INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Machine Learning Algorithms Based on Sampling Techniques for Raisin Grains Classification

Achmad Bisri [a,b,*], Mustafa Man [c]

[a] *Division of Information Technology, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Serang, 42118, Indonesia*
[b] *Department of Informatics Engineering, Universitas Pamulang, Tangerang Selatan, 15417, Indonesia*
[c] *Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, 21030, Malaysia*
*Corresponding author: *achmadbisri@uinbanten.ac.id*

***Abstract*— Raisin grains are among the agricultural commodities that can benefit health. The production of raisin grains needs to be classified to achieve optimal results. In this case, the classification is carried out on two types of grains, namely Kecimen and Besni. However, inaccurate sample data can affect the performance of the model. In this study, two sampling techniques are proposed: stratified and shuffled. The proposed classification model is RF, GBT, NB, LR, and NN. This study aims to identify the performance of classification models based on sampling techniques. Classification models are applied to the seven-features dataset, and modeling is done by cross-validation. The results of the models were tested with a different amount of test data. The performance of the models was evaluated related to accuracy and AUC. The best outcomes of all models based on stratified sampling were founded on tested data of 40 percent with a mean accuracy of 85.50% and an AUC of 0.921. In comparison, models based on shuffled sampling were founded on test data of 20 percent with a mean accuracy of 88.11% and an AUC of 0.935. On the other hand, classification models based on a stratified sampling of all data splits do not all models generate an excellent category. Whereas, based on shuffled sampling, all models resulted in the excellent category. Therefore, models based on shuffled sampling are superior to stratified sampling. The result of the significant test, RF, significantly differs based on sampling techniques.**

***Keywords*— Classification; data mining; machine learning; raisin grains; sampling techniques.**

Manuscript received 17 Jun. 2022; revised 24 Oct. 2022; accepted 28 Nov. 2022. Date of publication 31 Mar. 2023.

I. INTRODUCTION

Raisin (dried grape) is a dried fruit with fiber and high nutritional value. Raisin has a carbohydrate source that contains fiber, antioxidants, potassium, and iron. The benefits of consuming raisins can enhance health and prevent many chronic diseases like type 2 diabetes mellitus, cardiovascular disease, gastrointestinal disease, and dental caries [1]. The quality of the grapes can be determined by the characteristics of color, texture, aroma, phytochemicals, vitamins, and microbial stability [2]. Therefore, it is important to conduct research in order to obtain superior and high-quality raisin products [3].

Many researchers have already conducted studies on raisins and the like. In the production of raisins, the drying process is carried out using various techniques to preserve food [4], such as sunlight, oven, shade, and alkali. It can also be done with oil and temperature [5]. In addition, the drying process can be performed by pulsed vacuum drying (PVD) [6]. Raisin varieties were also identified using imaging [7]. Machine vision systems [8] with image processing are also performed to extract many raisin features such as size, color, texture, etc. A spectral extraction based on a resolution algorithm of the waveforms is proposed [9], and near-infrared spectra with machine learning [10] [11] [12], and neural network [13] [14] models are used for raisin classification. Raisin segmentation was also performed [15] using a random forest (RF), deep learning (DL), and support vector machine.

Karimi et al. [16] Used machine vision systems with image processing to identify the quality and purity of raisins. Machine learning algorithms are used to build classification models, namely the support vector machine (SVM) model and the neural network (NN) technique. Model performance is measured based on accuracy. In the results obtained, the SVM model has more efficient and effective results than the NN.

Zhao et al. [15] Used segmentation techniques for raisin extraction with aspects of shape features like roundness, area, *x* and *y*-axis values, length, and circumference of the axes. A machine learning algorithm approach is proposed to predict

the numbers of raisins, namely RF, SVM, and DL. Of the three proposed algorithms, the deep learning model has better performance than other models.

Khojastehnazhand et al. [8] Studied the quality of raisins in bulk comprising a combination of good and poor raisins with wood. The data model used is divided into two modes, namely mode 1 with 6 classes and mode 2 with 15 classes. The proposed classification methods for the supervised learning category are the SVM model and linear discriminant analysis (LDA) algorithm. The outcome shows that the SVM model with the grey level run length matrix (GLRM) performs better and more accurately.

Cinar et al. [17] Developed a system for distinguishing varieties of raisin grown in Turkey with two classifications, namely Kecimen and Besni. The image processing technique is used for feature extraction of the morphological features of the image of raisins. The morphological features used are similar to the classification of rice varieties [18]. Statistical information is also used for each feature with minimum, maximum, average, and standard deviation values. From the extraction of morphological features, it produces a raisin dataset with seven features and one label as a class. Then, the classification models are built using machine learning algorithms, namely logistic regression (LR), multilayer perceptron (MLP), and SVM. The performance of the models that have been built is measured based on the level of accuracy. The result obtained with the highest accuracy of 86.44% is on the support vector machine model.

Tarakci [19] applies the KNN and WKNN models by using the Euclidean distance on the raisin distribution of the dataset which is normalized with min-max between the two classes. The model is formed by applying cross-validation up to 10-fold. The outcome of the confusion matrix from the performance of the model is measured and evaluated related to the accuracy, recall, sensitivity, error rate, specificity, and F1-score. The performance of the proposed model, KNN is better than WKNN in the raisin grains dataset. However, it is not as good as the SVM model in previous related studies.

