KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 1, Page: 36 - 44, March 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Comparative Study of Decision Tree and Random Forest Models for Oil Palm Productivity Prediction After Replanting

Sukardi[1]✉, Yuhandri[2], Sarjon Defit[3]
[1,2,3] Master of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia
ardysukardi03@gmail.com

Abstract

Oil palm is a strategic commodity in Indonesia that can be affected by various factors such as plant age, soil conditions, rainfall, and maintenance variations between farmers. Over time, oil palm productivity decreases, so it is necessary to predict the productivity of oil palm rejuvenation. Based on this, the purpose of this study is to apply and compare the Decision Tree and Random Forest algorithms to predict the level of oil palm productivity after rejuvenation. The prediction process was carried out at the Koperasi Unit Desa (KUD) Tirta Kencana, Kuantan Singingi Regency. The Decision Tree algorithm is a supervised prediction model, meaning it requires a training dataset whose role replaces past human experience in making decisions. The Random Forest algorithm is also able to present several decision trees used in the prediction process. The dataset in this study amounted to 241 farmer data sourced from the KUD Tirta Kencana in Kuantan Singingi Regency. The comparative results of these two methods show that both the Decision Tree and Random Forest algorithms are capable of predicting precisely and accurately. The comparative results show that the random forest method outperforms the decision tree method with an accuracy of 99%. The contribution of this research provides knowledge with the application of data mining science by comparing the performance of the decision tree and random forest algorithms in the process of plant productivity management at KUD Tirta Kencana.

Keywords: Oil Palm Productivity, Data Mining, Decision Tree, Random Forest, Productivity Prediction

1. Introduction

Technological developments in recent decades have accelerated significantly and brought fundamental changes to various sectors of life, such as communication, transportation, health, education, and industry. Advances in information technology play an important role in supporting large-scale data management and analysis processes, enabling more efficient, accurate, and evidence-based decision-making. Furthermore, technology's ability to process and integrate big data from various sources enables organisations to identify patterns, trends, and strategic information that were previously difficult to obtain through conventional approaches, thereby improving the quality of planning and decision-making in various sectors. [1] [2].

The palm oil industry in Indonesia is one of the strategic sectors that contributes significantly to the national economy and community welfare, especially in rural areas. The development of this sector is supported by government policies, investment, and the application of modern technology, but it still faces various challenges, such as unstable productivity, differences in land characteristics, and complex environmental impacts. These conditions require a data-driven and systematic analytical approach to ensure that palm oil plantation management can be carried out in a sustainable manner and adapt to environmental and climate changes. [3], A number of studies show that the use of analytical technology and data-based approaches can help the agribusiness sector understand the factors that affect productivity and formulate more effective management strategies. [2].

Data mining is one of the most widely used approaches in analysing large-scale data to discover patterns, hidden relationships, and valuable information. The data mining process is carried out through systematic stages, including data pre-processing, modelling, evaluation, and interpretation of results, so that the information produced has a high level of reliability. With the development of computing technology and analytical capabilities, data mining has been widely applied in various fields, such as business, health, finance, and the agricultural and plantation sectors, to support data-driven decision making. [4]. The application of data mining in the agricultural sector enables a more comprehensive analysis of production factors and the potential for increasing crop yields.[5]

One of the algorithms commonly used in data mining is Decision Tree, a supervised learning method that forms a decision tree structure based on data attributes. This algorithm has the advantage of easy interpretation of results because each decision can be traced through

Submitted: 17-01-2026 | Revised: 19-02-2026 | Accepted: 30-02-2026 | Published: 30-03-2026

the branches of the tree that are formed. Decision Tree is able to identify the factors that most influence a decision through a systematic process of node selection and data separation. Previous research has shown that Decision Tree algorithms, such as CART and C4.5, have been successfully used in various classification and prediction cases, including student graduation prediction and disease risk analysis, with relatively good accuracy rates. [6] [7]

In addition to Decision Trees, the Random Forest algorithm is a data mining method that develops the concept of decision trees by combining multiple trees (ensemble) to improve prediction accuracy and stability. Random Forest works by training a number of decision trees using randomly selected subsets of data and features, then combining the prediction results through a voting or averaging mechanism. This approach is effective in reducing the risk of overfitting that often occurs in single tree models and improving the generalisation capabilities of the model. Various studies show that Random Forest produces more reliable prediction performance than single algorithms, as demonstrated in studies on earthquake prediction and heart disease risk prediction.[7] [8] [9]

Based on the above description, Decision Tree and Random Forest have great potential to be applied in analysing and predicting oil palm productivity after rejuvenation, particularly at KUD Tirta Kencana in Kuantan Singingi Regency. The instability of productivity caused by differences in plant age, soil conditions, rainfall, and variations in maintenance between farmers makes it difficult to accurately predict production yields. Decision Tree offers advantages in interpretability and analysis of dominant factors that affect productivity, while Random Forest excels in improving the accuracy and stability of predictions through an ensemble approach. Therefore, the objectives and contributions of this study are to apply and compare the Decision Tree and Random Forest algorithms in predicting oil palm productivity levels after rejuvenation, thereby enriching the understanding of the effectiveness of each algorithm as a basis for strategic decision-making in data-based oil palm plantation management.

