Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 1, April 2023, pp. 192~199
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i1.pp192-199

Feature selection optimization based on genetic algorithm for support vector classification varieties of raisin

Yudi Ramdhani[1], Dhia Fauziah Apra[2], Doni Purnama Alamsyah[3]
[1,2]Department of Informatic Engineering, Adhirajasa Reswara Sanjaya University, Bandung, Indonesia
[3]Entrepreneurship Department, BINUS Business School Undergraduate Program, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Oct 4, 2022
Revised Nov 21, 2022
Accepted Nov 30, 2022

Keywords:

Cross validation
Genetic algorithm
Grape
Optimize selection
Raisins
Split validation
Support vector machine

ABSTRACT

Grapes are one of the fruit plants that grow that propagate in certain fields. Grapes can be processed into juice, wine, raisins, and so on. Raisins are dried grapes. Raisins have a distinctive taste and aroma. Raisins are a concentrated and nutritious source of carbohydrates, containing antioxidants, potassium, fiber and iron. To increase the accuracy value, the optimize selection genetic algorithm (GA) is used. This research was conducted modeling using the support vector machine (SVM) and SVM algorithms based on optimize selection GA by using the raisin (raisin varieties) dataset obtained from the UCI machine learning repository. The research dataset is divided into training data and testing data. The data sharing will be carried out using the cross validation and split validation operators. Data validation with 10-Fold-validation on the SVM algorithm has the best level of performance among 5 other algorithms such as; Naïve Bayes, K-nearest neighbor (K-NN), decision tree (DT), neural network, and random forest (RF). The SVM algorithm produces accuracy and area under the curve (AUC) values of 87.11% for accuracy and 0.928 for AUC. Optimization in this study using optimize selection GA. SVM based on optimize selection GA produces accuracy and AUC values of 87.67% for accuracy and 0.930 for AUC.

Corresponding Author:

Yudi Ramdhani
Department of Informatic Engineering, Adhirajasa Reswara Sanjaya University
Bandung, Indonesia
Email: yudi@ars.ac.id

1. INTRODUCTION

Indonesia is well-known for its abundant natural resources. One of them is the end product of its plantations, as evidenced by the abundance of plantation products in Indonesia. Plantation products are one of the state assets that play an important role in regional and national economic development, particularly in efforts to increase employment opportunities, equalize income, and improve people's living standards [1]. Wine is a commodity with added value. That is, it can be consumed in the form of fresh fruit, grape juice, beverages (wine), and raisins. Grapes are climbing plants with a unique feature in that their branches can produce dense fruit. Grapes can be grown in cold, subtropical, or tropical climates. The vines originated in Europe's plains, North America, Iceland, cold areas near the North Pole, and Greenland, and then spread to Asia, including Indonesia. Local grapes are regarded as a commercially valuable crop in Indonesia [2]. Working at home or work from home (WFH) is one of the most effective things during the COVID-19 pandemic. One of the businesses that is currently being looked at by the public is the cultivation of imported grape seeds. Imported vines have good prospects in the future. Based on this, the grape community is more numerous than other fruit plant communities [3].

There are dozens of grape varieties found throughout Indonesia. Those still in the form of fresh or processed fruit, such as wine and raisins, can be found at the Banjarsari Experimental Garden in Pasuruan. Isabella grapes have also been developed in Palu, Central Sulawesi, with similar good results to imported wines, though wine development in Palu was eventually halted due to marketing constraints. Despite its shortcomings in comparison to subtropical regions, Indonesia as a tropical country has several advantages. Grape productivity is lower in the tropics than in the subtropics. Wine production in subtropical regions can reach 20 tons per hectare per year, whereas in tropical countries like Indonesia, it is only half that. However, the grape harvest in Indonesia can reach three harvests per year, whereas it is only once in subtropical countries [4], [5]. Grape (Vitis vinifera L.) is a fruit plant that grows by vines in certain fields. Grapes are certainly rich in benefits and are included in non-climacteric fruits [6]. Grapes can be processed into juice, wine, raisins, and so on. Raisins are dried grapes. Raisins have a distinctive taste and aroma. Raisins contain a fairly high concentration of sugar. During the decrystallization process, the fruit will be soaked in juice or boiling water to dissolve the sugar. This process also makes the raisin skin rough. Raisins are used as cake decorations, chocolate mixes, candy or bread [7]. Iron, potassium, vitamin B6, manganese, boron, selenium, vitamin C, calcium, magnesium, phosphorus, and sodium are all found in raisins [8]. Turkey ranks among the top wine-producing countries in the world. Turkey is a country with a long history of wine production and a large tourism industry, but it has yet to capitalize on the importance of wine tourism. Turkey has favorable conditions for grape cultivation and wine production due to its geographical location [9]. Turkey is currently the sixth largest wine producer in the world, with an average production of 4,080,932 tonnes and an average surface area of 440,829 hectares (ha) [10], [11]. As the second largest producer of raisins in the world, holding 25% of the total raisin production, and accounting for almost 40-45% of the volume traded, being a world leader in its exports [12].

