

AVOCADO DAMAGE DETECTION SYSTEM (Persea Americana Mill) USING GAS SENSOR WITH THE STRATIFIED K-FOLD CROSS VALIDATION METHOD

Rahmat Sabani *¹

Program Studi Teknik Pertanian, Universitas Mataram
Correspondensi author email: rahmat.sabani@gmail.com

Sukmawaty

Program Studi Teknik Pertanian, Universitas Mataram

Ansar

Program Studi Teknik Pertanian, Universitas Mataram

Murad

Program Studi Teknik Pertanian, Universitas Mataram

Hanifah Ayu

Program Studi Teknik Pertanian, Universitas Mataram

Abstract

The purpose of this research is to build a model that can detect damage to avocados, using gas sensors that are accurate, precise, specific, and sensitive. The specific goal is to get the best model from the several machine learning methods used. The best model will be obtained through a classification algorithm. The input data to be processed is gas data in units of ppm measured from damaged and normal avocados that have been stored for 7 days. Gas data recording uses the TGS 2600, MQ-3, MQ-4, MQ-2, and MQ-8 sensors. Each data quality will be improved by cleaning and normalizing data. The quality-enhanced data is then used as input to the machine learning system. The machine learning model used is the Ada-boost, Decision Tree, Multi Layer Perceptron (MLP) classification model and the Stratified K-Fold Cross Validation method. The trained model is then tested with a comparison of the number of test data and training data which is 70:30. The test results show the accuracy of the avocado damage detection system using ppm data obtained using the gas sensor and the Stratified K-Fold Cross Validation method with system accuracy on the Ada-boost value having an accuracy of 100%, 100%, 100% and 50%, with a method accuracy value classifier with 4x repetitions.

Keywords: Avocado, Damage, Detection System, Classification Model, Stratified K-Fold Cross Validation method

INTRODUCTION

Avocado fruit (*Persea americana* Mill) is a fruit plant that is in great demand by consumers. This plant is spread across several regions in Indonesia, one of which is

¹ Correspondence author.

West Nusa Tenggara, which is one of the areas producing avocados, which is quite high in quantity. Data from the Central Agency for Fruit Production Statistics states that avocado production in Indonesia in 2021 will reach 669,260 metric tons, while the amount of avocado production in West Nusa Tenggara in 2021 will reach 8,889 metric tons (Central Statistics Agency, 2022).

Avocados have yellow or greenish-yellow flesh that is not fishy but aromatic and slightly fibrous (Andrianto, 2013). In previous studies of avocado ripening, there was an overall taste assessment given to the ripe fruit before harvest, which has shown that the taste of the avocado improves as ripening progresses (e.g., Harding, 1954; Lee et al., 1983).

There are several problems in the avocado classification process, whether the fruit is ripe, unripe, or damaged. Consumer acceptance of a product begins with an assessment of its appearance, taste, aroma, and texture (Astuti, 2018). One method that can be used is using an electronic nose (e-nose). The e-nose contains an array of sensors to detect the quality of the avocado based on the gas produced (Murad et al., 2022).

Changes in fruit quality are related to the length of storage time. In general, fruit is divided into two groups, namely climacteric and non-climacteric. After the fruit is harvested, it continues to carry out the respiration process. In climacteric fruit, the respiration rate is higher than in non-climacteric fruit. A high respiration rate will cause significant changes in the physical or chemical properties of the fruit (Kusumiyati et al., 2018). So the higher the level of ripeness of an avocado, the shorter its shelf life because the faster it will spoil and rot (Aprilliani et al., 2021). Ripe fruit is not necessarily sweet, and vice versa. Humans only know that sweet fruit is ripe (Adhimantoro, 2014). Therefore, this research will develop a system that can detect the ripeness level of avocados based on several parameters, namely weight loss, water content, total dissolved solids, and gas content, using the stratified K-fold cross-validation method.

The imbalanced dataset problem is often encountered in various domains, namely when the amount of data in each class is not balanced. Imbalanced data occurs when the amount of data in one class is much higher (majority class) or lower (minority class) compared to other classes. Imbalanced data will be more difficult to process and analyze for purposes such as classification, prediction, and so on. This happens because the data analysis model is not designed to consider class distribution to increase the accuracy of the model. Many studies that report data analysis with imbalanced data often provide erroneous results. So special handling is needed before the imbalanced data can be used for the data analysis process.

