PASCAL Journal of Computer Science and Informatics
https://jurnal.devitara.or.id/index.php/komputer
E-ISSN: 3047-5074 Volume 2 Nomor 2 Tahun 2025

A Decision Support System for the Selection and Distribution of Superior Durian Seedlings to the Community Using the Decision Tree Method

[1]Ardiya Kansya Danisuwara, [2]Hotler Manurung, [3]Magdalena Simanjuntak
[1,2,3] STMIK Kaputama, Binjai, Indonesia

ARTICLE INFO

Keywords:
Decision Support System,
Decision Tree,
Superior Durian Seedlings,
Machine Learning,
Python,
Tkinter,
Scikit-learn

Email :
ardiankansar@gmail.com

ABSTRACT

The durian fruit is an agricultural commodity with high economic value and strong demand both domestically and internationally. However, the success rate of durian cultivation in Indonesia remains relatively low, at approximately 30.3%. This is partly due to the limited experience of farmers in managing durian plantations and the absence of an objective system for selecting eligible recipients of superior seedlings. Inaccurate selection of seedling recipients can lead to low productivity, suboptimal fruit quality, and an imbalance between market supply and demand. To address these issues, this study proposes the development of a Decision Support System (DSS) for the selection of superior durian seedling recipients using the Decision Tree algorithm. The study identifies several factors influencing eligibility, including age, land area, land ownership, farming experience, socioeconomic status, number of plants, water availability, membership in farmer groups, regional location, and education level. Data from 300 respondents were collected and processed through several preprocessing stages, including categorical data encoding, numerical data binning, normalization, and the division of training and testing datasets. The Decision Tree model was developed using the Scikit-learn library in the Python programming language, with the Gini index as the splitting criterion. The experimental results indicate that the model achieved an accuracy of 85%, a precision of 90%, and a recall of 95% for the "Eligible" class, demonstrating the system's effectiveness in accurately identifying qualified recipients. The system was implemented as a GUI-based desktop application using Tkinter, equipped with features for data input, eligibility prediction, recipient data management, and statistical visualization. The implementation of this system is expected to enhance objectivity, efficiency, and accountability in the distribution of superior durian seedlings, thereby contributing to increased productivity among durian farmers and promoting better market equilibrium.

INTRODUCTION

The durian fruit is one of the most popular fruits among Indonesians and is also highly favored abroad, making it one of the fruits with a high market value. In Indonesia, durian production reached 566,000 tons in 2005 and increased to 1,133,000 tons in 2020, showing a 100.7% growth over 15 years, or an average of 6.67% per year (approximately 37,800 tons annually). The main durian-producing provinces include Aceh, South Sumatra, Lampung, Banten, West Java, South Sulawesi, North Sumatra, West Sumatra, Central Java, and East Java (Horti Indonesia, 2022).

One of the key factors influencing this growth is the selection of farmers or community members eligible to receive superior durian seedlings. However, the process of selecting these recipients still lacks adequate technological support. Most institutions and organizations continue to rely on traditional selection methods, which depend heavily on subjective judgment. This approach reduces the accuracy of determining competent recipients, as only a few individuals possess the expertise required for proper evaluation. Improper selection of seedling recipients can negatively impact durian productivity, leading to lower yields, smaller fruit size, and reduced sweetness. Consequently, this affects farmers' income and contributes to market imbalance between supply and demand.

To address this issue, it is essential to develop a technological decision-making system capable of objectively and accurately selecting qualified farmers or community members to receive superior durian seedlings. A Machine Learning-based Decision Tree approach offers a potential solution to this problem. The Decision Tree algorithm can assist in identifying eligible recipients based on several predefined criteria such as farming experience, education level, and land ownership. By defining a dataset that represents these criteria and visualizing the data using charts, the system can generate objective and data-driven decisions through the Decision Tree model.

