.159
6	Melinjo	0.442	14	Kale	0.702	22	Mushroom	0.943
7	Petai	0.351	15	Cauliflower	0.607			
8	Mustard Greens	0.617	16	Cucumber	0.629			

TABLE 19. MSA Results for Vegetables Category

Based on Table 18 and Table 19, variables with MSA values > 0.5 can be analyzed further, while variables with MSA < 0.5 cannot be continued for further analysis. Thus, the research variables that can be used in further analysis consist of 50 variables with details of 3 staple food commodities, 16 side dishes consisting of 5 livestock products and 11 marine and aquaculture commodities, 16 fruit commodities, and 15 vegetable commodities.

2. Grouping of Regencies and Cities in East Java based on Local Food Commodities

Before grouping districts and cities in East Java based on their local food commodities, it is necessary to do a calculation to determine the optimal number of clusters for each type of food. The optimal cluster in this study was selected based on the largest Pseudo-F statistic value from each method used, namely Single Linkage and K-Means. Then, from the two cluster methods, one of the best cluster methods was selected based on the smallest icd rate value for further interpretation. Calculation of the optimal number of clusters of staple foods, side dishes, vegetables, and fruits, respectively, is shown in Table 20.

Category	**Method**	**Clus-ter**	R^2	**icd rate**	**Pseudo-F**	**Category**	**Method**	**Clus-ter**	R^2	**icd rate**	**Pseudo-F**
Staple food	Single	2	0.184	0.816	8.133	Fruit	Single	2	0.926	0.074	447.255
		3	0.224	0.776	5.039			3	0.949	0.051	327.263
		4	0.492	0.508	10.962			4	0.960	0.040	273.759
		5	0.496	0.504	8.128			5	0.973	0.027	301.287
	K-Means	2	0.525	0.475	39.800		K-Means	2	0.876	0.124	254.323
		3	0.658	0.342	33.700			3	0.205	0.795	4.502
		4	0.795	0.205	43.889			4	0.189	0.811	2.642
		5	0.764	0.236	26.642			5	0.159	0.841	1.565
Side dish	Single	2	0.427	0.573	26.823	Vegetable	Single	2	0.552	0.448	44.374
		3	0.624	0.376	29.039			3	0.756	0.244	54.173
		4	0.759	0.241	35.769			4	0.904	0.096	106.303
		5	0.781	0.219	29.475			5	0.923	0.077	99.277
	K-Means	2	0.098	0.903	3.858		K-Means	2	0.948	0.052	657.268
		3	0.138	0.862	2.807			3	0.583	0.417	24.513
		4	0.307	0.693	5.013			4	0.291	0.708	4.656
		5	0.389	0.611	5.250			5	0.031	0.272	0.264

TABLE 20. Optimal Cluster Number Selection

Based on Table 20, the optimal number of clusters for each type of commodity is shown in Table 21 below:

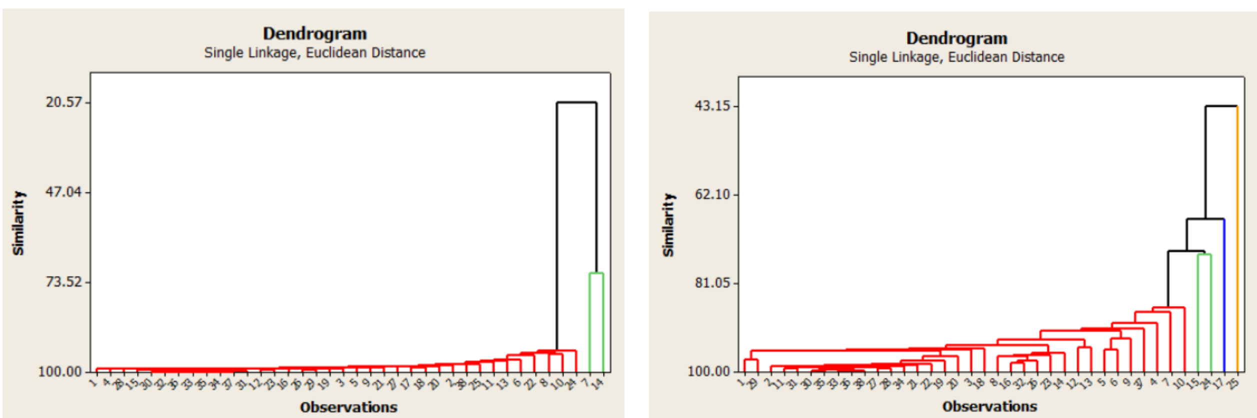
Type	Method	Cluster
Staple food	K-Means	4
Side dish	Single Linkage	4
Fruit	Single Linkage	2
Vegetables	K-Means	2

TABLE 21. Selection of Optimal Cluster Amount from All Types of Commodities

Based on Table 21, a dendogram is generated for the hierarchical method of each optimal cluster method which shows the grouping of foodstuffs based on districts and cities in East Java. The optimal cluster dendogram of staple foods, side dishes, and fruits is shown in Fig. 4.