In general, previous research has provided reports on image processing, segmentation, and feature extraction on raisin grains and method improvements that focus on accuracy. The methods used in previous research are popular methods used extensively to build models. In comparison, the sample used in the previous study was a raisin grains dataset obtained from various sources classified using two-class or multi-class. However, at the preprocessing level, the sample data's inaccuracy can affect the classification model's suboptimal performance.

In this study, the raisin grains dataset was used as a sample of publicly available data and was obtained from https://www.muratkoklu.com/datasets/. In addition, the sampling technique used is stratified and shuffled sampling. Whereas five popular machine learning algorithms are applied to build classification models, namely the random forest (RF) model, logistic regression (LR) algorithm, gradient boosted trees (GBT) algorithm, naive Bayes (NB), and neural network (NN). This study aims to accurately identify the performance of sampling techniques and classification models.

This paper is divided into four parts. To begin with the introduction in this section. In addition, it relates to methods. Moreover, it refers to results and discussions. Furthermore, in conclusion.

II. Material and Method

This section describes the materials and methods comprising a scheme of experimental settings, proposed method, dataset, performance evaluation, and significant test.

A. Scheme of Experimental Settings

This is the preferred setup employed by most of the modeling. Fig. 1 shows an overview of experimental settings. To begin with, the raisin grains dataset is split into testing and training data. In the distribution of training data using a sampling technique approach, namely stratified and shuffled sampling, to be trained with machine learning algorithms to build classification models. The data is divided into several parts with the training sequence (60, 70, 80, and 90) and testing (40, 30, 20, and 10) data in percent. In addition, the five machine learning algorithms, namely RF, GBT, NB, LR, and NN are applied to the training data in the modeling process.

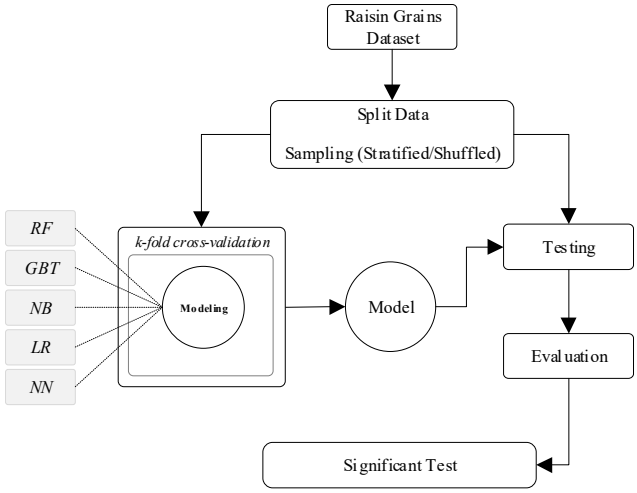


Fig. 1 Scheme of experimental settings

During modeling, validation is performed using *k*-fold cross-validation as shown in Fig. 2 in which the data is partitioned into *k* subsets with the same proportion. In this section, *k*=10 was used to perform cross-validation up to 10-fold. One subset is used as test data and the other subsets are combined into one as training data.

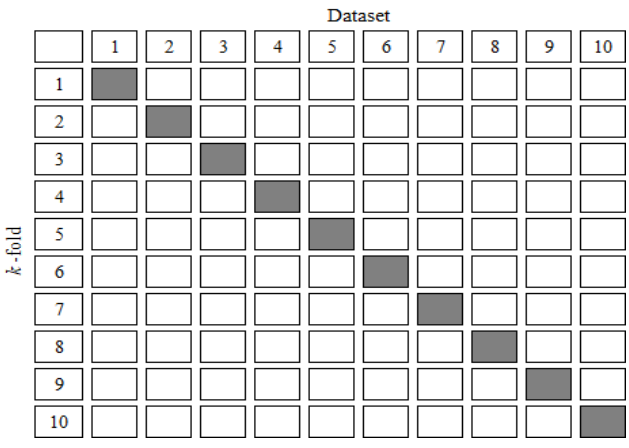


Fig. 2 Illustrates of *k*-fold cross-validation

The cross-validation procedure is repeated *k*-fold with each *k* subset in turn until the number of iterations is fulfilled. The accuracy results of these iterations are averaged. Furthermore, after the models are formed, the models are tested, and the test results are evaluated against performance indicators, namely accuracy, recall or sensitivity, specificity, precision or PPV, NPV, F-measure, and AUC. In conclusion, the resulting performance was statistically tested using a *t*-test to determine the significant difference between the classification models based on stratified and shuffled sampling and split data.

B. Sampling Techniques

Sampling is a process of taking or selecting a subset of data from a dataset to represent and use in data modeling by creating a representative model based on the sampled data. Sampling can have an impact on the relevance of model performance. In this study, samples were taken to build a model with two types of sampling, namely stratified and shuffled sampling.

1) Stratified Sampling: Stratified sampling is the process of selecting samples of the main dataset, which is carried out representatively on each part of the strata is carried out randomly, proportionally, and evenly, which is estimated based on the percentage of data distribution that is considered the same from the two sample classes for modeling data [20]. Fig. 3 is an illustration of stratified sampling, in which each sample is taken randomly on the same part of each stratum.

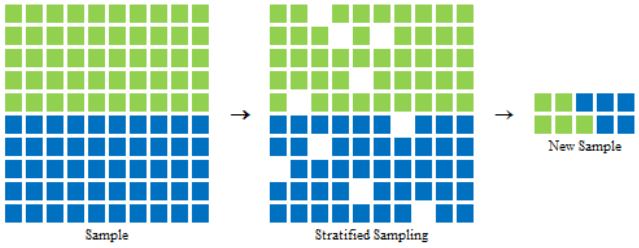


Fig. 3 Illustrates of stratified sampling

2) Shuffled Sampling: Shuffled sampling is the selection of a sample data subset of the main dataset, which is done randomly and forms a new set of data modeling. Fig. 4 is an illustration of shuffled sampling, where each sample is taken randomly without the need to look at the stratum status of each part of the sample.