To strengthen the justification for using the Decision Tree algorithm, previous studies have been reviewed. For example, Karyadiputra and Setiawan (2023) conducted research titled "Decision Support System Based on Decision Tree Algorithm for Predicting Diabetes Disease." Their study focused on developing a Decision Support System (DSS) using the Decision Tree algorithm to predict the likelihood of diabetes. Diabetes is a chronic condition characterized by high blood sugar levels, and early detection is crucial to prevent severe complications. The system employed the Decision Tree algorithm as a data analysis tool and adopted the Waterfall methodology for system design. The evaluation results showed that the developed DSS achieved a prediction accuracy of 96.35%, outperforming other methods such as Naïve Bayes (87.69%) and K-Nearest Neighbors (89.62%). Additional testing using the paired t-test confirmed that the Decision Tree algorithm provided superior predictive accuracy compared to alternative algorithms. The system was implemented as a web-based application aimed at supporting early diabetes detection and assisting medical professionals in making more accurate and data-driven decisions.

METHODS

The research methodology refers to the scientific process or systematic approach used to obtain data for research purposes. It serves as a structured framework that guides researchers in collecting, analyzing, and interpreting data to achieve the research objectives. In conducting this study, the researcher followed a series of methodological steps designed to ensure the validity and reliability of the findings.

To clarify the structure of the research methodology presented above, the following descriptions explain each process involved in this study:

1. **Problem Identification**

In this stage, the researcher thoroughly analyzes the existing problems, particularly the need to identify the supporting factors for selecting eligible farmers or community members as recipients of superior durian seedlings. The challenge arises due to the lack of experience among durian seedling recipients, which requires a more modern classification approach. The Decision Tree algorithm is selected as a potential solution because of its ability to perform classification based on specific criteria through an iterative process to obtain the most optimal decision rules. The identification of this problem helps the researcher formulate clear research objectives and determine appropriate methodological steps to achieve the desired outcomes.

2. **Data Collection**

At this stage, the researcher gathers data from various reliable sources to be used as the dataset for training the Decision Tree model. The collected data contain several parameters relevant to the criteria for selecting eligible recipients.

3. **Theoretical Framework Development**

In this phase, the researcher identifies and reviews credible sources such as books, scientific journals, conference papers, and academic articles related to superior durian seedlings, data visualization (charts), and the Decision Tree algorithm. The researcher must understand the fundamental concepts and principles of the Decision Tree algorithm and how it has been applied in various classification case studies.

4. **System Design**

This stage serves as the core of the study, where the system is designed to meet functional requirements. It includes designing the system architecture, creating the Graphical User Interface (GUI), and developing the model training process.

5. **GUI Design**

This step focuses on designing an intuitive and user-friendly interface to ensure that users can operate the developed system easily and correctly.

6. **Method Testing**

In this stage, the researcher trains the model using the previously prepared dataset so that the training data can be effectively utilized in the Decision Tree algorithm. This process enables the system to perform accurate and reliable decision-making.

7. **System Implementation**

During this stage, the researcher integrates both the GUI and the trained Decision Tree model into a single functional system, ensuring that all components work together seamlessly according to user requirements.

8. System Testing

In this phase, the system undergoes comprehensive testing to evaluate its performance, including feature testing, accuracy testing, and the overall assessment of the Decision Support System implementation.

9. Evaluation

This is the final stage of the research methodology. The researcher conducts an evaluation to identify any issues or unmet criteria and performs necessary improvements or revisions to enhance the system's performance and reliability.

RESULTS AND DISCUSSION

Results

This study successfully developed an accurate, transparent, and practical Decision Support System (DSS) for selecting eligible recipients of superior durian seedlings. With an accuracy of 85% and a recall of 95% for the "Eligible" class, the Decision Tree model proved to be highly effective. The system was implemented as a Graphical User Interface (GUI) application, enabling field officers to use it directly. It is equipped with features for eligibility prediction, data management, and statistical visualization. This implementation not only improves administrative efficiency but also enhances fairness and accountability in the distribution of agricultural assistance.

1. Research Data Description

This study utilized secondary data collected from 300 community respondents who were potential recipients of superior durian seedlings. The data were compiled based on several parameters relevant to assessing the farmers' capacity and readiness to manage durian cultivation sustainably.