Data mining is the process of finding patterns and correlations in large data sets to predict outcomes [13]-[15]. Data mining has its roots in artificial intelligence, particularly in machine learning (ML) as well as in statistical analysis to solve a problem that involves prediction, classification and segmentation, meaning that large amounts of data can be processed and used more efficiently [16]-[18]. Data mining classification techniques are used to measure the level of accuracy in a dataset. Classification is the job of evaluating data objects to put them into certain categories based on the number of categories available. Classifier builds a model based on existing training data, and then uses that model to classify the new data. Classification can be defined as the job of doing training or learning on an objective function that maps each set of attributes (features) to a number of available class labels [19]. There are many good classification techniques in the literature including artificial neural networks, k-nearest-neighbors classifier, decision trees, Bayesian classifier and support vector machine (SVM) algorithms. Of these techniques, SVM is one of the best known techniques for optimizing the expected solution [20]. SVM algorithm is one of the supervised machine learning algorithms based on statistical learning theory [21]. This algorithm selects from the training sample a subset of characteristics so that the classification of the character subset is equivalent to dividing the entire dataset. SVM has been used to solve different classification problems successfully in many applications [22], [23]. The accuracy of the target detection classifier can be guaranteed by the global optimal solution. However, it has some drawbacks, such as the long-established detection model. When processing large-scale data, time complexity and space complexity increase linearly with increasing data [24], [25]. In comparison, SVM is better able to solve smaller sample, nonlinear and high dimensional problems compared to other classification algorithms [26], [27].

Previous research conducted a classification model of hand movements based on electromyogram signals has been successfully developed using a machine support vector algorithm resulting in an overall accuracy value of 97.4% for training, and 88.0% for testing [28]. The findings of this study validate the performance of the machine algorithm's quadratic support vector metric (SVM squared) when applied to student satisfaction predictions, correct within 97.8% (Accuracy) in predictions, with recall (sensitivity) 96.5% and F1 score 0.968 [29]. The aim of this study was to build a classification model that might predict the early stage of Alzheimer's disease. There are 3 algorithms used, namely SVM, Naïve Bayes (NB), and K-nearest neighbors (K-NN). The current findings reveal that the SVM-based classification model can accurately distinguish cognitively impaired Alzheimer's patients from normal healthy individuals with 96.6% accuracy [30]. In this study, the classification of the Besni and Kecimen raisin varieties produced in Turkey was carried out using the SVM algorithm with a dataset of 900 data.

2. METHOD

2.1. Research method

To solve the problem of classifying raisin varieties in this study, several methods were used, including training on data separation and data testing using two methods (cross validation and split validation). A comparison of classification algorithms is also performed in order to determine the best classification

algorithm. The next step is to improve the classification by optimizing the dataset's features and weights. This research stage concludes with an evaluation to determine which algorithm will be used for classification and which optimization algorithm can improve classification value.

a) Problem identification

Raisins certainly have many varieties. To classify these varieties of raisins, an appropriate algorithm model is needed, so that it can help experts in classifying raisin varieties. In this study, researchers classified raisin varieties that grow in Turkey in the form of Kecimen and Besni raisin varieties.

b) Data collection

The data used in this study is public data, namely raisin dataset. The dataset is achieved through the UCI machine learning website in 2021. The raisin dataset consists of 900 data records and 8 attributes. This dataset is divided into two classes, namely the Kecimen class and the Besni class.