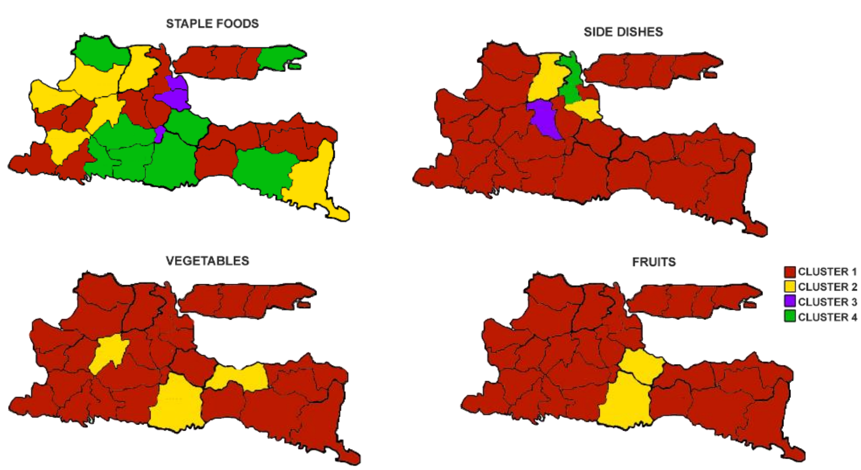
FIGURE 4. Optimal Cluster Dendrogram

(a) Fruit dendogram (b) Side-dish Dendogram



The results of the grouping are presented in the form of a map presented in Fig. 5 as follows:

FIGURE 5. Optimal Cluster Map



Based on Fig. 5, there are four colors that show the grouping of districts and cities in East

Java based on food potential. The red color shows the grouping of region I based on the category of food potential. The yellow color shows the grouping of region II based on the category of food potential. The purple color shows the grouping of Region III based on the category of food potential. The green color shows the grouping of region IV based on the category of food potential.

3. Characteristics of Clusters of Local Food Commodities in East Java

Based on the cluster analysis in the previous section, it was found that the level of food potential in each district and city is very diverse, which is indicated by the final cluster means value. One of the efforts to realize food security in Indonesia can be done by optimizing the potential of each region, especially in East Java. By identifying the characteristics of each cluster of food potential levels, food security can be carried out effectively and on target. The characteristics of each cluster for each food category are presented in Table 22, Table 23, Table 24, and Table 25.

Group I Members: Pacitan, Trenggalek, Lumajang, Bondowoso, Situbondo, Probolinggo, Mojokerto, Jombang, Madiun, Magetan, Gresik, Bangkalan, Sampang, Pamekasan Characteristic: 1. The region that produces rice and corn is the third highest out of the four existing clusters for East Java Province. 2. The second-highest potato-producing area of the four existing clusters is for East Java Province.
Group II Members: Ponorogo, Banyuwangi, Ngawi, Bojonegoro, Lamongan Characteristic: 1. The highest rice-producing area is in East Java. 2. The second highest corn producing area in East Java 3. The lowest potato-producing area is in East Java, even in some districts there is no recorded production.
Group III Members: Kediri City, Blitar City, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City, Batu City Characteristic: 1. The lowest producer of rice and corn commodities in East Java 2. The third-highest potato-producing area of the four clusters in East Java.
Group IV Members: Tulungagung, Blitar, Kediri, Malang, Jember, Pasuruan, Tuban, Sumenep Characteristic: 1. The region that produces the highest corn and potato commodities in East Java 2. The second highest rice producing area in East Java