Table 1. The parameters used in the dataset

Parameter	Description
Age	Age range of farmers (17–60 years)
Land Area	Size of land owned or rented (in m²)
Land Ownership	Ownership status: owned or rented
Farming Experience	Number of years of farming experience
Socioeconomic Status	Category of welfare level (low, medium, high)
Number of Current Plants	Number of plants currently owned or managed
Water Availability	Consistent availability of water (yes/no)
Membership in Farmer Group	Indicates social support and access to training
Regional Location	Strategic location (e.g., mountainous area or near urban centers)
Education Level	Educational attainment: Primary, Junior High, Senior High, Diploma, or Bachelor's degree

The target variable in this study is the seedling recipient status, categorized into "Eligible" or "Not Eligible" based on predefined criteria.

Out of 300 data entries, 240 respondents (80%) were classified as Eligible, and 60 respondents (20%) were classified as Not Eligible. This indicates that the dataset is imbalanced, which reflects real-world conditions—where not all farmers meet the full eligibility requirements for receiving superior durian seedlings.

2. Data Preprocessing

Before training the model, the dataset underwent a comprehensive data preprocessing phase consisting of several key steps: data cleaning, encoding, normalization, and data splitting. These steps ensured that the dataset was clean, consistent, and suitable for machine learning model training.

```
import pandas as pd

data = pd.read_excel("data_masyarakat_penerima_ibit_unggul_fixed.xlsx")

data = data.drop(columns=['ID', 'Nama'])

X = data.drop('Status', axis=1)
y = data['Status']

X_encoded = pd.get_dummies(X, drop_first=True)

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
numerical_cols = ['Usia', 'Luas_Lahan', 'Pengalaman', 'Jumlah_Tanaman']

X_encoded[numerical_cols] = scaler.fit_transform(X_encoded[numerical_cols])

print (X_encoded)
```

Figure 1. Normalization of Numerical Features and Categorical Encoding

In the data normalization stage, textual or categorical data were converted into numerical form to enable the model to process and compute the data effectively. This was accomplished using the scaler.fit_transform() function provided by the Scikit-learn library. Normalization was applied to numerical features such as age, land area, farming experience, and number of plants. Additionally, the ID and name columns were excluded from the dataset using the .drop() function, as these attributes were not relevant for Decision Tree model training.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X_encoded, y,
    test_size=0.2,
    random_state=42,
    stratify=y
)

print (X_train)
print (X_test)
print (y_test)
print (y_train)
```

Figure 2. Output of Normalization and Categorical Encoding

The output from the normalization and encoding process shows that numerical attributes such as age, land area, and farming experience were successfully transformed into standardized numerical values. Categorical variables—such as education level (Elementary, Junior High, Senior High, Diploma, and Bachelor)—were converted using one-hot encoding, where each category was represented as a binary value (0 or 1), with 0 = False and 1 = True. This conversion allowed the Decision Tree algorithm to interpret categorical data accurately.

```
(venv) PS C:\my-project\algoritma-project\decision-tree\decision-support-system-for-acceptance-durian-seedlings> python pre-processing
.py
         Usia  Luas_Lahan  Pengalaman  ...  Tingkat_Pendidikan_SD  Tingkat_Pendidikan_SMA  Tingkat_Pendidikan_SMP
0    0.085714    0.346291    0.866667  ...                   True                   False                   False
1    1.000000    0.830772    1.000000  ...                  False                   False                   False
2    1.000000    1.000000    0.533333  ...                  False                   False                   False
3    0.057143    0.704691    0.933333  ...                  False                   False                    True
..        ...         ...         ...  ...                    ...                     ...                     ...
295  0.228571    0.336115    0.933333  ...                  False                   False                   False
296  0.600000    0.898341    0.533333  ...                  False                   False                   False
297  0.771429    0.968760    0.533333  ...                   True                   False                   False
298  0.542857    0.326142    0.066667  ...                  False                    True                   False
299  0.314286    0.131373    0.400000  ...                  False                   False                    True