c) Data preprocessing

At the data pre-processing stage, the dataset is checked in the form of missing values, remove duplicate data, and normalize. Remove duplicate data is done to delete the same data. Normalization is done with the Z-Transformation method so that the attribute variables have the same value range, which is between 0 to 1.

d) Data validation

At the data validation stage, research data will be divided into training data and testing data. The data sharing will be done using cross validation and split validation. Data sharing using cross validation is carried out to determine the best performance of the model to be tested, while split validation is carried out to test a particular model.

e) Comparation of algorithm

Comparison of algorithms is used to obtain the algorithm that is considered the best in the process of classifying raisin varieties. In the algorithm comparison stage, several algorithms are tested. This study uses 6 algorithms, namely; Naïve Bayes, K-NN, decision tree (DT), neural network, SVM, and random forest (RF). Based on these 6 algorithms, it will be known which algorithm is the best in classifying raisin varieties through the accuracy value it produces.

f) Support vector machine

At this stage the SVM algorithm became the best model in the classification of raisin varieties. The SVM model was determined based on the highest level of accuracy and area under the curve (AUC) among the 6 algorithms used in the classification of raisin seeds. The algorithm will be tested for the model using split validation. The model will be tested using the split ratio parameter 0.5-0.9 so that the average value will be obtained.

g) Comparation of optimal algorithm

At the comparison stage of the optimization algorithm, testing is carried out with several optimization features. This study uses 2 optimization features, namely optimize selection and optimize weight. Each of these optimization features uses 3 algorithms, namely GA, backward, and forward.

h) Genetic algorithm (GA)

GAs is inspired by biological evolution. Mutation and crossover are two of the most commonly used GA operators. Mutation and crossover are two of the most commonly used GA operators. Mutation works on a single solution and generally alters a feature at random or according to some pre-defined criterion. Crossover, on the other hand, uses two parent solutions to create two offspring, resulting in new and improved solutions [31]. In general, the mathematical model is based on an initial chromosome population of n individuals. There are three operations in each iteration from a maximum number of t epochs: reproduction, mutation, and selection. The best individuals evaluated by the fitness function are assumed as a solution for a given problem at the end of the algorithm [32].

i) Evaluation

At this evaluation stage, the best accuracy and AUC values will be known in the classification of raisin varieties. The researcher saw a comparison of the results of accuracy and AUC with a split ratio of 0.5 to 0.9 from the SVM algorithm and the SVM algorithm based on optimize selection GA and conducted a paired two sample for means T-test using Microsoft Excel to find out whether there was a difference between before optimization and after optimization raisin variety classification optimization.

2.2. Proposed method

In this study, a method is proposed for the classification of raisin GA as a feature selection and SVM algorithm as a classification of raisin varieties. The proposed method can be seen in Figure 1. The initial stage in this research is the collection of the raisin dataset. After that, checking the dataset, and normalizing the data. Normalization in the dataset with the aim of blocking data in a simple range using the z-transformation method. The next step is to separate the data into training data and testing data. The training data is applied to generate a model from the SVM algorithm, while for testing the dataset it is applied to generate accuracy values. The

next step is to compare the algorithms. Algorithm comparison is done to compare several algorithms in classifying so that the best algorithm model is obtained.

Feature selection used in the study using GA. The genetic algorithm makes a population consisting of many selected individuals with the most values relevant to the classification so as to improve the performance of the classification accuracy value of raisin arieties [33]. Furthermore, the features that have been selected by the genetic algorithm are classified using the SVM algorithm. In Figure 1, the researcher describes the proposed method scheme for the classification of raisin varieties. The results of the evaluation of the classification of the raisin variety with the proposed model have the maximum value with feature optimization using GA so that it can affect the maximum classification results carried out by the SVM algorithm in classifying raisin varieties into the Kecimen class and Besni class.