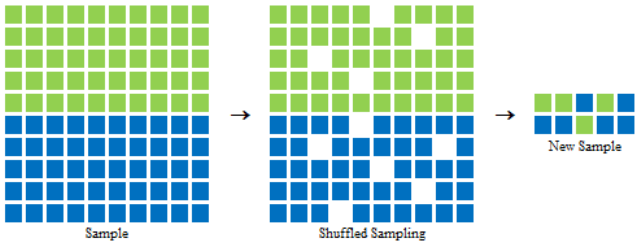


Fig. 4 Illustrates of shuffled sampling

C. Machine Learning Algorithms

1) The Random Forest (RF): RF model is a powerful machine learning method [21] which is also known as a method that has a similar learning concept to ensemble techniques that can reduce generalization errors. Random forest can solve classification and regression problems. Random forest in the decision tree classification can reduce overfitting and increase precision.

2) The Gradient Boosted Trees (GBT): GBT algorithm is a machine learning algorithm that is reliable [22] in building a decision tree model for classification and regression with boosting techniques. Boosting the GBT ensemble can improve the performance of the decision tree model by combining weak classifier outputs to produce a strong unity [23]. However, it has the disadvantage that it is difficult to determine the optimal parameter value and requires a longer processing time to obtain the optimal outcome.

3) The Naïve Bayes (NB): NB model is a probabilistic classification that uses Bayes' theorem [24], which can be used to predict a class as a probabilistic posterior class that has a lot of information whose modeling uses a multivariate normal distribution. The NB formulation (1) is as follows.

$$P(Y|X) = \frac{P(Y)*P(X|Y)}{P(X)} \quad (1)$$

This assumes that X is a set attribute. Y is assumed to be the label/target/class of a result. $P(Y)$ is the prior likelihood of the training set. $P(Y|X)$ is a conditional likelihood or called a posterior probability. $P(X|Y)$ is a conditional class likelihood determined from an outcome. $P(X)$ is the likelihood of evidence that has the same class value as the outcome.

4) Logistic Regression (LR): LR model is an algorithm that uses a logistic function with probabilistic modeling related to binary [25]. The logistic formulation (2) is as follows.

$$Log\ (p/(1-p)) = \alpha + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn \quad (2)$$

Where $p/(1\text{-}p)$ is the likelihood of an event. P is interpreted as the likelihood of the event occurring and 1-p is interpreted as the probability that the event not occurring. While α represents an intercept, β represents the regression coefficient, and X represents the independent variable.

5) Neural Network: NN model is an algorithm that imitates the architecture of the biological process of neurons that connects the function between input and output variables. A neural network has a simple mathematical model [26] with the following formulation (3).

$$Y = 1 + 2X1 + 3X2 + 4X3 + \ldots \quad (3)$$

Where Y is an output from the calculation of input attributes X_1, X_2, and X_3. The value of 1 is an intercept. Whereas 2, 3, and 4 are coefficients for the inputs of X_1, X_2, and X_3.

D. Dataset

The public dataset used in this study was obtained from https://www.muratkoklu.com/datasets/. The raisin grains dataset has seven features of numeric type, namely area, perimeter, major_axis, minor_axis, eccentricity, convex_area, and extent. This dataset has a balanced classification and is organized into groups of two classes containing a total of 900 instances consisting, of which 450 are Kecimen, and 450 are Besni classifications.

E. Performance Evaluation

In this section, the performance evaluation is carried out as an evaluation of the classification model using the outcome of the confusion matrix presented in Table I below.

TABLE I
THE CONFUSION MATRIX

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

Table 1 above consists of TP, TN, FP, and FN. True positive (TP): the class predicted to be Kecimen is actually Kecimen, false positive (FP): the class predicted to be Kecimen is actually Besni, false negative (FN): the class predicted to be Besni is actually Kecimen, true negative (TN): the class predicted to be Besni is actually Besni. The formula for evaluation [27] related to the results of the confusion matrix is accuracy (Acc), AUC, F-measure (F), sensitivity (SN), positive predictive value (PPV), specificity (SP), and negative predictive value (NPV).

$$Accuracy\ (Acc) = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

$$Recall = SN = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity\ (SP) = \frac{TN}{TN+FP} \quad (6)$$

$$Precision = PPV = \frac{TP}{TP+FP} \quad (7)$$

$$NPV = \frac{TN}{TN+FN} \quad (8)$$

$$F-measure\ (F1) = \frac{2TP}{2TP+FP+FN} \quad (9)$$

$$\text{AUC} = \frac{1+TP_{rate}-FP_{rate}}{2} \quad (10)$$

- Accuracy (Acc): The total number of cases accurately estimated (4).
- Sensitivity (SN) or Recall: The proportion of positive rates on the classification that was correctly predicted (5).
- Specificity (SP): The proportion of negative rates on the classification that was correctly predicted (6).
- Precision (PPV): The proportion of a total number of correctly categorized positive classes predicted (7).
- Negative Predictive Value (NPV): The proportion of a total number of correctly categorized negative classes predicted (8).
- F-measure: The harmonic average of PPV and SN (9).
- Area Under Curve (AUC): An indicator used to measure and evaluate a model with a two-dimensional curve that depicts the balance between true positives and false positives (10).