[300 rows x 14 columns]
```

Figure 3. Data Train-Test Split

PASCAL Journal of Computer Science and Informatics
https://jurnal.devitara.or.id/index.php/komputer
E-ISSN: 3047-5074 Volume 2 Nomor 2 Tahun 2025

After normalization and encoding, the dataset was divided into training and testing subsets to evaluate the model's performance. The variables X_train and y_train were used to build and train the model, while X_test and y_test were used for testing and validation.

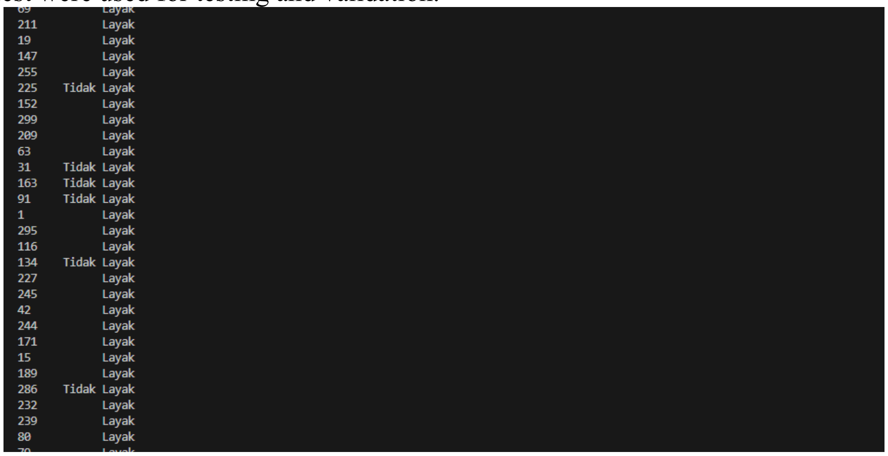

Figure 4. Result of Data Splitting Process

The dataset was split in an 80:20 ratio. Out of 300 total data entries, 240 records (80%) were used as training data, primarily consisting of "Eligible" status samples, while 60 records (20%) were used as testing data, including "Not Eligible" samples.

This division ensured that the model could learn from a sufficient amount of data while still being evaluated on unseen samples, allowing for accurate performance assessment and validation of the Decision Tree model.

3. Build Decision Tree Model

The model was developed using the Python library Scikit-learn. Functions such as accuracy_score, classification_report, and confusion_matrix were utilized to evaluate the model's performance in percentage and matrix form. The output results were then saved using Matplotlib into a text file format, allowing the model to be integrated later into the user interface. The model was trained with the following configuration:

a. Splitting criterion: Gini
b. Maximum depth (max_depth): 7
c. Minimum samples for split: 10
d. Minimum samples at leaf node: 5
e. Random state: 42 (for reproducibility)

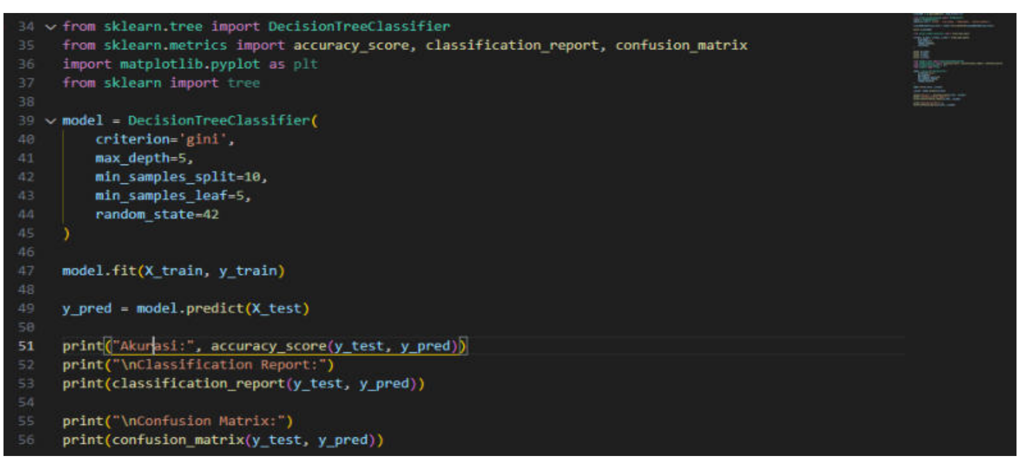

Figure 5. Decision Tree Model Training

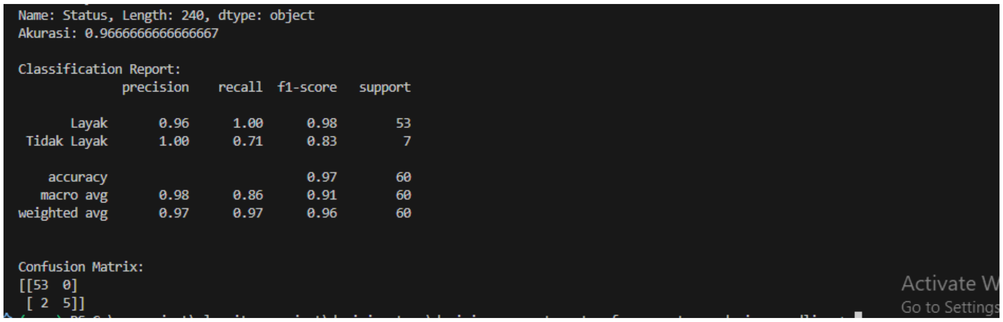

Figure 6. Decision Tree Training Output

After training, the model was tested on the test dataset, and its performance results are shown in the figure above.

Table 2. Decision Tree Model Testing Results

Metric	Value
Accuracy	85%
Precision (Eligible)	90%
Recall (Eligible)	95%
F1-Score (Eligible)	92%

The results indicate that the Decision Tree model performs quite well, achieving 85% accuracy in classifying the data. The precision and recall values show that the model can effectively identify "Eligible" (Layak) samples with minimal misclassification, while the F1-Score of 92% reflects a good balance between precision and recall.