The most popular way to overcome the imbalanced data problem is to carry out resampling by changing the amount of data in each class until it reaches a balanced amount of data in all classes. Resampling is an effective technique. One

resampling method is undersampling. This method has been applied in several studies. Undersampling methods include Random Under Sampling (RUS), Edited Nearest Neighbors (ENN), One Side Selection (OSS), Neighborhood Cleaning Rule (NCR), and so on (Indrawati, 2021). Next, the resampling dataset will be classified using the Adaboost, Decision Tree, and Multi-Layer Perceptron (MLP) classification methods. After that, the test results will be compared using the stratified K-fold cross-validation method to obtain results in the form of accuracy, precision, and recall values. The best algorithm is then implemented into an information system prototype. The testing technique for the method will be carried out using stratified K-fold cross-validation.

Adaboost is combined with other methods to improve classification performance. Of course, intuitively, combining several methods will help if they are different from each other. Adaboost and its variants have been successfully applied in several fields (domains) because of their strong theoretical basis, accurate predictions, and great simplicity (Qadrini et al., 2021). In this way, it is hoped that a system will be able to classify avocados accurately and quickly during processing, which will later simplify the marketing process based on their quality.

Based on several parameters used, it is hoped that we can detect the ripeness level of avocados using machine learning. Machine learning is used to teach machines how to handle large amounts of data while still being efficient and to make machines learn themselves without being explicitly programmed (Batubara and Awangga, 2020). The problem formulation in this research is how to determine the level of ripeness, starting from unripe to ripe avocado fruit, based on physical and chemical properties using a classification algorithm and the Stratified K-Fold Cross Validation method.

From the explanation above, the problem can be formulated, namely how to create a system that can classify damage to avocados using TGS and MQ gas sensors with the Stratified K-Fold Cross Validation method. The aim of this research is to form a system from the results of machine learning on sensors to be able to classify avocado damage.

RESEARCH METHOD

In general, this research was carried out by conducting a study of relevant literature regarding avocado damage, the tools and algorithms that have been used to detect avocado damage, and the advantages and disadvantages of the methods that have been developed. The literature used is articles that have been published in national journals. Next, the design of an avocado damage detection tool using a gas sensor was carried out. The quality of the data obtained from the acquisition process is then improved. Improving data quality includes eliminating interference signals (noise) in signals and normalizing data.

After the quality of the data was improved, the process continued with the development of an avocado damage detection system. The system developed uses the Python programming language and is still a backend. The system built accommodates all research needs in the form of learning using the stratified K-fold cross-validation method. The results of learning using the stratified K-fold cross-validation method are then tested and analyzed to make reports in the form of scientific publications and research reports.

This research was conducted at the Bioprocess Engineering Laboratory, Faculty of Food Technology and Agroindustry, University of Mataram. The materials used in this research were local avocados that were normal and damaged due to storage. Research tools include a gas sensor and an Arduino. Development and testing of damage detection system models using the Python programming language.

RESULT AND DISCUSSION

Avocado Gas Data

After measuring the data set points, we continued with the gas measurement process for the avocado fruit samples. The number of avocados used was 40 with storage conditions in an open room. The avocado used is an avocado that is ripe. Storage was carried out for 4 days and the condition of the avocado was normal (1) or damaged (2). The data measured in this data recording process is in the form of ADC data. Raw data from measurement results are shown in Table 3.

Table 3. Raw data for gas measurements for avocado fruit samples for 29 days 3

No	====PuTTY log 2021.09.30 21:21:38 =====
3	TGS2600= 120 MQ-3= 610 MQ-4= 509 MQ-2= 178 MQ-8= 225
5	TGS2600= 120 MQ-3= 610 MQ-4= 509 MQ-2= 178 MQ-8= 225
6	TGS2600= 120 MQ-3= 610 MQ-4= 509 MQ-2= 178 MQ-8= 225
7	TGS2600= 121 MQ-3= 610 MQ-4= 508 MQ-2= 179 MQ-8= 225
8	TGS2600= 119 MQ-3= 609 MQ-4= 508 MQ-2= 177 MQ-8= 224
9	TGS2600= 119 MQ-3= 609 MQ-4= 507 MQ-2= 178 MQ-8= 224
.	.
.	.
.	.
622	TGS2600= 114 MQ-3= 598 MQ-4= 509 MQ-2= 175 MQ-8= 219
623	TGS2600= 114 MQ-3= 599 MQ-4= 508 MQ-2= 174 MQ-8= 219
624	TGS2600= 114 MQ-3= 598 MQ-4= 509 MQ-2= 175 MQ-8= 219
625	TGS2600= 114 MQ-3= 598 MQ-4= 509 MQ-2= 175 MQ-8= 219
626	TGS2600= 114 MQ-3= 598 MQ-4= 508 MQ-2= 174 MQ-8= 219

It can be seen from the raw data obtained that there is data that is not needed in the next process either because the data read is truncated or has all zero values. This condition requires a data cleaning process before it is used in the next process.

