INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

Machine Learning-Based Naïve Bayes Classification of Pineapple Productivity: A Case Study in North Sumatra

Received:
21 April 2025
Accepted:
10 July 2025
Published:
31 August 2025

[1*]Suendri, [2]Rima Aprilia, [3]Ramadiani Br. Rambe, [4]Nur Haryani Zakaria
[1,2,3]Information Systems, Universitas Islam Negeri Sumatera Utara
[4]School of Computing, University Utara Malaysia
E-mail: [1]suendri@uinsu.ac.id, [2]rimaaprilia@uinsu.ac.id, [3]ramadianirambe@gmail.com, haryani, [4]zahurin @uum.edu.my
*Corresponding Author

Abstract— Background: Pineapple is a major agricultural commodity in Indonesia, especially in North Sumatra, where increasing demand calls for improved productivity. Although machine learning has been widely applied in agriculture, most prior studies on pineapple focus on fruit quality assessment or employ complex, less interpretable models, leaving a gap in lightweight and practical approaches for productivity classification. **Objective:** This study aims to evaluate the novelty and effectiveness of the Naïve Bayes algorithm in classifying pineapple productivity based on agronomic characteristics, addressing the underexplored use of this method for productivity prediction in pineapple cultivation. **Methods:** A descriptive quantitative approach was applied using secondary data from the Labuhan Batu Agricultural Extension Center, consisting of 52 records with seven agronomic parameters. The dataset was divided into 31 training and 21 testing samples, and the Naïve Bayes model was implemented using RapidMiner 7.1, with performance measured by accuracy. The small dataset size is recognized as a limitation that may affect generalizability. **Results:** The Naïve Bayes model achieved an accuracy of 86.67%, effectively distinguishing between productive and unproductive pineapples and demonstrating its suitability for agricultural classification tasks even with limited data. **Conclusion:** This study highlights the novelty and practicality of applying Naïve Bayes for pineapple productivity classification, offering an interpretable and computationally efficient alternative to more complex models. Future work should address dataset limitations by incorporating larger and more diverse samples and exploring hybrid or ensemble approaches to further enhance performance and support precision agriculture.

Keywords— Pineapple; Classification; Naïve Bayes; Machine Learning; North Sumatra

Corresponding Author:
Suendri,
Sistem Informasi,
UIN Sumatera Utara,
Email: suendri@uinsu.ac.id
Orchid ID: https://orcid.org/0000-0003-2736-646X

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

I. INTRODUCTION

Most of Indonesia's population relies on farming as their primary livelihood, making the country an agrarian nation [1], [2]. The rapid growth of the global population has led to increased food consumption, necessitating advancements in agricultural production to meet demand [3], [4]. Over the past two decades, agricultural activities in Indonesia have expanded beyond food and crop production to include processing, marketing, and distribution of agricultural products and livestock [2], [5]. The agricultural sector remains a primary contributor to the national economy by increasing Gross Domestic Product (GDP), supporting national trade, reducing unemployment, and supplying raw materials for various industries [6], [7].

Among Indonesia's agricultural exports, pineapple is a key commodity, particularly in North Sumatra. According to the Central Statistics Agency (BPS), pineapple exports in 2019 reached 236,226 tons, reflecting an increase from the previous year [5]. In 2018, Indonesia was the ninth-largest pineapple producer globally, with an annual production of 1.39 million tons [4]. Pineapple production supports both domestic consumption and export markets, contributing significantly to farmers' livelihoods and the national economy [8], [9].

To address the challenges of food security and sustainability, technological innovation in agriculture is essential. Machine learning (ML) has emerged as a transformative tool, leveraging big data and high-performance computing to optimize agricultural processes [1], [3], [10]. ML models have been widely applied for crop yield prediction, disease detection, resource optimization, and automated decision support in farming [7], [6], [9], [11]. Various computational approaches, such as neural networks, decision trees, random forests, and ensemble methods, have demonstrated success in agricultural applications [9], [12]–[14].