Table II shows the success criteria of a classification model that is determined related to the area under the curve (AUC) with a two-dimensional area [28], and the AUC value is in the range of 0.50 – 1.00. Where the higher the AUC value, the better the model's performance [29].

TABLE II
THE CRITERIA OF CLASSIFICATION

Category	Ranges
Excellent Classification	0.9 up to 1.0
Good Classification	0.8 up to 0.9
Fair Classification	0.7 up to 0.8
Poor Classification	0.6 up to 0.7
Failure	0.5 up to 0.6

F. Significant Test

In this section, hypothesis testing is carried out on the performance of the resulting model using a *t*-Test: paired two samples for means based on the *p*-value. The threshold determined for the *p*-value in this study is 0.05. There are two possibilities in the hypothesis test: the null hypothesis (H_0) and the alternative hypothesis (H_a). In H_0 there is no meaningful difference between the two paired samples and on the contrary, H_a there is a significant difference and rejects H_0. By assuming reject H_0, then H_0 is not valid [30].

III. Results And Discussion

In this study, the computing platform used has the following specifications, namely Intel® Core™ i3 2.00GHz (4 CPUs), 8-gigabyte memory, Windows 10 64-bit, and RapidMiner version 9.10.001 as a data analytics. The outcomes of this experiment are summarized in several tables, and Table III shows the test results based on the stratified sampling technique.

TABLE III
THE RESULT OF THE CONFUSION MATRIX BASED ON STRATIFIED SAMPLING

Tests	Models	TP	FP	FN	TN	SN	SP	PPV	NPV	F	ACC	AUC
40%	RF	164	41	16	139	91.11	77.22	80.00	89.68	85.19	84.17	0.916
	GBT	158	33	22	147	87.78	81.67	82.72	86.98	85.18	84.72	0.905
	NB	175	50	5	130	97.22	72.22	77.78	96.30	86.42	84.72	0.922
	LR	164	31	16	149	91.11	82.78	84.10	90.30	87.47	86.94	0.933
	NN	166	33	14	147	92.22	81.67	83.42	91.30	87.60	86.94	0.927
		Mean				91.89	79.11	81.60	90.91	86.37	85.50	0.921
30%	RF	117	30	18	105	86.67	77.78	79.59	85.37	82.98	82.22	0.903
	GBT	108	20	27	115	80.00	85.19	84.38	80.99	82.13	82.59	0.894
	NB	130	34	5	101	96.30	74.81	79.27	95.28	86.96	85.56	0.911
	LR	119	24	16	111	88.15	82.22	83.22	87.40	85.61	85.19	0.921
	NN	121	25	14	110	89.63	81.48	82.88	88.71	86.12	85.56	0.913
		Mean				88.15	80.30	81.87	87.55	84.76	84.22	0.908
20%	RF	74	21	16	69	82.22	76.67	77.89	81.18	80.00	79.44	0.885
	GBT	74	21	16	69	82.22	76.67	77.89	81.18	80.00	79.44	0.878
	NB	86	26	4	64	95.56	71.11	76.79	94.12	85.15	83.33	0.887

Tests	Models	TP	FP	FN	TN	SN	SP	PPV	NPV	F	ACC	AUC
	LR	80	20	10	70	88.89	77.78	80.00	87.50	84.21	83.33	0.900
	NN	79	21	11	69	87.78	76.67	79.00	86.25	83.16	82.22	0.890
		Mean				87.33	75.78	78.32	86.04	82.50	81.56	0.888
	RF	40	13	5	32	88.89	71.11	75.47	86.49	81.63	80.00	0.870
	GBT	42	11	3	34	93.33	75.56	79.25	91.89	85.71	84.44	0.894
10%	NB	43	13	2	32	95.56	71.11	76.79	94.12	85.15	83.33	0.881
	LR	41	13	4	32	91.11	71.11	75.93	88.89	82.83	81.11	0.888
	NN	42	13	3	32	93.33	71.11	76.36	91.43	84.00	82.22	0.873
		Mean				92.44	72.00	76.76	90.56	83.86	82.22	0.881

Where the scores are calculated based on the confusion matrix, namely sensitivity, specificity, PPV, NPV, F-measure, accuracy, and AUC. This was done based on four distributions of model data with tested data 10, 20, 30, and 40 percent by applying the five proposed classification models. In the case sensitivity (recall) of test data, 10 and 40 percent, the result is greater than 90%. Whereas, using test data 20 and 30 percent scored less than 90%. The specificity, when tested data of 30 percent, gives a score that is greater than 80% from other test data. For precision (PPV) with tested data of 30 and 40 percent, the score is greater than 80%, and for other tests data, less than 80%. For the NPV, using 20 and 30 percent tested data gives a mean score of about 85%, and using 10, and 30 percent gives a score of about 90%. Whereas for F-measure, all the test data had a score greater than 80% and a mean score of 84%. The performance of the classification models is measured related to the accuracy of all tested data, giving a score greater than 80% with a mean of 83%. On the other hand, when tested data 40 and 30 percent for AUC, the model's performance yields a score greater than 0.9 in the excellent category. In comparison, the other test data with AUC scores are in a good category. The results given in Table IV show the experimental results based on the shuffled sampling technique.