After obtaining the measurement data, proceed with the data processing. There are two stages carried out, namely data cleaning and data conversion. In this process, Python programming is used with the Pandas and Numpy libraries.

Data Cleaning

Data cleaning is carried out based on data recording conditions with incorrect results, as shown in Table 3. The process begins with reading the data, then dividing the data to be able to retrieve only the measured ADC value from each sensor. The Pandas library is used in this process. The result is a data table with size $M \times N$, where M represents the number of rows of data and N represents the columns, namely tgs26, mq3, mq4, mq2, and mq8. So each new table is $M \times 5$ in size. Example data from data cleaning is shown in Table 4.

Table 4. ADC measurement data from cleaning results

Index	Tgs26	MQ3	MQ4	MQ2	MQ8
5	120	610	508	179	225
6	120	610	509	178	225
7	121	610	508	179	225
8	119	609	508	177	224
9	119	609	507	178	224
.
.
.
308	114	598	509	175	219
309	114	599	508	174	219
310	114	598	509	175	219
311	114	598	509	175	219
312	114	598	508	174	219

The graph of the ADC data measured after cleaning is shown in Figure 3. It can be seen that the measured ADC data is in steady state condition. However, because the data displayed is still ADC data, it must be converted into ppm data.

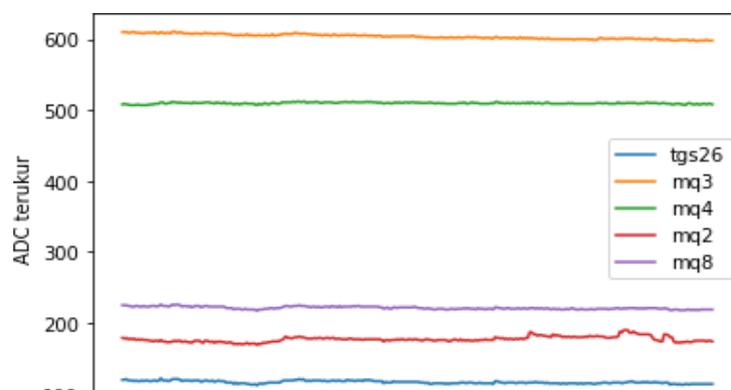


Figure 3. Graph of ADC Measurement Cleaning Results

Data Conversion

Data conversion is carried out to obtain the ppm value for each measured ADC value. The data needed to carry out the conversion is the measurement range data for each sensor, set points, measured ADC value, ADC 1023 scale, maximum input voltage V_{max} 5 V, and DAC value. The DAC value is obtained using the following equation:

$$DAC = \frac{ADC_{terukur}}{1024} \times V_{maks}$$

After obtaining the DAC value, it is then converted to ppm using the equation:

$$ppm_{sensor} = \frac{range_{maks_sensor} - range_{min_sensor}}{V_{maks}} \times DAC$$

The ADC value to be converted is first normalized. The aim is to make the initial value of each measured sensor value 0. Normalization is carried out by subtracting all measured ADC values from the set point value, then the results are reduced by the minimum value for each measurement result.

$$\begin{aligned} \Delta ADC &= ADC_{terukur} - set_point_{sensor} \\ \Delta ADC &= \Delta ADC - \min(\Delta ADC) \end{aligned}$$

The results of converting ADC to ppm are shown in Table 5.