Among these, classification algorithms like support vector machines, Bayesian networks, and particularly Naïve Bayes have proven effective in addressing data-driven agricultural challenges due to their simplicity, accuracy, and efficiency in handling large datasets [10], [15], [16]. The Naïve Bayes algorithm, with its probabilistic foundation and minimal computational requirements, is especially suitable for agronomic data analysis and crop classification [10], [15]–[17].

Machine learning applications in agriculture now include automatic irrigation systems, drone-based field analysis, crop monitoring, precision farming, animal identification, and health monitoring [2], [9], [11], [13]. Data mining, as an integral process in ML, enables the extraction of useful knowledge from large agricultural datasets using techniques such as association rules, clustering, and classification [6], [7], [11]. The Naïve Bayes algorithm is particularly valued for its ability to manage large amounts of agronomic data efficiently, especially when interpretability and computational efficiency are required [10], [15], [16].

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

However, most research on machine learning and pineapple has primarily focused on fruit quality assessment—such as sugar content or ripeness—using nondestructive techniques like VIS/NIR spectroscopy or image processing [8], [9], [14]. Studies have also explored growth stage prediction and yield estimation using various ML models [9], [12]–[14]. While these approaches are important, they often overlook productivity classification, which is crucial for cultivation planning and economic output [10], [11].

Moreover, the application of Naïve Bayes for classifying pineapple productivity using structured agronomic data remains limited. Most existing studies favor more complex models, such as neural networks or ensemble methods, which may lack interpretability or require extensive computational resources [9], [12], [13]. Only a few recent works have demonstrated the effectiveness of Naïve Bayes in crop classification and recommendation systems, highlighting its potential for practical, lightweight decision support [10], [15], [17].

This study addresses this gap by proposing a Gaussian Naïve Bayes-based classification model trained on real-world agronomic data from the Labuhan Batu Agricultural Extension Center. By utilizing features such as harvest age, fruit size, and yield per cycle, the research evaluates the feasibility and accuracy of Naïve Bayes in predicting pineapple productivity. The expected outcome is a practical, interpretable decision support tool tailored for pineapple farmers and agricultural stakeholders in North Sumatra.

II. RESEARCH METHOD

This research employed a descriptive quantitative approach, utilizing secondary data collected from several pineapple-producing areas in North Sumatra, such as the Labuhan Batu Agricultural Extension Center (BPP). The study aimed to classify pineapples based on specific agronomic characteristics using the Naïve Bayes classification algorithm.

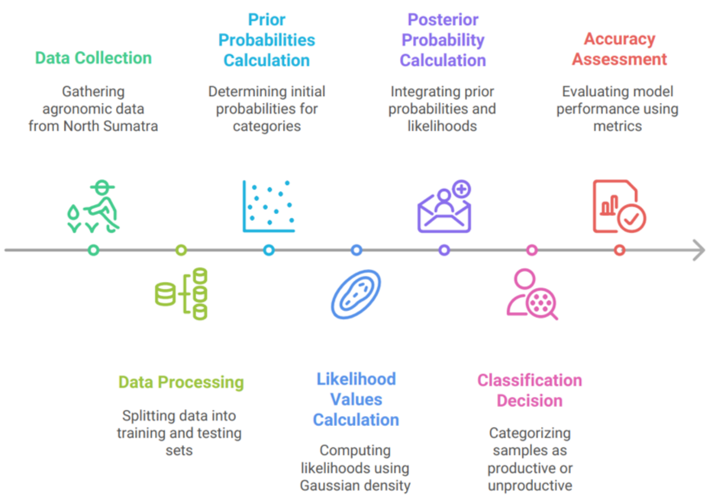

Fig 1. Research Workflow

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

The research workflow (Fig 1) began with Data Collection, where numerical secondary data on seven agronomic parameters harvest age, planting area, fruit diameter, fruit weight, fruit length, plant height, and production were gathered from pineapple-producing areas in North Sumatra. This raw data was then processed. In the Data Processing phase, the dataset was split into training and testing sets. The training data was used to calculate prior probabilities for each classification category and likelihood values for each parameter using the Gaussian density function. Next, the posterior probability for each sample was computed using Bayes' theorem, which integrated the prior probabilities and likelihood values. This calculation led to the Classification Decision, where each pineapple sample was categorized as either "productive" or "unproductive." Finally, in the Accuracy Assessment phase, the model's performance was evaluated using the testing dataset, with key metrics such as classification accuracy and Area Under the Curve (AUC) being used to assess the model's effectiveness.