2. Methods

The research method is presented in the form of a structured flowchart to illustrate the research steps systematically. This framework aims to provide clear guidance on the process undertaken, from data collection to the evaluation of prediction results. The research method framework is shown in Figure 1. [10]

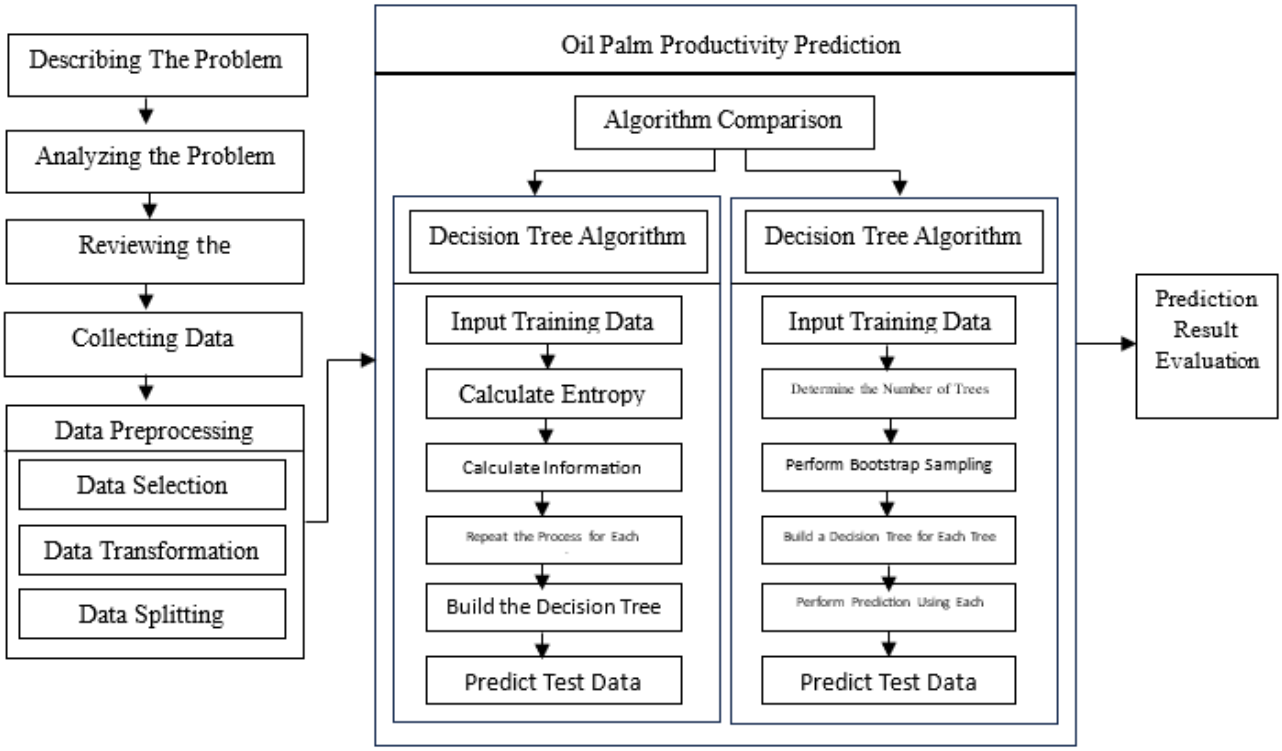

Figure 1. Research Framework

Figure 1 shows the research framework illustrating the stages the author undertook in this study. These stages were then divided into sub-stages to facilitate the author's research. The following section will explain in detail the stages the author undertook in this study.