TABLE IV
THE RESULT OF CONFUSION MATRIX BASED ON SHUFFLED SAMPLING

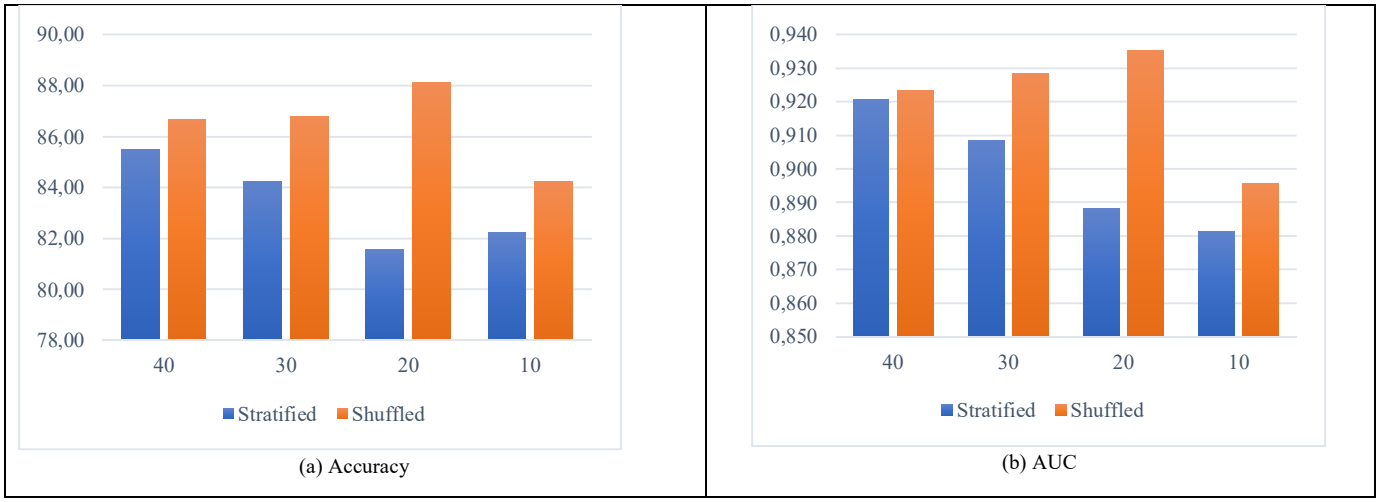
Tests	Models	TP	FP	FN	TN	SN	SP	PPV	NPV	F	ACC	AUC
	RF	169	27	17	147	90.86	84.48	86.22	89.63	88.48	87.78	0.925
	GBT	144	17	42	157	77.42	90.23	89.44	78.89	83.00	83.61	0.901
40%	NB	176	37	10	137	94.62	78.74	82.63	93.20	88.22	86.94	0.922
	LR	163	23	23	151	87.63	86.78	87.63	86.78	87.63	87.22	0.936
	NN	167	25	19	149	89.78	85.63	86.98	88.69	88.36	87.78	0.933
		Mean				88.06	85.17	86.58	87.44	87.14	86.67	0.923
	RF	120	18	19	113	86.33	86.26	86.96	85.61	86.64	86.30	0.934
	GBT	122	19	17	112	87.77	85.50	86.52	86.82	87.14	86.67	0.911
30%	NB	130	26	9	105	93.53	80.15	83.33	92.11	88.14	87.04	0.926
	LR	121	16	18	115	87.05	87.79	88.32	86.47	87.68	87.41	0.938
	NN	120	17	19	114	86.33	87.02	87.59	85.71	86.96	86.67	0.934
		Mean				88.20	85.34	86.55	87.34	87.31	86.81	0.929
	RF	82	11	10	77	89.13	87.50	88.17	88.51	88.65	88.33	0.942
	GBT	86	14	6	74	93.48	84.09	86.00	92.50	89.58	88.89	0.941
20%	NB	85	17	7	71	92.39	80.68	83.33	91.03	87.63	86.67	0.921
	LR	82	11	10	77	89.13	87.50	88.17	88.51	88.65	88.33	0.939
	NN	85	14	7	74	92.39	84.09	85.86	91.36	89.01	88.33	0.933
		Mean				91.30	84.77	86.31	90.38	88.70	88.11	0.935
	RF	40	8	8	34	83.33	80.95	83.33	80.95	83.33	82.22	0.909
	GBT	40	5	8	37	83.33	88.10	88.89	82.22	86.02	85.56	0.899
10%	NB	43	9	5	33	89.58	78.57	82.69	86.84	86.00	84.44	0.873
	LR	38	5	10	37	79.17	88.10	88.37	78.72	83.52	83.33	0.903
	NN	41	6	7	36	85.42	85.71	87.23	83.72	86.32	85.56	0.895
		Mean				84.17	84.29	86.10	82.49	85.04	84.22	0.896

As explained earlier that this value is generated from the confusion matrix with tested data of 10, 20, 30, and 40 percent. For sensitivity, the results greater than 90% are only on test data of 30 percent, other than the mean is about 87%. The results are almost identical for specificity and precision, which is about 85% and 86%. For NPV, it produces a score of 90% on test data of 20 percent, other than that it is less than 90%. For F-measure, the four tested data have a mean score of 87%. Likewise, the accuracy of the four tested data with a mean score of 86%. In addition, on the AUC score, all the mean classification results have an excellent category, and only one in the tested data of 10 percent has a good classification category.

Table V provides a summary of the accuracy, and Fig. 5 displays the accuracy means for models based on stratified and shuffled sampling. On stratified sampling, all models perform similarly, with a performance about of 83%. Whereas on shuffled sampling about 86%. The difference in the accuracy means is about 3%. On the other hand, shuffled sampling has the highest accuracy and area under the curve (AUC) on 20 percent of test data.