A. Data Collection

The data used in this study were numerical, consisting of the following parameters:

1. Harvest age (days)
2. Planting area (ha)
3. Fruit diameter (mm)
4. Fruit weight (kg)
5. Fruit length (cm)
6. Plant height (cm)
7. Production (kg) per planting period

These parameters were chosen for their relevance in determining the productivity of pineapples [18].

B. Data Processing

1. Training and Testing Data Preparation:
   a. The dataset was divided into training and testing sets.
   b. The training data were used to calculate prior probabilities and likelihood values [19].
2. Calculating Probabilities:
   a. Prior Probability: The probability of each classification category was calculated based on the training data.
   b. Likelihood Values: For each parameter, the mean and standard deviation were computed. The Gaussian density function was used to calculate the likelihood values for numerical data [20].

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

3. Posterior Probability Calculation:

Using Bayes' theorem, the posterior probability was calculated by combining prior probability and likelihood values [21].

4. Classification Decision:

Based on the posterior probability, each sample was classified as productive or unproductive [22].

C. Accuracy Assessment

The model's accuracy was determined by comparing predictions against actual outcomes in the testing dataset. Key performance metrics included classification accuracy and the Area Under the Curve (AUC) [23]. While the model demonstrated good performance with an accuracy of 86.67%, it is important to note the limitation of the small dataset size (52 records), which may lead to overfitting and reduced generalizability. Moreover, cross-validation techniques such as k-fold or leave-one-out were not applied in this study to avoid further fragmenting the limited data. Future studies are encouraged to incorporate cross-validation to enhance the robustness and reliability of the model's evaluation.

III. RESULT AND DISCUSSION

A. Data collection

The pineapple dataset consisted of 52 records, with 31 records allocated for training. Based on data processing conducted at the Labuhan Agricultural Extension Center (BPP), the pineapples were categorized into two groups: 22 classified as "Yes" and 15 as "No". During the testing phase, the dataset was split into training and testing subsets, utilizing the Naïve Bayes algorithm. The training data was employed to construct a probability table, while the testing data was used to evaluate the established probabilities.

B. Read Training Data

The first step taken was to read the training data (table 1) and then group the variables based on pineapple classification between discrete data and continuous data [24]. The data obtained shows that there is no discrete data and only seven continuous data, including:

Continuous data

1. Harvest duration (days)
2. Cultivation area (hectares)
3. Diameter of the fruit (millimeters)
4. Weight of the fruit (kilograms)
5. Length of the fruit (centimeters)
6. Height of the plant (centimeters)
7. Pineapple yield per planting cycle (kilograms)

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

Table 1. Data Training

No	Type	Age of Harvest (days)	Planted Area (rante)	Fruit Diameter (mm)	Fruit Weight (kg)	Fruit Length (cm)	Plant Height (cm)	Pineapple Production (fruit)	Productive Category
1	An1	254	8	90,3	1,35	17	99	6480	Ya
2	An3	285	10	96,1	1,55	22	74	8100	Ya
3	An4	238	8	87	1	18	97	5200	Tidak
4	An6	238	8	86,6	1,25	17	96	6500	Ya
5	An7	392	8	92,9	0,83	18	128	2400	Tidak
…	…	…	…	…	…	…			
37	An52	254	4	92,6	1,23	17	99	3160	Ya

C. Calculating Probabilities *Prior*

Then the second step is to search for classification using the method *naïve bayes*, namely: finding the probability of each pineapple. Predictions of pineapple productivity will be determined by two categories (table 2), namely "Yes" and "No". Calculation of probability by finding the number of Yes and No data from the total training data, then dividing it by the total data [25].