2.1. Decision Tree Algorithm

The Decision Tree algorithm is a machine learning method used to build classification models through the systematic formation of decision tree structures. The construction process begins with calculating the entropy value in the initial dataset to measure the level of data uncertainty and determine the most optimal attribute as a separator. This stage is carried out repeatedly until a complete classification model is formed and is able to produce predictions based on the resulting decision rules. Each stage in the Decision Tree algorithm is carried out in a structured and sequential manner to ensure that the model formation process runs optimally. The calculations performed at each stage aim to determine the most informative attributes so that they can be used as the main nodes in the decision tree structure. The end result of this process is a productivity classification model with a high level of accuracy. [11] [12]

The Decision Tree algorithm stages are a model formation process that aims to break down data into a decision tree structure. The Decision Tree algorithm stages can be explained as follows : [13]

a. Calculating Entropy

Entropy is used to measure the level of uncertainty or impurity in a data set. A high entropy value indicates that the data has a high level of uncertainty, while a low value indicates more homogeneous data. The formula for calculating entropy can be written as Equation 1.

$$Entropy\ (S) = \sum_{i=1}^{n} -p_i log_2 p_i \quad (1)$$

Description :
S = Case Collection.
n = Number of S Partitions.
p_i = The probability obtained from the number of classes divided by the total number of cases.

b. Calculating Information Gain

Information Gain is used to determine which attributes are best at separating data based on target classes. The higher the gain value, the greater the ability of that attribute to reduce data uncertainty. The formula for calculating Information Gain is shown in Equation 2. [14]

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy\ (S_i) \quad (2)$$

Description :
A = Attribute.
|Si| = Number of partitions in i
|S| = Number of cases in S.

c. Repeat the Process for Each Branch

After the best attribute is selected as the root of the tree, the dataset is divided into several branches based on the value of that attribute. For each branch formed, the process of selecting the best attribute is repeated using the data subset in that branch. This step is performed recursively until all data in the branch has the same class or there are no attributes left to divide.

d. Build a Decision Tree

The results of the data division process are converted into a decision tree structure. Each node represents an attribute used for testing, while each branch shows the attribute value, and each leaf shows the result or decision class. This tree is built from the root to the leaves based on the division results in the previous steps.

e. Classification Results

Once the tree is complete, it is used to classify new data. The classification process is carried out by traversing the tree from root to leaf according to the attribute values in the test data. The final result is the decision class displayed at the leaf node as the output of the model.

2.2 Random Forest Algorithm

The Random Forest algorithm is an extension of the Decision Tree algorithm that combines a number of decision trees to improve prediction accuracy. Each tree is built independently using random subsets of data taken from the training dataset through a bootstrap sampling process. The stages in building a model using the Random Forest algorithm are explained in the image on the right.[15] [16]

The Random Forest algorithm combines the results of several Decision Trees that are trained with randomly selected data and features. Each tree in the random forest provides a prediction result and a final decision. The stages of the Random Forest algorithm can be explained as follows.

a. Determine the Number of Trees

The first step in the Random Forest algorithm is to determine how many decision trees will be built. The more trees there are, the more stable and accurate the prediction results tend to be. Sometimes, more trees will also increase computation time.

b. Create Several Bootstrap Samples

Based on the training data that has been prepared, the algorithm takes several data subsets using the bootstrap sampling method. This method means that sampling is done randomly with replacement, so that one data point can appear more than once in a subset. Each bootstrap subset will be used to build a decision tree.

Building a Decision Tree for Each Bootstrap Sample

Each bootstrap data subset is used to train a different decision tree. This process is carried out independently so that each tree has a unique structure. Based on this method, the variation between trees increases and the risk of overfitting can be reduced.

c. Continuing Tree Formation

After the best features are selected, the algorithm divides the data according to these criteria and continues the branch formation process until the tree is fully grown. This process continues until the termination condition is met, for example, when there are no more features that can be used for separation. Tree pruning is rarely performed so that each tree can grow to its maximum potential. [17]

d. Making Predictions on Each Tree

Once all trees have been formed, each tree will provide prediction results for the test data provided. In classification cases, the final result is determined by counting the majority vote from all trees. Meanwhile, for regression, the result is taken from the average prediction.

e. Generating Final Predictions

The final stage in the Random Forest algorithm is to combine the prediction results from all the trees that have been built. Each tree provides its own prediction result, and then all the results are combined to produce a final decision. This combination process is done by voting for classification or averaging for regression, so that the final result is more stable and accurate than a single decision tree. The Random Forest final prediction formula can be expressed in Equation 3. [18] [19]

$$l(y) = \arg max_c(\sum_{n=1}^{n} I h_n(y) = c) \quad (3)$$

Explanation::
$l(y)$ = Final prediction result for data y..
$h_n(y)$ = Prediction result from tree n.
I = Indicator function that has a value of 1 if the prediction $h_n(y)$
N = total number of trees in the forest.
$\arg max_c$ = selects class c that has the most votes

3. Results and Discussions

The discussion stages and results were carried out to form a structured system workflow. This process included identifying requirements and creating a system model. Each stage described the logical steps in building an efficient system. Data analysis is conducted to understand the patterns and relationships between variables in oil palm plantation productivity data. This analysis is performed to examine the interrelationships between factors that influence production outcomes. Algorithms are selected based on their ability to process data and generate accurate predictions. The analysis steps are illustrated in Figure 2.