TABLE V
RESUME OF ACCURACY BASED ON CLASSIFICATION MODELS

Models	Stratified					Shuffled				
	40%	30%	20%	10%	*Mean*	40%	30%	20%	10%	*Mean*
RF	84.17	82.22	79.44	80.00	81.46	87.78	86.30	88.33	82.22	86.16
GBT	84.72	82.59	79.44	84.44	82.80	83.61	86.67	88.89	85.56	86.18
NB	84.72	85.56	83.33	83.33	84.24	86.94	87.04	86.67	84.44	86.27
LR	86.94	85.19	83.33	81.11	84.14	87.22	87.41	88.33	83.33	86.57
NN	86.94	85.56	82.22	82.22	84.24	87.78	86.67	88.33	85.56	87.08
Mean	85.50	84.22	81.56	82.22	83.38	86.67	86.81	88.11	84.22	86.45



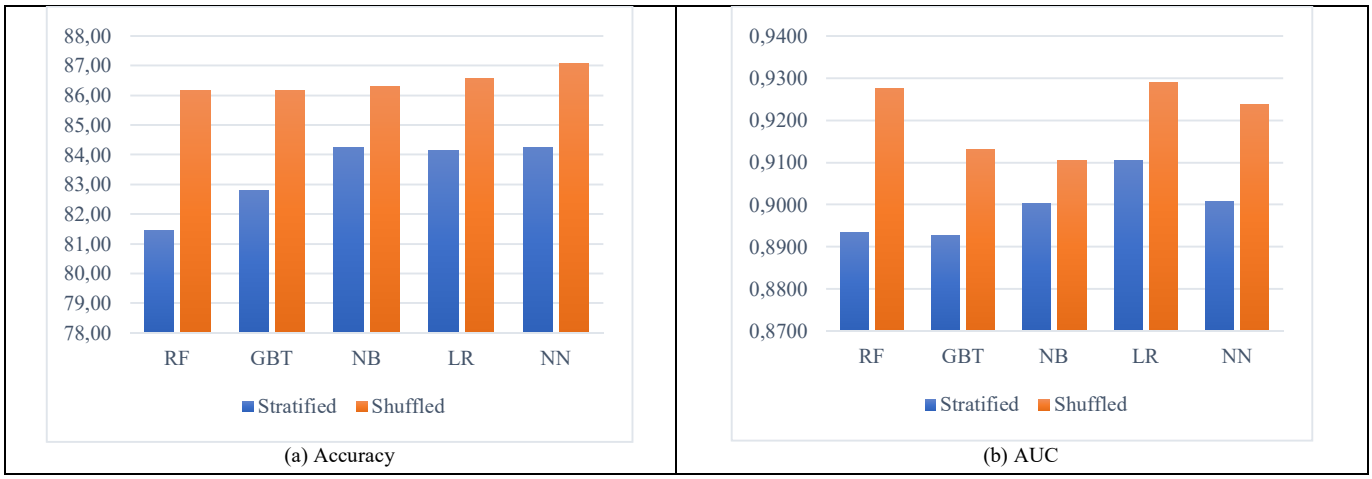
(a) Accuracy (b) AUC

Fig. 5 Accuracy and AUC of test data based on stratified and shuffled sampling

Table VI and Fig. 6 show the mean scores of AUC for classification models based on sampling techniques. On stratified sampling, there are three models whose performance is greater than 0.9, namely NB, LR, and NN. Next, the performance of all models in shuffled sampling is greater than 0.9 in the excellent category. The five classification models RF, GBT, NB, LR, and NN applied with shuffled sampling have superior performance compared to classification models using stratified sampling.

TABLE VI
RESUME OF AUC BASED ON CLASSIFICATION MODELS

Models	Stratified						Shuffled					
	40%	30%	20%	10%	*Mean*	*Classification*	40%	30%	20%	10%	*Mean*	*Classification*
RF	0.916	0.903	0.885	0.870	0.8935	*Good*	0.925	0.934	0.942	0.909	0.9275	*Excellent*
GBT	0.905	0.894	0.878	0.894	0.8928	*Good*	0.901	0.911	0.941	0.899	0.9130	*Excellent*
NB	0.922	0.911	0.887	0.881	0.9003	*Excellent*	0.922	0.926	0.921	0.873	0.9105	*Excellent*
LR	0.933	0.921	0.900	0.888	0.9105	*Excellent*	0.936	0.938	0.939	0.903	0.9290	*Excellent*
NN	0.927	0.913	0.890	0.873	0.9008	*Excellent*	0.933	0.934	0.933	0.895	0.9238	*Excellent*
Mean	0.921	0.908	0.888	0.881	0.8996		0.923	0.929	0.935	0.896	0.9208	



(a) Accuracy (b) AUC

Fig. 6 Accuracy and AUC of classification models based on stratified and shuffled sampling

Furthermore, the statistical test is shown in Tables VII and VIII using a *t*-test paired two-tailed. *T*-tests were used to analyze the relationship between paired samples, where the *p*-value is a probability. If the *p*-value <0.05, then H_0 is rejected, there is a meaningful difference. On the other hand, if the *p*-value >0.05, then H_0 is accepted, and there is no meaningful difference.

Table VII shows a significant test for test data between stratified and shuffled sampling with the distribution of test data from 10 to 40 percent. Based on the accuracy results, there are three *p*-values in the test data that are less than 0.05, namely 10, 20, and 30 percent. However, based on the AUC, which has a *p*-value of less than 0.05, it is found in the test data, 20 and 30 percent, which means there is a meaningful difference in the *t*-test between the two sampling techniques.