Table 2. Prior Probability

Pineapple Probability	
Probability	
Of	**No**
31	21
31/37	21/37

D. Count *Likelihood value*

The likelihood for each class is calculated by multiplying the Gaussian probabilities of each attribute, following the approach in [26]. Counting steps *likelihood*:

1. Calculating Values *Mean* and Standard Deviation

The next step that must be taken is to determine the average or mean value and standard deviation for each attribute, including: harvest age (days), planting area (ha), fruit diameter (mm), fruit weight (kg), fruit length (cm), and pineapple production in one planting period (kg). The Gaussian Naïve Bayes algorithm was applied, where the mean and standard deviation for each feature were computed for both classes. These statistical parameters were then used to derive likelihood values using the Gaussian probability density function. Table 3 and Table 4 present the

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

computed values. *Mean* (table 3) and standard deviation (table 4) for each category (Yes and No) of the seven attributes as follows:

Table 3. Mean

MEAN							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
OF	313,0454	8,2727	90,3	1,2440	17,2409	84,6363	5193,6363
NO	298,0666	7,7333	88,7866	0,8966	14,7333	90,4	4089,3333

Table 4. Standar Deviation

STANDARD DEVIATION							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
OF	59,5934	2,6935	16,7635	0,1385	1,9335	16,5026	1789,6994
NO	49,4201	2,6850	4,7755	0,1549	1,8790	19,4157	1439,8981

2. Prediction calculation with *Naïve Bayes* with Gaussian Dentity Function

After obtaining the mean and standard deviation values for each attribute, the next stage is: calculating *method naïve bayes* with the identity formula *Gauss* (table 5). To classify pineapple as productive or unproductive, for example if you know: type of pineapple, harvest age 440 days, planting area 11 ha, fruit diameter 96 mm, fruit weight 1.9 kg, fruit length 20 cm, plant height 90 cm, pineapple production 8900 pieces.

Table 5. Dentitis Gaus

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
OF	0,005344749	0,145627111	0,091988154	1,45005E-05	0,103680254	0,093176145	0,001104915
NO	0,000918309	0,116180976	0,058354673	7,96739E-10	0,00573091	0,090542171	3,96305E-05

Using the Naïve Bayes formula, posterior probabilities were computed:

a. $P(Yes) = 9.28 \times 10^{-15}$

b. $P(No) = 5.78 \times 10^{-23}$

Since the posterior probability for "Yes" is significantly higher, the sample is classified as productive.

3. Calculating Probabilities *Posterior*

Based on the likelihood values calculated in the previous section, posterior probabilities were computed to determine the most likely class for the input sample. The posterior probability for the "productive" class (Yes) was found to be 0.999999994, while the probability for the "unproductive" class (No) was 0.0000000062, as shown in the formulas above. Since the

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

probability of "Yes" is significantly higher and close to 1, the sample is classified as a productive pineapple. This confirms that the model is confident in its classification based on the given input values

$$P\ (X\ |Of) = \frac{9{,}28486_{-15}\ x\ 10^{-15}\ \ _{-23}}{(9{,}2846\ x\ 10+5{,}78945\ x\ 10)} = 0{,}999999994$$

$$P\ (X\ |No) = \frac{1{,}18663_{-15}\ x\ 10^{-21}\ \ _{-23}}{(9{,}28486\ x\ 10+5{,}78945\ x\ 10)} = 0{,}0000000062$$

4. Finding the Maximum Likelihood Value

The classification decision is based on the posterior probability values. Since the probability for the "Yes" class is 0.999999994, significantly higher than the probability for "No" (0.0000000062), the sample is classified as productive. The classified sample corresponds to a pineapple with the following attributes: harvest age of 440 days, planting area of 11 ha, fruit diameter of 96 mm, fruit weight of 1.9 kg, fruit length of 20 cm, plant height of 90 cm, and yield of 8900 fruits.

E. Analysis of Naïve Bayes with RapidMiner 7.1

This study utilized RapidMiner 7.1 to validate the accuracy of the Naïve Bayes classification algorithm in predicting pineapple productivity. The dataset, obtained from the Labuhan Batu Agricultural Extension Center (BPP), was first validated and then processed through a structured workflow in RapidMiner, as illustrated in Figure 2. The testing phase assessed the alignment between predicted and actual classes to ensure consistency and reliability.