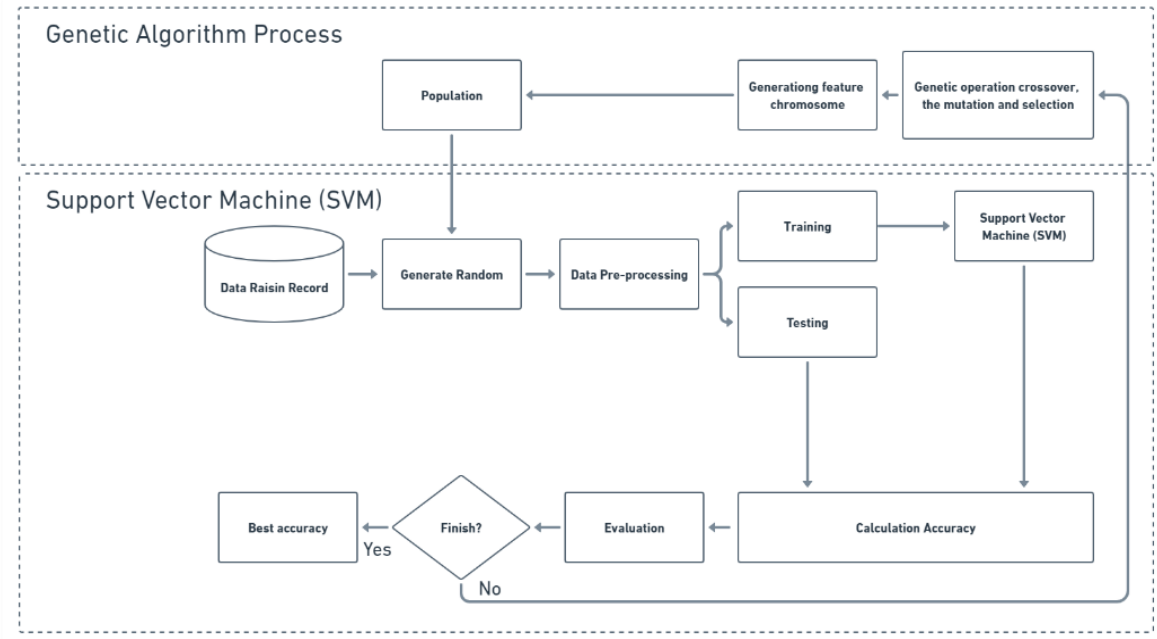

Figure 1. Proposed method

3. RESULTS AND DISCUSSION

At this stage, the experimental results of testing the classification of the raisin dataset are shown. The first step is to identify the problem. It is known that in classifying raisins, a method or algorithm with the best model is needed. Based on this, a research was conducted on the classification of raisin varieties. This research uses raisin's research dataset obtained from the UCI machine learning repository website. The dataset in this study has 900 data records of raisin varieties consisting of 8 attributes and 1 label consisting of 2 classes, namely Kecimen class and Besni class. This is shown in Table 1.

After data collection, the researcher preprocessed the data. At this stage, checking for missing values on the data is carried out to see if there are data that are not appropriate. After that, remove duplicates so that no data is the same, and normalize with the Z-Transformation method. The normalized data will have the same value range, which is between 0 to 1. The following is the result of the normalization that has been carried out which is shown in Table 2.

Table 1. Attribute

No	Attribute	Detail
1	Area	Gives the number of pixels in raisins
2	Perimeter	Measures the environment by calculating the distance between the currant border and the surrounding pixels
3	Major Axis Length	Gives the length of the main axis
4	Minor Axis Length	Gives small axis length
5	Eccentricity	Gives a measure of the eccentricity of the ellipse, which has the same moment as the raisin
6	ConvexArea	Gives the smallest number of convex skin pixels of the region formed by raisins
7	Extent	Gives the ratio of the area formed by the raisins to the total pixels in the bounding box
8	Class	Kecimen and Besni raisins

Table 2. Normalization of dataset

No	Class	Area	MajorAxis	MinorAxis	Eccentricity	ConvexAres	Extent	Parimeter
1	Kecimen	-0,007	0,098	-0,024	0,423	-0,016	1,106	0,066
2	Kecimen	-0,324	-0,209	-0,229	0,,224	-0,304	-0,288	-0,161
3	kecimen	0,078	0,098	0,237	0,186	0,062	-1,158	0,156
..	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…
899	Besni	0,147	0,391	-0,006	0,711	0,159	-0,761	0,338
900	Besni	-0,056	0,699	-0,784	1,393	-0,049	-1,262	0,391

After the data preprocessing process is complete, the next step is to compare the algorithms. Algorithm comparison was conducted to compare the 6 algorithms tested in this study. The algorithm used is; Naïve Bayes, K-NN, decision tree, neural network, SVM, and random forest. To determine the performance of the 6 algorithms, the data validation process is carried out using the 10-Fold validation method which produces accuracy, precision, recall, and AUC values. The following are the accuracy and AUC values shown in Table 3 generated by each algorithm. To make it easier to understand the difference in accuracy of the AUC Performance in the comparison algorithm, it is necessary to make a graph. The following is a graph of 6 algorithms in the algorithm comparison process shown in Table 3.