TABLE VIII
T-TEST OF CLASSIFICATION MODELS BASED ON STRATIFIED AND SHUFFLED SAMPLING

Models	Accuracy			AUC		
	p-Value	Results	Significant	*p*-Value	Results	Significant
RF	0.047902	< 0.05	Sig.	0.041931	< 0.05	Sig.
GBT	0.235448	> 0.05	Not	0.266889	> 0.05	Not
NB	0.025302	< 0.05	Sig.	0.348270	> 0.05	Not
LR	0.087542	> 0.05	Not	0.090331	> 0.05	Not
NN	0.102391	> 0.05	Not	0.056562	> 0.05	Not

TABLE VII
T-TEST OF TEST DATA BASED ON STRATIFIED AND SHUFFLED SAMPLING

Tests (%)	Accuracy			AUC		
	p-Value	Results	Significant	*p*-Value	Results	Significant
40	0.224025	> 0.05	Not	0.284406	> 0.05	Not
30	0.014726	< 0.05	Sig.	0.002149	< 0.05	Sig.
20	0.004790	< 0.05	Sig.	0.001012	< 0.05	Sig.
10	0.008581	< 0.05	Sig.	0.138886	> 0.05	Not

TABLE IX
T-TEST OF CLASSIFICATION MODELS BASED ON AUC

	Stratified						Shuffled				
Models	RF	GBT	NB	LR	NN	Models	RF	GBT	NB	LR	NN
RF		0.933618	0.037386	0.000158	0.033018	RF		0.078384	0.104722	0.719215	0.497468
GBT	0.933618		0.367701	0.113897	0.477817	GBT	0.078384		0.847835	0.169978	0.356060
NB	0.037386	0.367701		0.003795	0.874137	NB	0.104722	0.847835		0.019434	0.022281
LR	0.000158	0.113897	0.003795		0.014987	LR	0.719215	0.169978	0.019434		0.017854
NN	0.033018	0.477817	0.874137	0.014987		NN	0.497468	0.356060	0.022281	0.017854	

Table VIII shows the *t*-test of classification models based on stratified and shuffled sampling. For accuracy with a *p*-value less than 0.05, there are two models, RF and NB. Whereas, there is only one model based on AUC, namely RF. This means that there is a meaningful difference between stratified and shuffled sampling in the RF. Whereas, for the others, there is no significant difference. Table IX shows a *t*-test between classification models based on AUC. Four models are distinguished for the stratified sampling: RF, NB, LR, and NN. Whereas, in shuffled sampling, there are three models: NB, LR, and NN.

IV. CONCLUSION

This study discusses machine learning algorithms based on sampling techniques for raisin grains classification, aiming to identify the performance of sampling techniques and classification models accurately. Three models based on stratified sampling have a performance with a mean score of AUC greater than 0.9. Whereas applying shuffled sampling has excellent performance for all models. The best outcomes of all models based on stratified sampling were founded on tested data of 40 percent with a mean accuracy of 85.50% and an AUC of 0.921. In comparison, models based on shuffled sampling were founded on test data of 20 percent with a mean accuracy of 88.11% and an AUC of 0.935.

Moreover, for the *t*-test, RF significantly differs based on sampling techniques. It can be concluded that models based on a stratified sampling of all data split, not all models perform excellently. Moreover, in models based on shuffled sampling, all performances are excellent. In this case, the classification of raisin grains by applying the shuffled sampling technique with 20% testing is reliable. In future work, enhancing the performance of a more accurate model needs parameter optimization.

REFERENCES

[1] M. J. Schuster, X. Wang, T. Hawkins, and J. E. Painter, "A Comprehensive review of raisins and raisin components and their relationship to human health," *J. Nutr. Heal.*, vol. 50, no. 3, p. 203, 2017, doi: 10.4163/jnh.2017.50.3.203.

[2] R. Khiari, H. Zemni, and D. Mihoubi, "Raisin processing: physicochemical, nutritional and microbiological quality characteristics as affected by drying process," *Food Rev. Int.*, vol. 35, no. 3, pp. 246–298, Apr. 2019, doi: 10.1080/87559129.2018.1517264.

[3] T. Uzun and B. Hallaç, "Physicochemical Characteristic of The Organic Raisins Served to Markets in Midyat Town of Mardin Province and Beşiri Town of Batman Province in Turkey," *Erwerbs-Obstbau*, vol. 63, no. S1, pp. 61–69, Oct. 2021, doi: 10.1007/s10341-021-00582-6.

[4] A. Rahimi, A. Heshmati, and A. Nili-Ahmadabadi, "Changes in pesticide residues in field-treated fresh grapes during raisin production by different methods of drying," *Dry. Technol.*, vol. 0, no. 0, pp. 1–14, May 2021, doi: 10.1080/07373937.2021.1919140.

[5] G. Singh, N. Kaushal, O. Tokusoglu, and A. Singh, "Optimization of process parameters for drying of red Grapes (Vitis vinifera) to raisin:

A design expert laden approach," *J. Food Process. Preserv.*, no. September 2020, pp. 1–8, Jan. 2021, doi: 10.1111/jfpp.15248.

[6] J. Wang, A. S. Mujumdar, H. Wang, X.-M. Fang, H.-W. Xiao, and V. Raghavan, "Effect of drying method and cultivar on sensory attributes, textural profiles, and volatile characteristics of grape raisins," *Dry. Technol.*, vol. 39, no. 4, pp. 495–506, Feb. 2021, doi: 10.1080/07373937.2019.1709199.