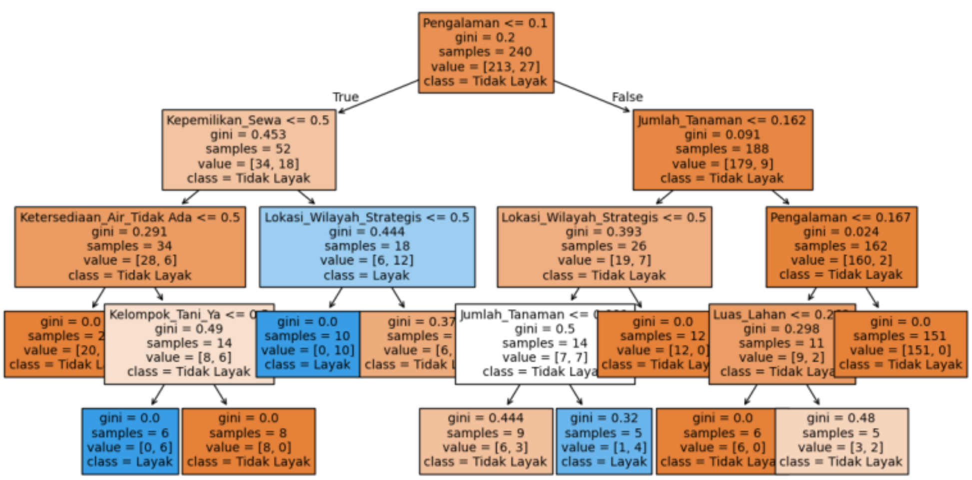

Figure 7. Decision Tree Visualization

Discussion

System Implementation in the Application

To facilitate practical implementation in the field, the system was developed as a desktop application with a Graphical User Interface (GUI) using Tkinter. This application includes several features such as a farmer data input form, automatic eligibility prediction, and recipient data management, including options to download and view data.

1. Application System Dashboard

The application dashboard serves as the main interface of the system, functioning as the area where users can input values that will be utilized by the model for classification. It includes an eligibility prediction button, which, when clicked, displays the classification results based on the data entered.

Figure 8. Application System Dashboard

2. Recipient Data Page

The recipient data page contains all community data that has been successfully classified based on eligibility status. This page presents a complete list of individuals who meet the selection criteria, making it easier to identify potential aid recipients. The displayed data not only includes basic personal information but also the values of parameters previously entered in the community data form, such as age, land availability, water availability, and education level.

In addition, this page serves as an information center for data evaluation and verification. With a structured data storage system, decision-makers can easily review the classification results and consider key factors supporting eligibility decisions. Presenting the data in tabular or visual form also helps accelerate the analysis process, ensuring that the decisions made are more objective, transparent, and accurate.

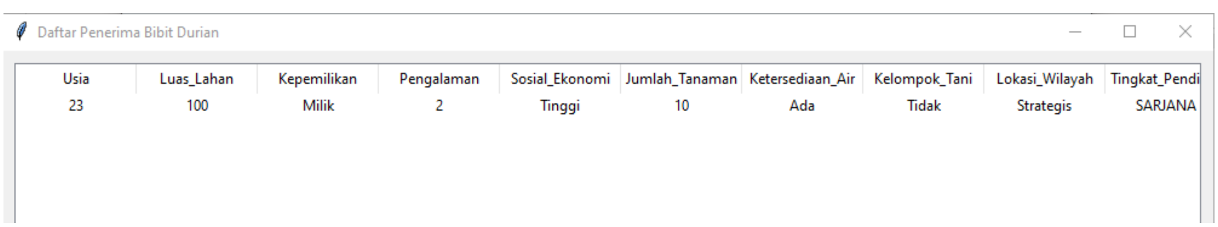

Figure 9. Recipient Data Page of Durian Seedling Distribution System

3. Statistics Page

The statistics page displays the progress of data during the model training process, showing the distribution of community data based on key variables in the form of charts. This page highlights the most influential variables identified during model training, allowing users to compare the predicted results with the actual data.