Table 5 Data on conversion of ADC to ppm for sample 29 days 3

Index	Tgs26	MQ3	MQ4	MQ2	MQ8
7	9	13	1	10	8
8	7	12	1	8	7
9	7	12	0	9	7
10	7	13	0	8	6
11	6	13	0	9	6
.
.
.
308	2	1	2	6	2
309	2	2	1	5	2
310	2	1	2	6	2
311	2	1	2	6	2
312	2	1	1	5	2

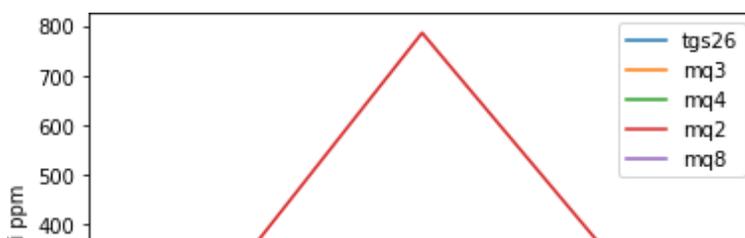


Figure 4. Sensor ppm value

The conversion results make the range of ppm values for each sensor visible and in the range 0–600. Meanwhile, the normalization results make the initial ppm value for each sensor 0. These results are shown more clearly in Figure 4.

Modeling and Simulation

The modeling and simulation process is carried out on the converted data. At this stage, all ppm data from the five gas sensors for seven days is used for each sample. Each sample is labeled raw (1) and cooked (2). The data used is as shown in Table 6.

Table 6. Input data

	Tgs26	MQ3	MQ4	MQ2	MQ8	Code
0	0,05	0,03	3,67	22,92	2,92	1
1	0,94	30,71	6,63	151,56	46,17	1
2	0,01	3,06	32,52	11,68	0,88	1
3	0,05	0,39	8,53	23,68	1,99	1
4	0,23	4,95	4,42	99,15	4,8	1
5	0,31	3,08	8,21	111,15	7,76	1
6	0,08	0,82	4,74	12,63	0,82	1
7	0,59	5,43	63,78	785,28	21,83	1
8	0,06	0,74	16,42	60,63	1,76	1
9	0,1	3,54	14,84	68,2	15,67	1
10	0,01	3,62	5,68	20,21	1,11	1
11	0,03	2,63	12,63	6	2,96	1
12	0,33	0,57	4,74	19,26	40,43	2
13	0,25	7,23	23,54	262,08	47,38	2
14	0,37	16,44	155,48	123,14	25,82	2

T
he
simul
ation
proc
ess
uses
Adab
oost,
Decis
ion
Tree,
Multi
-
Layer
Perc

eptron (MLP), and the Stratified K-Fold Cross Validation method. With this method, the best model will be obtained to determine the classification of normal or damaged avocados based on the gas they produce. Determining the best model is based on the values of accuracy (Acc), precision (Pr), sensitivity (Se), and specificity (Sp). The models that will be tested are Adaboost, Decision Tree, Multi-Layer Perceptron (MLP), and the Stratified K-Fold Cross Validation method. In this simulation, a 70:30 comparison scheme for training data and test data from a total of 40 samples is used.

During the data processing process, there is an imbalance between raw and ripe fruit. Balanced data is required for the classification process to be carried out. A very effective method to use to overcome the problem of imbalanced data is to carry out sampling by changing the amount of data in each class until it reaches a balanced amount of data in all classes.

Application of the Stratified K-Fold Cross Validation Method in Selecting Data

At this stage, coding is carried out using the Visual Studio Code and PuTTY applications. The stages of the sampling method carried out in the Visual Studio Code and PuTTY applications are entering the sampling method coding, bringing up the initial dataset, entering the random under sampling method coding, and then bringing up the dataset that has been carried out by the sampling method. The sampling method reduces the avocado data into 6 pieces, namely 3 raw avocados (1) and 3 ripe avocados (2), so that the data class becomes balanced.

At the classification stage, experiments were carried out using all classification methods, including the Adaboost method, GaussianNB, Decision Tree Classifier, Random Forest Classifier, Logistic Regression, SVC, K-Neighbors Classifier, and MLP Classifier. However, among the 7 classification methods, only 3 can be used, namely the Adaboost (adaptive boosting), Decision Tree Classifier, and Multi Layer Perceptron (MLP) method, because when it is run 10x, this method has no accuracy value. equals 0. The following classification results for avocado damage detection are shown in Table 4.7 below:

Table 7. Results of Avocado Damage Detection Classification

Method Name	Accuracy Value of Experiment Results				Average Results
	1	2	3	4	
Adaboost	1,0	1,0	1,0	0,5	0,875
Decision Tree Classifier	0,5	1,0	1,0	0	0,625
Multi Layer Perceptron (MLP)	1,0	0,5	0,5	0	0,5

It can be seen that the accuracy value of the Adaboost method in predicting the level of damage to avocados with imbalanced data is 87.5%, the Decision Tree Classifier method is 62.5% and the Multi Layer Perceptron (MLP) method is 50%. Of the three methods, the Adaboost method has the highest level of accuracy.