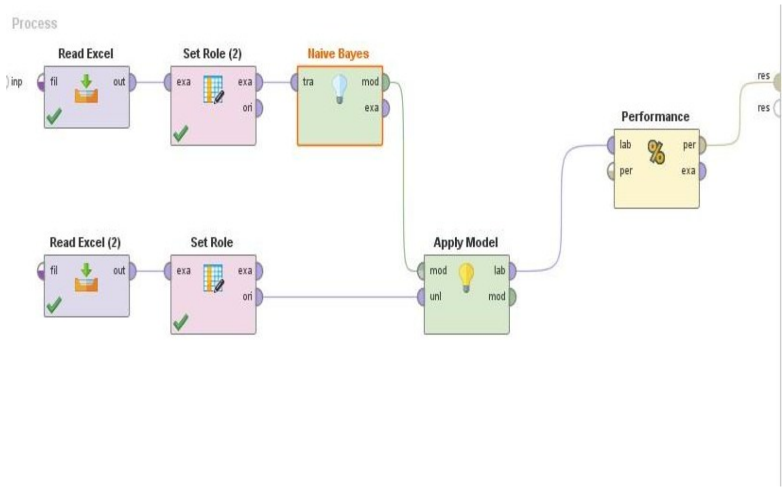

Fig 2. RapidMiner

Based on the evaluation, the model achieved an accuracy rate of 86.67%, indicating strong performance despite the relatively small dataset. This supports previous findings that machine learning models, including Naïve Bayes, are capable of producing accurate classifications in

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

agricultural domains [16], [15]. Moreover, studies incorporating IoT and machine learning have shown promising results in enhancing precision agriculture practices [27], [28]. Exploring hybrid models that integrate ML and deep learning can further improve prediction capabilities, particularly for crop yield and quality assessments [29], [30], [31], [32].

IV. CONCLUSION

This study demonstrated that the Naïve Bayes classification algorithm is capable of effectively classifying pineapple productivity using agronomic features. With an accuracy of 86.67%, the model provides a lightweight and interpretable solution suitable for supporting early decision-making in agricultural practices, particularly for pineapple cultivation in North Sumatra. However, the findings must be interpreted in light of certain limitations, including the small dataset size and the absence of cross-validation, which may affect generalizability. Future work should consider expanding the dataset, integrating validation techniques, and comparing multiple classification algorithms to improve robustness. Overall, the research highlights the practical potential of Naïve Bayes in agricultural prediction tasks and its value in enhancing data-driven farming strategies.

Author Contributions: *Suendri:* Conceptualization, Methodology, Writing - Original Draft, Supervision. *Rima Aprilia:* Investigation, Data Curation, Writing - Original Draft. *Ramadiani Br. Rambe:* Data Analysis, Literature Review, Editing. *Nur Haryani Zakaria:* Literature Review, Editing

All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Operational Grant for Higher Education (BOPTN) from the Ministry of Religious Affairs, 2023.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: Data is available upon request by contacting the corresponding author.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Suendri: https://orcid.org/0000-0003-2736-646X
Rima Aprilia: https://orcid.org/0000-0002-1860-3876
Ramadiani Br. Rambe: https://orcid.org/0009-0005-1273-7215
Nur Haryani Zakaria: https://orcid.org/0000-0001-5971-1307

REFERENCES

[1] S. K. Tripathi, A. S. Raghav, and S. K. Singh, “Machine learning in agriculture: A comprehensive updated review,” Computers and Electronics in Agriculture, vol. 202, p. 107437, 2023, **doi:** https://doi.org/10.3390/s21113758

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

[2] U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, and N. Iqbal, “Precision Agriculture Techniques and Practices: From Considerations to Applications,” Sensors, vol. 19, no. 17, p. 3796, Sep. 2019, **doi:** https://doi.org/10.3390/s19173796