Based on the comparison of these algorithms, it is known that the SVM algorithm has the highest algorithm performance value compared to other algorithms, which is 87.11% for accuracy and 0.928 for AUC. The following table confusion matrix generated SVM algorithm classification can be seen in Table 4. From the results of testing the AUC value of the SVM algorithm model is 0.928. Based on the test value, it shows that the SVM algorithm model achieves excellent classification. Receiver operating characteristic (ROC) curves are also generated by Rapidminer. After knowing the best performance of the SVM algorithm in classifying raisin varieties, data validation was carried out using split validation to test the algorithm. The following are the results of data validation using split validation with a split ratio of 0.5 to 0.9 contained in Table 5.

Table 3. Result of algorithm comparison

Algorithm	Validation	Accuracy	AUC
Naïve Bayes	Cross	83.67%	0.92
K-NN	Cross	85.11%	0.91
Decision Tree	Cross	85.11%	0.866
Neural Network	Cross	86.67%	0.927
SVM	Cross	87.11%	0.928
Random Forest	Cross	85.56%	0.926

Table 4. Confusion matrix SVM

	True Kecimen	True Besni	Class Precision
Pred. Kecimen	405	71	85.08%
Pred. Besni	45	379	89.39%
Class Recall	90.00%	84.22%	

Table 5. Split ratio 0,5-0,9 SVM

Algorithm	Validation	Ratio	Accuracy	AUC
SVM	Split	0.5	88.44%	0.944
SVM	Split	0.6	86.11%	0.927
SVM	Split	0.7	84.81%	0.914
SVM	Split	0.8	82.22%	0.89
SVM	Split	0.9	82.22%	0.871
	Average		84.76%	0.9218

Based on Table 4, it can be seen that the SVM algorithm with a split ratio of 0.5 to 0.9 has an average value of 84.76% for accuracy and 0.9218 for AUC. Validation with a split ratio of 0.5 has the highest accuracy and AUC values of 88.44% and 0.944, respectively. The following table of the resulting confusion matrix can be seen in Table 6. From the results of testing the AUC value of the SVM algorithm model with a split ratio of 0.5 is 0.944. Based on the test value, it shows that the SVM algorithm model with a split ratio of 0.5 achieves a very good classification. ROC curves are also generated by Rapidminer. To increase the accuracy value of the SVM algorithm, the optimization feature is used. In this study, we compare the optimization features,

namely optimize selection and optimize weight. Validation is carried out using the 10-Fold validation method. The following are the accuracy and AUC values of each optimization feature contained in Table 7 and Table 8.

Table 6. Confusion matrix split ratio 0,5 SVM

	True Kecimen	True Besni	Class precision
Pred. Kecimen	206	33	86.19%
Pred. Besni	19	192	91.00%
Class Recall	91.56%	85.33%	

Based on Table 6 and Table 7, it can be seen that optimize selection and optimize weight have succeeded in increasing the accuracy and AUC values of the SVM algorithm in the classification of raisin varieties. The optimize selection and optimize weight features with GA have the highest accuracy and AUC values compared to other optimization method features, besides that the accuracy and AUC values produced have the same value, namely 87.67% for accuracy and 0.930 for AUC. Based on this, the researcher chose to test the SVM algorithm based on optimize selection GA in classifying raisin varieties. The following is a table of the overall values of accuracy, precision, recall, and AUC of each parameter split ratio 0.5 to 0.9 SVM algorithm based on optimize selection GA can be seen in Table 9.

The following is a test of the SVM algorithm based on optimize selection GA with a split ratio of 0.5 to 0.9 Table 9. It is known that the SVM algorithm based on optimize selection GA with a split ratio of 0.5 to 0.9 has an average value of 91.56% for accuracy, 94.79% for precision, 87.98% for recall, and 0.953 for AUC. Validation with a split ratio of 0.9 has a high final result compared to other split ratios. Table 10 comparison of the accuracy of SVM and SVM algorithms based on optimize selection GA with a split ratio of 0.5 to 0.9.