A. Testing

The test results of the avocado damage detection system based on input data using the Stratified K-Fold Cross Validation method are shown in Table 8. The system testing model uses the Adaboost, Decision Tree, Multi Layer Perceptron (MLP) method based on the confusion matrix value.

Table 8. Test results of the adaboost avocado damage detection system
Adaboost Method

Test	Accuracy	Precision	Sensitivity	Spesificity
AB1	1.0	0	0	1.0
AB2	1.0	0	0	1.0
AB3	1.0	0	0	1.0
AB4	0.5	0	0	1.0

Table 9. Test results of the avocado damage detection system using the decision tree method
Decision Tree Method

Test	Accuracy	Precision	Sensitivity	Spesificity
DT1	0.5	0	0	1.0
DT2	1.0	0	0	1.0
DT3	1.0	0	0	1.0
DT4	0	0	0	0

Table 10. Test results of the avocado damage detection system using the multi layer perceptron method
Multi Layer Perceptron Method (MLP)

Test	Accuracy	Precision	Sensitivity	Spesificity
MLP1	1.0	0	0	1.0
MLP2	0.5	0	0	1.0
MLP3	0.5	0	0	1.0
MLP4	0	0	0	0

Analysis

From the results of the tests that have been carried out as shown in Table 8 to Table 10, it shows that the avocado damage detection system model was built using the Stratified K-Fold Cross Validation method with 3 classification algorithms, namely Adaboost, Decision Tree and Multi Layer Perceptron (MLP). In the table above, the Adaboost value has the best accuracy. This is based on the accuracy values, which are 100%, 100%, 100%, and 50%, which are the highest compared to the accuracy values of other classifier methods in 4 repetitions.

Regarding the sensors used in this research, it can be concluded that the five sensors can be used to detect damage to avocados based on the ppm value of the gas produced. System accuracy of 100%, 100%, 100%, and 50% indirectly indicates a relationship with the sensor's ability to detect gas produced by fruit. This is because

the sensors used are not sensors specifically made to measure gas produced by fruit but rather gases that are used to detect gas in general. In fact, each sensor has a high level of sensitivity to certain gases. However, the combination of these sensors is capable of detecting fruit damage based on the gas it produces.

CONCLUSION

Based on the results of the research and discussions that have been carried out, the following conclusions are obtained:

1. The avocado damage detection system uses ppm data obtained with a gas sensor and the stratified K-fold cross-validation method with system accuracy at the Adaboost value having the best accuracy. This is based on the accuracy values, which are 100%, 100%, 100%, and 50%, which are the highest compared to the accuracy values of other classifier methods in 4 repetitions.
2. Integration of gas sensors types Tgs 2600, MQ-3, MQ-4, MQ-2, and MQ-8 can be used as an e-nose to detect gas produced by avocados.
3. The Adaboost, Decision Tree, Multi Layer Perceptron (MLP) classifier model used provides better performance compared to the performance of the respective classifiers that form Adaboost, Decision Tree, Multi Layer Perceptron (MLP), and the Stratified K-Fold Cross Validation method.

REFERENCES

- Adhimantoro, S. 2014. Mengetahui Tingkat Kematangan Buah Dengan Ultrasonik Menggunakan Logika Fuzzy. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 3(1), 63-68.
- Aprilliani, F., Atmiasih, D., & Ristono, A. 2021. Evaluasi Tingkat Kematangan Buah Alpukat (*persea americana* Mill.) dengan Teknologi Pengolahan Citra. *Jurnal Penelitian Pascapanen Pertanian*, 18(1), 1-8.
- Astuti, W. P. 2018. Substitusi Tempe untuk Meningkatkan Rasa dan Kadar Protein pada Abon Biji Durian (*Durio zibethinus* Murr). Skripsi. Pendidikan Biologi, Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Raden Intan, Lampung.
- Azrita, M. W., Ahmad, U., & Darmawati, E. 2019. Rancangan Kemasan dengan Indikator Warna untuk Deteksi Tingkat Kematangan Buah Alpukat. *Jurnal Keteknikan Pertanian*, 7(2), 155-162.
- Badan Pusat Statistik. 2022. Produksi Tanaman Buah-Buahan 2021. <https://www.bps.go.id/indicator/55/62/1/produksi-tanaman-buah-buahan.html> diakses pada tanggal 23 Agustus 2022.
- Barbosa, L. N., Carciofi, B. A. M., Danderhau C. E., & Monteiro, A. R. 2011. Influence of Temperature on the Respiration Rate of Minimally Processed Organic Carrots