[3] D. P. Akurati, P. Devarahatti, S. K. Konakandla, C. S. S. Satwik, and B. Singh, “Transforming Agriculture with Machine Learning: A Review of Applications and Future Prospects,” in 2024 Parul International Conference on Engineering and Technology (PICET), IEEE, May 2024, pp. 1–5. **doi:** https://doi.org/10.1109/picet60765.2024.10716047

[4] FAO, "Agricultural production statistics 2010–2023," FAOSTAT Analytical Briefs, No. 96. Rome, 2024. Diakses: 29 Juli 2024. [Daring]. Tersedia di: https://openknowledge.fao.org/handle/20.500.14283/cd3755en

[5] Pusat Data dan Sistem Informasi Pertanian, "Outlook Komoditas Pertanian Hortikultura Nanas," Sekretariat Jenderal Kementerian Pertanian, 2023. Diakses: 30 Juli 2024. [Daring]. Tersedia di: https://satudata.pertanian.go.id/assets/docs/publikasi/Outlook_Nenas_2023.pdf

[6] N. Gandhi and L. J. Armstrong, “A review of the application of data mining techniques for decision making in agriculture,” in 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Dec. 2016, pp. 1–6. **doi:** https://doi.org/10.1109/ic3i.2016.7917925

[7] T. van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” Computers and Electronics in Agriculture, vol. 177, p. 105709, Oct. 2020, **doi:** https://doi.org/10.1016/j.compag.2020.105709

[8] R. G. de Luna et al., “A Machine Learning-Based Approach for Accurate Size Classification of Pineapple (Ananas Comosus),” in TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), IEEE, Oct. 2023, pp. 1034–1039. **doi:** https://doi.org/10.1109/tencon58879.2023.10322473

[9] R. Wan Nurazwin Syazwani, H. Muhammad Asraf, M. A. Megat Syahirul Amin, and K. A. Nur Dalila, “Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning,” Alexandria Engineering Journal, vol. 61, no. 2, pp. 1265–1276, Feb. 2022, **doi:** https://doi.org/10.1016/j.aej.2021.06.053

[10] D. Vernanda, Z. Nurizati, and A. Hidayat, “Klasifikasi Buah Nanas Menggunakan Metode Naïve Bayes,” Melek IT : Information Technology Journal, vol. 10, no. 2, pp. 111–120, Dec. 2024, **doi:** https://doi.org/10.30742/melekitjournal.v10i2.355

[11] M. Waqas, A. Naseem, U. W. Humphries, P. T. Hlaing, P. Dechpichai, and A. Wangwongchai, “Applications of machine learning and deep learning in agriculture: A

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

comprehensive review," Green Technologies and Sustainability, vol. 3, no. 3, p. 100199, Jul. 2025, **doi:** https://doi.org/10.1016/j.grets.2025.100199

[12] H. Afzal et al., "Incorporating soil information with machine learning for crop recommendation to improve agricultural output," Scientific Reports, vol. 15, no. 1, Mar. 2025, **doi:** https://doi.org/10.1038/s41598-025-88676-z

[13] Ö. Şimşek, "Harvesting sustainability: Innovations and practices in modern agriculture," Green Technologies and Sustainability, vol. 3, no. 3, p. 100192, Jul. 2025, **doi:** https://doi.org/10.1016/j.grets.2025.10019

[14] M. Mohd Ali, N. Hashim, S. Abd Aziz, and O. Lasekan, "Pineapple (Ananas comosus): A comprehensive review of nutritional values, volatile compounds, health benefits, and potential food products," Food Research International, vol. 137, p. 109675, Nov. 2020, **doi:** https://doi.org/10.1016/j.foodres.2020.109675

[15] D. Munjal, L. Singh, M. Pandey, and S. Lakra, "A Systematic Review on the Detection and Classification of Plant Diseases Using Machine Learning," International Journal of Software Innovation, vol. 11, no. 1, pp. 1–25, Dec. 2022, **doi:** https://doi.org/10.4018/ijsi.315657