[7] R. Sadeghi, E. Seyedabadi, and R. M. Moghaddam, "Evaluation of Microwave and Ozone Disinfections on the Color Characteristics of Iranian Export Raisins Through an Image Processing Technique," *J. Food Prot.*, vol. 82, no. 12, pp. 2080–2087, Dec. 2019, doi: 10.4315/0362-028X.JFP-19-296.

[8] M. Khojastehnazhand and H. Ramezani, "Machine vision system for classification of bulk raisins using texture features," *J. Food Eng.*, vol. 271, no. September 2019, p. 109864, Apr. 2020, doi: 10.1016/j.jfoodeng.2019.109864.

[9] Y. Zhao, X. Xu, and Y. He, "A Novel Hyperspectral Feature-Extraction Algorithm Based on Waveform Resolution for Raisin Classification," *Appl. Spectrosc.*, vol. 69, no. 12, pp. 1442–1456, Dec. 2015, doi: 10.1366/14-07617.

[10] J. Guo *et al.*, "Near-infrared spectroscopy combined with pattern recognition algorithms to quickly classify raisins," *Sci. Rep.*, vol. 12, no. 1, p. 7928, May 2022, doi: 10.1038/s41598-022-12001-1.

[11] L. Feng, S. Zhu, C. Zhang, Y. Bao, P. Gao, and Y. He, "Variety Identification of Raisins Using Near-Infrared Hyperspectral Imaging," *Molecules*, vol. 23, no. 11, p. 2907, Nov. 2018, doi: 10.3390/molecules23112907.

[12] Y. Zhang, Y. Yang, C. Ma, and L. Jiang, "Identification of multiple raisins by feature fusion combined with NIR spectroscopy," *PLoS One*, vol. 17, no. 7, p. e0268979, Jul. 2022, doi: 10.1371/journal.pone.0268979.

[13] K. Mollazade, M. Omid, and A. Arefi, "Comparing data mining classifiers for grading raisins based on visual features," *Comput. Electron. Agric.*, vol. 84, pp. 124–131, Jun. 2012, doi: 10.1016/j.compag.2012.03.004.

[14] A. Bakhshipour, A. Jafari, and A. Zomorodian, "Vision based features in moisture content measurement during raisin production," *World Appl. Sci. J.*, vol. 17, no. 7, pp. 860–869, 2012.

[15] Y. Zhao, M. L. Guindo, X. Xu, X. Shi, M. Sun, and Y. He, "A Novel Raisin Segmentation Algorithm Based on Deep Learning and Morphological Analysis," *Eng. Agrícola*, vol. 39, no. 5, pp. 639–648, Oct. 2019, doi: 10.1590/1809-4430-eng.agric.v39n5p639-648/2019.

[16] N. Karimi, R. Ranjbarzadeh Kondrood, and T. Alizadeh, "An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms," *Measurement*, vol. 107, pp. 68–76, Sep. 2017, doi: 10.1016/j.measurement.2017.05.009.

[17] İ. Çinar, M. Koklu, and Ş. Taşdemir, "Classification of Raisin Grains Using Machine Vision and Artificial Intelligence Methods," *Gazi J. Eng. Sci.*, vol. 6, no. 3, pp. 200–209, Dec. 2020, doi: 10.30855/gmbd.2020.03.03.

[18] I. Cinar and M. KOKLU, "Classification of Rice Varieties Using Artificial Intelligence Methods," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 3, pp. 188–194, Sep. 2019, doi: 10.18201/ijisae.2019355381.

[19] F. Tarakci and I. A. Ozkan, "Comparison of classification performance of kNN and WKNN algorithms," vol. 20, no. 02, pp. 32–37, 2021, [Online]. Available: https://sujes.selcuk.edu.tr/sujes/article/view/536.

[20] T. Zaman, "An efficient exponential estimator of the mean under stratified random sampling," *Math. Popul. Stud.*, vol. 28, no. 2, pp. 104–121, Apr. 2021, doi: 10.1080/08898480.2020.1767420.

[21] R. Rachmatika and A. Bisri, "Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 417–422, 2020, doi: http://dx.doi.org/10.26418/jp.v6i3.43097.

[22] A. Bisri and R. Rachmatika, "Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 309, Nov. 2019, doi: 10.22146/jnteti.v8i4.529.

[23] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM; 2016: 785–94." 2016.

[24] M. J. Zaki and W. J. Meira, *Data Mining and Machine Learning Fundamental Concepts and Algorithms*, Second. 2020.

[25] L. Niu, "A review of the application of logistic regression in educational research: common issues, implications, and suggestions," *Educ. Rev.*, vol. 72, no. 1, pp. 41–67, 2020, doi: 10.1080/00131911.2018.1483892.

[26] C. C. Aggarwal, "Machine Learning for Text: An Introduction," in *Machine Learning for Text*, Cham: Springer International Publishing, 2018, pp. 1–16.

[27] V. Kotu and B. Deshpande, *Data Science: Concepts and Practice*, Second Edi. Morgan Kaufmann, 2019.

[28] J. P. Quintas, F. Machado e Costa, and A. C. Braga, "ROSY Application for Selecting R Packages that Perform ROC Analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12251 LNCS, 2020, pp. 199–213.

[29] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[30] G. James, DanielaWitten, T. Hastie, and R. Tibshirani, *Springer Texts in Statistics An Introduction to Statistical Learning wth application in R*. 2013.