Table 11 AUC comparison of SVM and SVM algorithms based on optimize selection GA with a split ratio of 0.5 to 0.9. To make it easier to understand the difference in accuracy of the results of the confusion matrix calculation and the AUC performance of the SVM and SVM methods based on the optimize selection GA split ratio of 0.9, it is necessary to make a graph. The following is a graph of the comparison of accuracy and AUC values between SVM and SVM based on optimize selection GA. After testing the SVM algorithm and SVM based on optimize selection GA, the last step to be taken is to do a T-Test paired two samples. The t-test was carried out by researchers to find out whether there was a difference in the average value of the raisin variety before and after optimization. The following are the results of the T-Test paired two samples using Microsoft Excel shown in Table 12. Based on the T-Test, the significance value can be compared with 0.05. The significance value generated in the T-Test is 0.026571244 which is smaller than 0.05, meaning that there is a difference between before optimize and after optimize.

Table 7. Feature optimize selection

Algorithm	Method	Validation	Accuracy	AUC
SVM	GA	Cross	87.67%	0.93
SVM	Forward	Cross	87.22%	0.93
SVM	Backward	Cross	87.11%	0.93

Table 8. Feature optimize weight

Algorithm	Method	Validation	Accuracy	AUC
SVM	GA	Cross	87.67%	0.93
SVM	Forward	Cross	87.22%	0.93
SVM	Backward	Cross	87.33%	0.93

Table 9. Result of accuracy, precision, recall, AUC SVM + optimize selection GA

Algorithm	Method	Ratio	Accuracy	Precision	Recall	AUC
SVM	GA	0.5	90.22%	92.89%	87.11%	0.939
SVM	GA	0.6	90.00%	92.35%	87.22%	0.95
SVM	GA	0.7	90.37%	95.80%	84.44%	0.954
SVM	GA	0.8	92.78%	95.29%	90.00%	0.953
SVM	GA	0.9	94.44%	97.62%	91.11%	0.969
		average	91.56%	94.79%	87.98%	0.953

Table 10. Accuracy SVM and SVM + GA

Validation	Accuracy	
	Algoritm SVM	Algoritma SVM + GA
0.5	88.44%	90.22%
0.6	86.11%	90.00%
0.7	84.81%	90.37%
0.8	82.22%	92.78%
0.9	82.22%	94.44%

Table 11. AUC SVM and SVM + optimize selection GA

Validation	AUC	
	Algoritm SVM	Algoritma SVM + GA
0.5	0.944	0.939
0.6	0.927	0.95
0.7	0.914	0.954
0.8	0.953	0.953
0.9	0.871	0.969

Table 12. T-test paired two samples

	Variable 1	Variable 2
Mean	0.8476	0.91562
Variance	0.000706765	0.000385702
Observation	5	5
Pearson Correlation	-0.838455569	
Hypothesized Mead	0	
Df	4	
T stat	-3.428537169	
P(T<=t) one-tail	0.013285622	
T Critical one-tail	2.131846786	
P(T<=t)two-tail	0.026571244	
T Critical two-tail	2.776445105	

4. CONCLUSION

This research was conducted modeling the SVM and SVM algorithms based on optimize selection GA by using the raisin (raisin varieties) dataset obtained from the UCI machine learning repository. The SVM algorithm produces accuracy and AUC values of 87.11% for accuracy and 0.928 for AUC.To improve the performance of the accuracy value of the SVM algorithm, optimization is carried out with the selection feature with the GA method resulting in accuracy and AUC values of 87.67% for accuracy and 0.930 for AUC. Based on the tests that have been obtained on the raisin dataset, it can be seen that the support vector machine algorithm based on optimize selection GA has a good accuracy of 87.67%, so it can be used as a reference for methods by a programmer which will be implemented when making a program regarding the classification of raisin varieties.

REFERENCES

[1] M. Romero, M. Wollni, K. Rudolf, R. Asnawi, and B. Irawan, "Promoting biodiversity enrichment in smallholder oil palm monocultures–Experimental evidence from Indonesia," *World Dev*, vol. 124, p. 104638, 2019, doi: 10.1016/j.worlddev.2019.104638.