[16] E. Im, J. Lee, S. An, and G. Gim, "A Study on Prediction Performance Measurement of Automated Machine Learning," International Journal of Software Innovation, vol. 11, no. 1, pp. 1–11, Dec. 2022, **doi:** https://doi.org/10.4018/ijsi.315656

[17] W. D. Septiani and M. Marlina, "Comparison Of Decision Tree, Naïve Bayes, And Neural Network Algorithm For Early Detection Of Diabetes," Jurnal Pilar Nusa Mandiri, vol. 17, no. 1, pp. 73–78, Mar. 2021, **doi:** https://doi.org/10.33480/pilar.v17i1.2213

[18] B. Brumen, I. Rozman, and A. Černezel, "Observing a Naïve Bayes Classifier's Performance on Multiple Datasets," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2014, pp. 263–275. **doi:** https://doi.org/10.1007/978-3-319-10933-6_20

[19] Y. Lian, J. Hao, W. Zeng, and Q. Luo, "A survey of visual insight mining: Connecting data and insights via visualization," Visual Informatics, p. 100271, Aug. 2025, **doi:** https://doi.org/10.1016/j.visinf.2025.100271

[20] J. Botero-Valencia et al., "Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives," Agriculture, vol. 15, no. 4, p. 377, Feb. 2025, **doi:** https://doi.org/10.3390/agriculture15040377

[21] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," Technologies, vol. 9, no. 3, p. 52, Jul. 2021, **doi:** https://doi.org/10.3390/technologies9030052

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

[22] M. U. Maheswari and R. Ramani, “A Comparative Study of Agricultural Crop Yield Prediction Using Machine Learning Techniques,” in 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, Mar. 2023, pp. 1428–1433. **doi:** https://doi.org/10.1109/icaccs57279.2023.10112854

[23] A. Urtubia, J. R. Pérez-Correa, A. Soto, and P. Pszczólkowski, “Using data mining techniques to predict industrial wine problem fermentations,” Food Control, vol. 18, no. 12, pp. 1512–1517, Dec. 2007, **doi:** https://doi.org/10.1016/j.foodcont.2006.09.010

[24] J.-W. Chang and Y.-K. Kim, “An Efficient Clustering Method for High-Dimensional Data Mining,” in Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 276–285. **doi:** https://doi.org/10.1007/978-3-540-28645-5_28

R. M. Redwan, A. Saidin, and S. V. Kumar, “The draft genome of MD-2 pineapple using hybrid error correction of long reads,” DNA Research, vol. 23, no. 5, pp. 427–439, Jul. 2016, **doi:** https://doi.org/10.1093/dnares/dsw026

[26] G. Shanthakumari, A. Vignesh, R. Harish, and R. Roshan Karthick, “Advancements in Smart Agriculture: A Comprehensive Review of Machine Learning and IOT Approaches,” in 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), IEEE, Apr. 2024, pp. 1–6. **doi:** https://doi.org/10.1109/ic3iot60841.2024.10550268

[27] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” SN Computer Science, vol. 2, no. 3, Mar. 2021, **doi:** https://doi.org/10.1007/s42979-021-00592-x

[28] K. Yeturu, “Machine learning algorithms, applications, and practices in data science,” in Handbook of Statistics, Elsevier, 2020, pp. 81–206. Accessed: Aug. 31, 2025. [Online]. **doi:**: https://doi.org/10.1016/bs.host.2020.01.002

[29] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, “Machine Learning in Agriculture: A Comprehensive Updated Review,” Sensors, vol. 21, no. 11, p. 3758, May 2021, **doi:** https://doi.org/10.3390/s21113758

[30] T. Ayoub Shaikh, T. Rasool, and F. Rasheed Lone, “Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming,” Computers and Electronics in Agriculture, vol. 198, p. 107119, Jul. 2022, **doi:** https://doi.org/10.1016/j.compag.2022.107119

[31] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, Apr. 2018, **doi:** https://doi.org/10.1016/j.compag.2018.02.016

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24034

[32] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," Computers and Electronics in Agriculture, vol. 143, pp. 23–37, Dec. 2017, **doi:** https://doi.org/10.1016/j.compag.2017.09.037