TABLE 22. Cluster Characteristics by Staple Food Category

Group I Members: Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, Kediri City, Blitar City, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City, Batu City Characteristic: 1. This region is the lowest producer of fisheries and aquaculture in East Java. 2. This region is the second lowest livestock producer in East Java.
Group II Members: Sidoarjo, Lamongan Characteristic: 1. The total production of livestock products such as cattle, native chickens, laying hens, ducks, and fishery products such as catfish, tilapia, snapper, and grouper in this region is the highest in East Java. 2. The production of carp and milkfish is quite high in this area. 3. This area does not record skipjack and tuna production.
Group III Member: Jombang Characteristic: 1. The number of broiler production in this area is very high and the number of cattle production in this area is quite high. 2. This area does not record the production of skipjack, cob, tuna, shrimp, snapper, milkfish and grouper.
Group IV Member: Gresik Characteristic: 1. The area with the highest fishery and aquaculture production but the lowest livestock production is in East Java. 2. The amount of milkfish production is highest in East Java, even the highest compared to other commodities. 3. There is no recorded production of skipjack, tuna, and snapper.

TABLE 23. Cluster Characteristics by Category of Side dishes

Group I Members: Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Mojokerto, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, Kediri City , Blitar City, Malang City, Pasuruan City, Probolinggo City, Mojokerto City, Madiun City, Surabaya City, Batu City, Sidoarjo, Lamongan, Jombang, Gresik Characteristic: 1. The vegetable producing area is quite large in East Java but not as big as in Group II. 2. The majority of areas do not record the production of certain fruits.
Group II Members: Malang, Pasuruan Characteristic: 1. The area with the highest and varied fruit production yields is in East Java.

TABLE 24. Cluster Characteristics by Fruit Category

Group I Members: Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Mojokerto, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, Kota Kediri, Kota Blitar, Malang City, Pasuruan City, Probolinggo City, Mojokerto City, Madiun City, Surabaya City, Batu City, Sidoarjo, Lamongan, Jombang, Gresik, Pasuruan Characteristic: 1. The vegetable producing area is quite large in East Java but not as big as in Group II. 2. The majority of areas do not record any particular vegetable production.
Group II Members: Malang, Probolinggo, Nganjuk Characteristic: 1. The largest vegetable producing area in East Java with the most varied types.

TABLE 25. Cluster Characteristics by Vegetable Category

5. CONCLUSION

During the COVID-19 pandemic, it can be seen that there are still food menus recommended by East Java residents that do not meet B2SA standard. In order to maximize the B2SA movement in Indonesia, especially during the COVID-19 pandemic, it is necessary to have a massive B2SA campaign by providing information in the form of food menu categories according to B2SA standard which is the result of the performance of the best classification method in this study, namely random forest that is included in supervised learning. Of course, the food menu consumed by the community is a food potential that must be maximized for production in each region. In this study, it is found that the Single Linkage hierarchical cluster method shows maximum performance in grouping side dishes and fruits based on the regions in East Java, while the K-Means non-hierarchical cluster method gives the best performance in classifying staple foods and vegetables based on the regions in East Java. Therefore, to maximize the production of each type of potential food in each region that has been grouped based on the results of the performance of the two cluster methods which include unsupervised learning. In addition, synergy is needed between the community and the Indonesian government to maximize the B2SA movement in the COVID-19 pandemic so that the SDGs targets can be achieved.

CONFLICT OF INTERESTS

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

REFERENCES

[1] D. Laborde, W. Martin, J. Swinnen, R. Vos, COVID-19 Risks to Global Food Security, Science 369(6503) (2020), 500-502.

[2] Central Bureau of Statistics. East Java in Numbers 2020. Central Bureau of Statistics, East Java, (2020).

[3] S. Hartati, I. N. Tari, R. Widyastuti, The Development of Diverse, Nutritious, Balanced and Safe (B2SA) Formulation Menu Based on Local Food in Sukoharjo, Proceedings of the International Conference on Community Development (ICCD 2020), 2020.

[4] R. Johnson, D. Wichern, Applied Multivariate Statistical Analysis 6th Edition, Pearson Hall International Inc, New Jersey, 2007.

[5] L. Breiman, Random Forests, Mach. Learn. 45(1) (2001), 5-32.

[6] M. F. F. Mardianto, A. S. Syamsudin, D. A. Husada, R. Andriawan, Indonesia's Provinces Clustering Based on Industrial Concentration as a Reference in Establishing Vocational Programs for Vocational High Schools, Int. J. Innov. Creat. Change, 5(3) (2019), 917-935.

[7] S. P. Rahayu, S. Febryani, H. B Kusuma, Suhartono, D. D. Prasetyo, The Approach of Fuzzy C-Means Cluster

and Factor Analysis on Economic Potential Mapping of Regency / City in East Java Province, International Conference on Science and Applied Science (ICSAS), 2018.