[2] A. B. Suwardi and Z. I. Navia, "Sustainable use and management of wild edible fruit plants: A case study in the ulu masen protected forest, West Aceh, Indonesia," *Journal of Sustainable Forestry*, pp. 1–20, 2022, doi: 10.1080/10549811.2022.2123355.

[3] S. Lopez-Leon, D. A. Forero, and P. Ruiz-Díaz, "Recommendations for working from home during the COVID-19 pandemic (and beyond)," *Work*, vol. 66, no. 2, pp. 371–375, 2020, doi: 10.3233/WOR-203187.

[4] R. Rajan, M. Feza, K. Pandey, A. Aman, and V. Kumar, "Climate change and resilience in fruit crops," *Climate change and its effects on Agriculture*, pp. 337–354, 2020.

[5] C. Frankel, *Volcanoes and Wine: From Pompeii to Napa*. University of Chicago Press, 2019, doi: 10.7208/chicago/9780226603582.001.0001.

[6] A. Heshmati, S. Ghadimi, A. Ranjbar, and A. M. Khaneghah, "Assessment of processing impacts and type of clarifier on the concentration of ochratoxin A in pekmez as a conventional grape-based product," *LWT*, vol. 119, p. 108882, 2020, doi: 10.1016/j.lwt.2019.108882.

[7] D. Wang, C.-Q. Duan, Y. Shi, B.-Q. Zhu, H. U. Javed, and J. Wang, "Free and glycosidically bound volatile compounds in sun-dried raisins made from different fragrance intensities grape varieties using a validated HS-SPME with GC–MS method," *Food Chem*, vol. 228, pp. 125–135, 2017, doi: 10.1016/j.foodchem.2017.01.153.

[8] A. Utebaeva, R. Alibekov, V. Evlash, and V. Lonkin, "Functional jam based on the grapes-seedless raisins," *Industrial Technology and Engineering*, vol. 2, no. 35, pp. 25–32, 2020.

[9] E. K Dodd, "Roman and late antique wine production in the eastern Mediterranean: a comparative archaeological study at Antiochia ad Cragum (Turkey) and Delos (Greece)," *Roman and Late Antique Wine Production in the Eastern Mediterranean*, pp. 1–222, 2020, doi: 10.2307/j.ctvwh8c1m.

[10] N. Keskin, O. Kaya, F. Ates, and M. Turan, "Solutions: Effects on Hormones, Minerals, and Vitamins" pp. 1–12, 2022.

[11] Y. Ramdhani, R. T. Prasetio, and D. P. Alamsyah, "Decision support system application with survey 360 degree feedback," in *3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2021, pp. 1–5, doi: 10.1109/ICORIS52787.2021.9649489.

[12] H. Uysal and S. Karabat, "Forecasting and evaluation for raisin production in Turkey," in *XXX International Horticultural Congress IHC2018: International Symposium on Viticulture: Primary Production and Processing 1276*, 2018, pp. 217–222, doi: 10.17660/ActaHortic.2020.1276.31.

[13] J. F. P. da Costa and M. Cabral, "Statistical methods with applications in data mining: A review of the most recent works," *Mathematics*, vol. 10, no. 6, p. 993, 2022, doi: 10.3390/math10060993.

[14] R. T. Prasetio, Y. Ramdhani, and D. P. Alamsyah, "Scrum method in help-desk ticketing and project management system," in *3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2021, pp. 1–6, doi: 10.1109/ICORIS52787.2021.9649626.

[15] D. P. Alamsyah, Y. Ramdhani, T. Arifin, F. Febrilla, and S. Setiawan, "Prediction of immunotherapy success rate: particle swarm optimization approach," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Jun. 2022, pp. 1–5. doi: 10.1109/CONIT55038.2022.9847850.

[16] M. J. Zaki and W. Meira Jr, *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press, 2020, doi: 10.1017/9781108564175.

[17] D. P. Alamsyah, T. Arifin, Y. Ramdhani, F. A. Hidayat, and L. Susanti, "Classification of customer complaints: TF-IDF approaches," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Jun. 2022, pp. 1–5, doi: 10.1109/CONIT55038.2022.9848056.