PASCAL Journal of Computer Science and Informatics
https://jurnal.devitara.or.id/index.php/komputer
E-ISSN: 3047-5074 Volume 2 Nomor 2 Tahun 2025

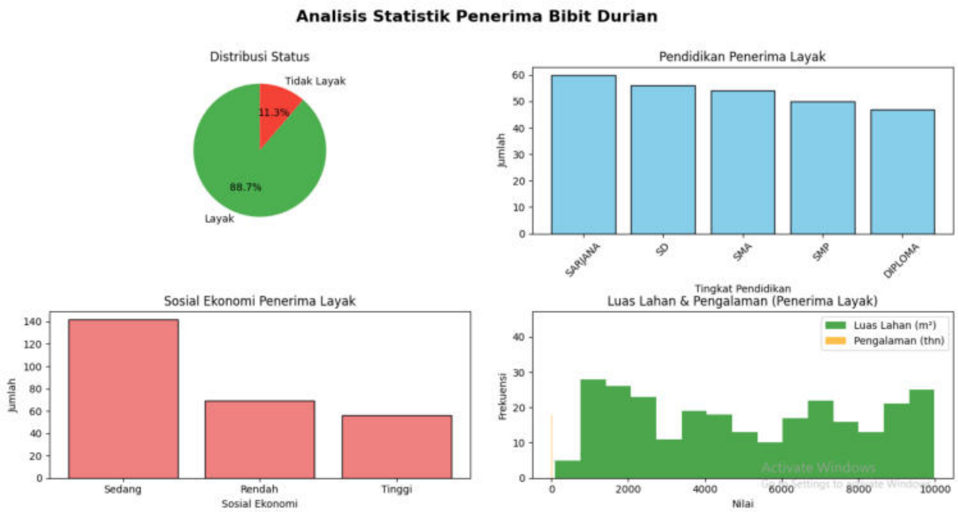

Figure 10. Statistical Visualization of Influential Variables

CONCLUSION

This study successfully developed a Decision Support System (DSS) for determining eligible recipients of superior durian seedlings using the Decision Tree algorithm, effectively classifying applicants into “Eligible” and “Not Eligible” categories based on criteria such as age, land area, land ownership, farming experience, socioeconomic conditions, number of existing plants, water availability, farmer group membership, regional location, and education level. The model achieved 85% accuracy, 90% precision, and 95% recall for the “Eligible” class, proving its effectiveness and reliability. Implemented as a desktop-based GUI application using Tkinter, the system enables easy data input, eligibility prediction, recipient management, and transparent result visualization. Future research is recommended to enhance performance by comparing other algorithms (e.g., Random Forest, Naïve Bayes, K-Nearest Neighbor), expanding the dataset for better generalization, applying resampling or SMOTE to address class imbalance, developing web or mobile-based versions for broader accessibility, and integrating the system with government or agricultural institutions to ensure more transparent, targeted, and sustainable seed distribution.

REFERENCES

Andriani, Anik. 2013. “Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa (Studi Kasus: AMIK ‘BSI Yogyakarta’).” *Seminar Nasional Teknologi Informasi dan Komunikasi* 2013(Sentika): 163–68.

Anugerah, Widiansyah. 2023. “Apa Itu Bibit Unggul? Penjelasan Lengkap Mengenai Bibit Unggul Dan Keuntungannya.” *17 september*.

Daqiqil Ibnu. 2021. *Machine Learning : Teori , Studi Kasus Dan Implementasi Menggunakan Python*. doi:10.5281/zenodo.5113507.

Dudes manulu, raegan surbakti saragih, paniel samputra sitorus. 2024. *Pemograman Python Dengan GUI Tkinter*.