[8] A. Lawi, S. L. Wungo, S. Manjang, Identifying Irregularity Electricity Usage of Customer Behaviors Using Logistic Regression and Linear Discriminant Analysis. 3rd International Conference on Science in Information Technology (ICSITech), 2017.

[9] M. Liu, M.Wang, J. Wang, D. Li, Comparison of Random Forest, Support Vector Machine and Back Propagation Neural Network for Electronic Tongue Data Classification: Application to the Recognition of Orange Beverage and Chinese Vinegar. Sensors Actuators B: Chem. 177 (2013), 970-980.

[10] K. Van Mierlo, K. De Ridder, A. Geeraerd, Identifying Belgian protein consumption typologies by means of clustering and classification to move towards personalized advices for sustainable and nutritious food choices. Appetite 166 (2021), 105583.

[11] J. A. Hartigan, Consistency of Single Linkage for High-Density Clusters, J. Amer. Stat. Assoc. 76(374) (2012), 388-394.

[12] G. A. Wilkin, X. Huang, K-Means Clustering Algorithms: Implementation and Comparison, Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007), 2007, 133-136.

[13] J. W. Osborne, Best Practices in Quantitative Methods, Sage Publications, Inc, California, 2008.

[14] D. W. Hosmer, S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression, John Wiley and Sons Inc, Canada, 2000.

[15] L. Breiman, A. Cutler, Manual--Setting Up, Using, and Understanding Random Forest V4.0, 2003.

[16] M. Schonlau, R. Y. Zou, The random forest algorithm for statistical learning, Stata J.: Promot. Commun. Stat. Stata, 20(1) (2020), 3-29.

[17] J. F. Hair, W. C. Black, B. J. Babin, R. E. dan Anderson, Multivariate Data Analysis Eighth Edition. Cengage Learning, Hampshire, 2019.

[18] S. T. Wierzchon, M. A. Klopotek, Modern Algorithms of Cluster Analysis, Studies in Big Data, Vol. 34, Springer, Gewerbestrasse, 2018.

[19] A. Afifi, S. May, R. A. Donatello, V. A. Clark, Practical Multivariate Analysis Sixth Edition, Taylor & Francis Group, Boca Raton, 2020.

[20] A. Kassambara. Practical Guide to Cluster Analysis in R (Unsupervised Machine Learning), STHDA, (2017).

[21] M. Mazziotta, A. Pareto, Use and Misuse of PCA for Measuring Well-Being, Soc. Indic. Res. 142 (2019), 451–476.

[22] H. Abdi, L. J. Williams, Principal component analysis. Wiley Interdiscip. Rev.: Comput. Stat. 2(4) (2010), 433-459.

[23] D. Susilastuti, A. Meri, L. S. Marhaeni, Functional Food Support Government Policies to Prevent Stunting. Palu, Proceeding International Conference of Innovation Science, Technology, Education, Children and Health (iCISTECH 2021), 2021.

[24] D. K. Wardhani, R. M. Raharja, R. Kusumawardani, Are the Early Children's School Meals Diverse, Nutritious, Balanced, and Safe?, Proceedings of the 2nd Early Childhood and Primary Childhood Education (ECPE 2020), 2020.

[25] V. Rawal, V. Bansal, P. Bansal, Prevalence of Undernourishment in Indian States: Explorations Based on NSS 68th Round Dat, Econ. Political Weekly, 54(15) (2019), 35-45.

[26] Mondelez, Snacking Habit Report Indonesia, https://observerid.com/mondelez-internationals-studies-on-snacking-habits-and-trends-in-indonesia/. (2017).

[27] O. O. Oguntibeju, E. J. Truter, A. Esterhuyse, The Role of Fruit and Vegetable Consumption in Human Health and Disease Prevention, Health and Disease Prevention, Diabetes Mellitus - Insights and Perspectives, 2013.

[28] D. Pem, R. Jeewon, Fruit and Vegetable Intake: Benefits and Progress of Nutrition Education Interventions-Narrative Review Article, Iran. J. Public Health, 44(10) (2015), 1309-1321.

[29] M. Ravichandran, M. Ganesh, G. Sridhar, A Study of Adaptive Cluster Distance Bounding for High-Dimensional Indexing, Cikitusi J. Multidiscip. Res. 6 (5) (2019), 57-61.