[18] D. P. Alamsyah, Y. Ramdhani, S. R. Rhamadhan, and L. Susanti, "Application of IoT and cloud storage in android-based smart home technology," in *2021 7th International HCI and UX Conference in Indonesia (CHIuXiD)*, 2021, pp. 33–38, doi: 10.1109/CHIuXiD54398.2021.9650605.

[19] D. Verma and N. Mishra, "Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques," in *2017 International Conference on Intelligent Sustainable Systems*, 2017, pp. 533–538, doi: 10.1109/ISS1.2017.8389229.
[20] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[21] Y. Wendong, L. Zhengzheng, and J. Bo, "A multi-factor analysis model of quantitative investment based on GA and SVM," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017, pp. 1152–1155, doi: 10.1109/ICIVC.2017.7984734.
[22] F. Q. Kareem and A. M. Abdulazeez, "Ultrasound medical images classification based on deep learning algorithms: a review," *Fusion: Practice and Applications, vol*, vol. 3, pp. 29–42, 2021, doi: 10.54216/FPA.030102.
[23] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 145–150, doi: 10.1109/ICOASE.2018.8548836.
[24] H. Dai, "Research on SVM improved algorithm for large data classification," in *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018, pp. 181–185, doi: 10.1109/ICBDA.2018.8367673.
[25] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 4, pp. 2715–2743, 2021.
[26] P. Tao, Z. Sun, and Z. Sun, "An improved intrusion detection algorithm based on GA and SVM," *Ieee Access*, vol. 6, pp. 13624–13631, 2018, doi: 10.1109/ACCESS.2018.2810198.
[27] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artif Intell Rev*, vol. 52, no. 2, pp. 803–855, 2019, doi: 10.1007/s10462-018-9614-6.
[28] D. Albitar, R. Jailani, M. S. A. M. Ali, and A. P. P. A. Majeed, "Classification of hand gestures from forearm electromyogram signatures from support vector machine," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, p. 260, 2021, doi: 10.11591/ijeecs.v24.i1.pp260-268, doi: 10.11591/ijeecs.v24.i1.pp260-268.
[29] O. Chamorro-Atalaya *et al.*, "Quadratic vector support machine algorithm, applied to prediction of university student satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 139–148, 2022, doi: 10.11591/ijeecs.v27.i1.pp139-148.
[30] S. T. Ahmed and S. M. Kadhem, "Alzheimer's disease prediction using three machine learning methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 3, pp. 1689–1697, 2022, doi: 10.11591/ijeecs.v27.i3.pp1689-1697.
[31] S. Ahmed, K. K. Ghosh, P. K. Singh, Z. W. Geem, and R. Sarkar, "Hybrid of harmony search algorithm and ring theory-based evolutionary algorithm for feature selection," *IEEE Access*, vol. 8, pp. 102629–102645, 2020, doi: 10.1109/ACCESS.2020.2999093.
[32] D. Połap, "An adaptive genetic algorithm as a supporting mechanism for microscopy image analysis in a cascade of convolution neural networks," *Appl Soft Comput*, vol. 97, p. 106824, 2020, doi: 10.1016/j.asoc.2020.106824.
[33] Y. Ramadhani, A. Mubarok, S. Hidayatullah, and W. Wiguna, "Attribute optimization: Genetic algorithms and neural network for voice analysis classification of parkinson's disease," no. Icri 2018, pp. 3074–3079, 2020, doi: 10.5220/0009947030743079.

BIOGRAPHIES OF AUTHORS

Yudi Ramdhani is Faculty Member and researcher from Adhirajasa Reswara Sanjaya University in Informatics Technic Program. Have focused research on Data Mining, Data Science and Decision Support System. He can be contacted at email: yudi@ars.ac.id.

Dhia Fauziah Apra she is a student of Informatics Engineering at Adhirajasa University Reswara Sanjaya. She has a high interest in data mining processing, has HTML, PHP, and graphic design skills. She can be contacted via email: dhiaapra04@gmail.com.

Doni Purnama Alamsyah he is faculty member from Bina Nusantara University, teaching at Entrepreneurship Department in Customer Behavior study. He received a Doctorate degree from Padjadjaran University in the field of Management Science. Currently has an interest and research focus on Consumer Behavior, and is very open to conducting research collaboration. He also can be contacted at email: doni.syah@binus.ac.id.