Horti Indoensia. 2022. “KEBERHASILAN TANAM DURIAN DI INDONESIA BARU 30,3%.” *Horti Indoensia*. https://www.hortiindonesia.com/berita/keberhasilan-tanam-durian-di-indonesia-baru-30-3.

Karyadiputra, Erfan, and Agus Setiawan. 2023. “Sistem Pendukung Keputusan Berbasis Decision Tree Algorithm Untuk Prediksi Penyakit Diabetes.” *Media Informasi dan Teknologi* 17(2023): 294–301. https://doi.org/10.24252/teknosains.v17i3.38383.

Latifah, Retnani, Emi Setia, and Priadhana Edi. 2019. “Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn.”

Mahendra, Gede Surya. 2023. *Sistem Pendukung Keputusan : Teori Dan Penerapannya Dalam Berbagai Metode*.

Nainggolan, Lusiana, Iskandar Zulkarnain, and Sri Kusnasari. 2020. “Sistem Pendukug Keputusan Pemilihan

Jenis Bibit Unggul Pada Tanaman Durian Varietas Matahari Dengan Menggunakan Metode Weighted Aggremented Sum Product Assesment (Waspas).” *Jurnal CyberTech* 3(3): 555–64.

Nasrulah, Asmaul Husna. 2021. “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS.” 7(2): 45–51.

Nugroho, Muhammad Yogie. 2019. “SISTEM PENDUKUNG KEPUTUSAN BERBASIS WEB UNTUK DIAGNOSA PENYAKIT JANTUNG KORONER MENGGUNAKAN ALGORITMA MACHINE LEARNING HYBRID DECISION TREE DAN NAIVE BAYES.” : 1–23.

Purnamawati, Annida, Monikka Nur Winnarto, and Mely Mailasari. 2022. “ANALISIS CART (CLASSIFICATION AND REGRESSION TREES) UNTUK PREDIKSI PENGGUNA SEPEDA BERDASARKAN CUACA.” 16(1): 14–19.

Quinlan, J R. 2007. “Induction of Decision Trees.” : 81–106.

Rani, Reka Rama, Fajri Profesio Putra, and Elvi Rahmi. 2025. “Sistem Pendukung Keputusan Untuk Menentukan Bibit Sawit Unggul Menggunakan Metode Analytic Hierarchy Process (AHP).” 3: 14–27.

Susanto, Ferry, Prodi Teknik Informatika, Lampung Utara, Ade Sherly, Novita Sari, and Agus Salim. 2018. “Sistem Pendukung Keputusan Dalam Menentukan Kualitas Jambu Biji Unggulan Menggunakan Metode Weighted Product.” 01(03).

Triasanti, Dini. 2000. “Konsep Dasar Python.” : 1–6.

Weiss, Ron. 2012. “Scikit-Learn : Machine Learning in Python Scikit-Learn : Machine Learning in Python.” (January).

Widiyati, Dwi Kinasih, Masna Wati, and Herman Santoso Pakpahan. 2018. “Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah Di Kabupaten Kutai Kartanegara.” *Jurnal Rekayasa Teknologi Informasi (JURTI)* 2(2): 125. doi:10.30872/jurti.v2i2.1864