- (*Daucus carota* L. cv. Brasilia). *Journal Food Science and Technology*, 31(1), 78-85.
- Batubara, N. A., & Awangga, R. M. 2020. Tutorial Object Detection Plate Number with Convolution Neural Network (CNN). Bandung: Kreatif Industri Nusantara.
- Brownlee, J. 2016. *Machine Learning Algorithms from Scratch with Python. Machine Learning Mastery*.
- Datasnips. 2021. Stratified K-Fold: What It Is & How to Use It. Towards Data Science. <https://towardsdatascience.com/stratified-k-fold-what-it-is-how-to-use-it-cf3d107d3ea2#:~:text=Stratified%20kfold%20cross%20validation%20is%20an%20extension%20of,the%20full%20dataset.%20Let%E2%80%99s%20look%20at%20an%20example> diakses pada tanggal 10 September 2022.
- Diantoro, K., Rahmadewi, R., & Ibrahim. 2020. Implementasi Sensor MQ 4 dan Sensor DHT 22 pada Sistem Kompos Pintar Berbasis IoT (SIKOMPI). *Jurnal Rekayasa dan Teknologi Elektro (Electrian)*, 14(3), 84-94.
- Frans. 2022. 18 Jenis Sensor Gas seri MQ. <https://www.anakteknik.co.id/rahasia1/articles/18-jenis-sensor-gas-seri-mq> diakses pada tanggal 12 Agustus 2022.
- Hapsari, R. K., & Indriyani, T. 2022. Implementasi Algoritma SMOTE Sebagai Penyelesaian Imbalance Hight Dimensional Datasets. Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika. Surabaya, 26 Maret 2022.
- Harianja, R. M. 2018. Alat Ukur Gas Hidrogen pada Air Menggunakan Sensor MQ-8. Tugas Akhir Program Studi D3 Metrologi dan Instrumentasi Departemen Fisika. Fakultas MIPA, Universitas Sumatera Utara, Medan.
- Hutagalung, D. D. 2018. Rancang Bangun Alat Pendeteksi Kebocoran Gas dan Api dengan Menggunakan Sensor MQ2 dan Flame Detector. *Jurnal Rekayasa Informasi*, 7(2), 43-53.
- Indrawati, A. 2021. Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset. *Jurnal Informatika dan Komputer (JIKO)*, 4(1), 38-43.
- Irawan, E., & Wahono, R. S. 2015. Penggunaan Random Under Sampling untuk Penanganan Ketidakseimbangan Kelas pada Prediksi Cacat Software Berbasis Neural Network. *Journal of Software Engineering*, 1(2), 92-100.
- Kumar, S. 2020. Understanding 8 types of Cross-Validation. Towards Data Science. <https://towardsdatascience.com/understanding-8-types-of-cross-validation-80c935a4976d> diakses pada tanggal 11 September 2022.
- Kusumiyati, K., Farida, F., Sutari, W., Hamdani, J. S., & Mubarok, S. 2018. Pengaruh Waktu Simpan terhadap Nilai Total Padatan Terlarut, Kekerasan dan Susut Bobot Buah Mangga Arumanis. *Jurnal Kultivasi*, 17(3), 766-771.

- Lee, S. K., Young, R. E., Schiffman, P. M., & Coggins, C. W. 1983. Maturity Studies of Avocado Fruit Based on Picking Dates and Dry Weight. *J. Amer. Soc. Hortic. Sci.*, 108(3), 390–394.
- Maarif, V., & Fadlilah, N. I. 2014. Pembuatan Alat Pengukur Tingkat Polusi Udara Berbasis Mikrokontroler AT89S51 Menggunakan Sensor TGS 2600. *Prosiding Seminar Nasional ReTII ke-9*. Yogyakarta, 13-14 Desember 2014.
- Merta, I. G. S., Widagda, I. G. A., & Paramarta, I. B. A. 2017. Perancangan Alat Ukur Kadar Alkohol Menggunakan Sensor MQ-3 Berbasis Mikrokontroler Atmega16. *Jurnal Buletin Fisika*, 18(2), 74-80.
- Murad, M., Sukmawaty, S., Ansar, A., Sabani, R., & Hidayat, S. 2022. Sistem Pendeteksi Kerusakan Buah Mangga Menggunakan Sensor Gas dengan Metode DCS-LCA. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 3(4), 186-194.
- Neter, J., Kutner, M., Wasserman, W., & Nachtsheim, C. 1996. *Applied Linear Statistical Models*. Fourth Edition. Boston, McGraw-Hill.
- Nisah, K., dan Barat, Y. M. 2019. Efek Edible Coating pada Kualitas Alpukat (*Persea americana* Mill.) Selama Penyimpanan. *Ar-Raniry Chemistry Journal*, 1(1), 11-17.
- Pah, Y. I., Mardjan, S. S., & Darmawati, E. 2020. Aplikasi Coating Gel Lidah Buaya pada Karakteristik Kualitas Buah Alpukat dalam Penyimpanan Suhu Ruang. *Jurnal Keteknikan Pertanian*, 8(3), 105-112.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Putri, K. 2019. 4 Library Python Terbaik untuk Machine Learning <https://teknologi.id/insight/4-library-python-terbaik-untuk-machine-learning/> diakses pada tanggal 8 Agustus 2022.
- Qadrini, L., Seppewali, A., & Aina, A. 2021. Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian*, 2(7), 1959-1966.
- Rohman, Y. A. 2019. Pengenalan Numpy, Pandas, Matplotlib. <https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib-b9obafd36co#:~:text=Matplotlib%20Matplotlib%20adalah%20library%20Python%20yang%20fokus%20pada,yang%20sesuai%20pada%20publikasi%20jurnal%20atau%20artikel%20ilmiah.> diakses pada tanggal 18 Agustus 2022.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang. *Indonesian Journal on Computer and Information Technology*, 5(1), 75-82.
- Saifudin, A., & Wahono, R. S. 2015. Pendekatan Level Data untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(2), 76-85.
- Siringoringo, R. 2018. Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor. *Journal Information System Development (ISD)*, 3(1), 44-49.

- Sukmawaty, S., Ahzani, M., & Putra, G. M. D. 2019. Karakteristik Buah Manggis, Alpukat, dan Jambu Biji pada Penyimpanan Suhu Rendah. *Jurnal Teknik Pertanian Lampung (Journal of Agricultural Engineering)*, 8(4), 280-292.
- Sumanto, B., Java, D. R., Wijaya, W., & Hendry, J. 2022. Seleksi Fitur terhadap Performa Kinerja Sistem E-Nose untuk Klasifikasi Aroma Kopi Gayo. *MATRIK: Jurnal Manajemen Teknik Informatika dan Rekayasa Komputer*, 21(2), 429-438.
- Syahrudin, A. N., & Kurniawan, T. 2018. Input dan Output pada Bahasa Pemrograman Python. *Jurnal Dasar Pemrograman Python STMIK*, January, 1-7.
- Tambunan, M.P.M., Ginting, Z., & Nurlaila, R. 2021. Pengaruh Suhu dan Waktu Hidrolisis terhadap Kadar Glukosa dalam Pembuatan Sirup Glukosa dari Biji Alpukat dengan Metode Hidrolisis Asam. *Chemical Engineering Journal Storage*, 1(3), 17-26.
- Trivusi. 2022. Metriks Evaluasi Sistem Menggunakan Confusion Matrix. <https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html#:~:text=Confusion%20Matrix%20merupakan%20metode%20evaluasi%20yang%20dapat%20digunakan,kombinasi%20berbeda%20dari%20nilai%20prediksi%20dan%20nilai%20aktual>. Diakses pada tanggal 8 Oktober 2022.
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011. The NumPy array: A Structure for Efficient Numerical Computation. *Journal Computing in Science & Engineering*, 13(2), 22-30.
- Yuliati, I. F., & Sihombing, P. R. 2021. Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 20(2), 417-426.
- Yusuf, S. A. A., & Hidayat, R. 2019. MFCC Feature Extraction and KNN Classification in ECG Signals. In 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE) (pp. 1-5). IEEE.