Figure 2. Data analysis process

Figure 2 ab ove shows the system analysis process flow. This analysis was conducted using two algorithms, namely Decision Tree and Random Forest. The results of both algorithms were then compared to obtain an evaluation matrix.. [20] [21]

3.1. Preprocesing

The data is processed through a pre-processing stage to ensure it is ready for further processing. Each step in the chart illustrates the necessary data cleaning and adjustment processes. This process ensures that the data is of good quality and consistency, as shown in the following figure : [22][23]

Figure 3. Data Pre-Processing Flowchart

Oil palm plantation productivity data was collected from the Tirta Kencana Village Unit Cooperative (KUD) in Kuantan Singingi District. Productivity data was recorded in kilograms per hectare. Each farmer's harvest was calculated and recorded. Productivity status was determined based on harvest levels. Assessment attributes were selected based on variables that affect productivity. These variables are coded X1 to X5 for inputs and Y for outputs. The description of each attribute explains its function and role in the analysis, as shown in Table 1.

Table 1. Determining Assessment Attributes

No	Code	Name	Description
1	X1	Plant Age	Indicates the age of oil palm plants in years, which affects productivity levels
2	X2	Seed Type	Describes the variety or type of oil palm seeds used by planters, such as PPKS DxP Simalungun or other types
3	X3	Plant Density	Describes the number of oil palm trees per hectare planted on plantation land..
4	X4	Fertiliser Dosage	states the amount of fertiliser applied to plants in kilograms per hectare..
5	X5	Fertilisation Frequency	Indicates how many times fertilisation is carried out in a given period..
6	Y	Status	Indicates the category of plant productivity level, such as low, medium, or high.

Data transformation is carried out to convert categorical attributes into numerical values. This conversion facilitates the use of data in analysis and calculations. Each sub-attribute is assigned a conversion value according to the criteria set out in Table 2.

Tabel 2. Attribute Conversion Indicator

No	Kode	Sub Atribut	Nilai Konversi
1	X2	PPKS DxP Langkat	0
		PPKS DxP Simalungun	1
		PPKS DxP Yangambi	2
2	Y	Many	0
		Enough	1
		Less	2

The table above shows the attribute conversion indicators used. Variable X2 is converted from seed type to a numerical value. Variable Y is converted from productivity status to a numerical value. After determining the conversion indicators, the original data is converted into numerical data. This process produces a dataset that is ready for further analysis.

Next is the Data Splitting process, where the transformed data is divided into two parts for analysis purposes. 80% of the data is used as training data and 20% as test data..

3.2. Analysis / Decision Tree Method Process

To calculate entropy, the amount of data in each class of attribute Y in the training data is displayed to show the class distribution. This information is useful for calculating the entropy of the dataset. The amount of data is presented in Table 3.

Tabel 3. Number in Attribute Y

No	Class	Jumlah
1	0	49
2	1	65
3	2	79
Total		**193**

The table above shows the distribution of data in each class of attribute Y in the training data. This information is used to calculate the entropy of the dataset. The calculation is performed using Equation 2.1 as follows:

$$Y: \left(-\frac{49}{193} \times log_2 \frac{49}{193}\right) + \left(-\frac{65}{193} \times log_2 \frac{65}{193}\right) + \left(-\frac{79}{193} \times log_2 \frac{79}{193}\right) = 1{,}558398129$$

The entropy calculation for attribute Y yields a value of 1.558398129. This value is obtained from the distribution of data counts in the training data. The next step is to analyse the other attributes.

The Information Gain calculation is performed to determine which attributes are most effective in dividing the dataset. The dataset is divided based on attribute values, and each subset is analysed to see its contribution to reducing uncertainty. The calculation uses Equation 2.2 as follows:

$$X_1: 1{,}558398129 - \left(\left(\frac{49}{193} \times 0\right) + \left(\frac{65}{193} \times 0\right) + \left(\frac{79}{193} \times 0\right)\right) = 1{,}558398129$$

The Information Gain calculation reduces the entropy of the initial dataset by the weighted entropy of the subset. The same process is performed for all attributes so that each attribute has its own Information Gain value. The results of these calculations are presented in Table 4.

Tabel 4. Information Gain Calculation Results

No	Kode	Gain
1	X1	1,558398129
2	X2	0,817379407
3	X3	1,558398129
4	X4	1,558398129
5	X5	0,976145427

The decision tree formation process is carried out based on the results of information gain calculations for each attribute. Based on these calculations, attribute X3 has the highest gain value and is therefore selected as the root node. This process forms the basis for determining the initial branching in the decision tree structure. Next, the data is separated based on the attribute values that have the highest information gain. Each branch from the separation results will be analysed again to determine other attributes that affect data classification. This stage produces new branches that describe class differences based on the combination of these attribute values. The decision tree formation process and final structure can be seen in Figure 4.

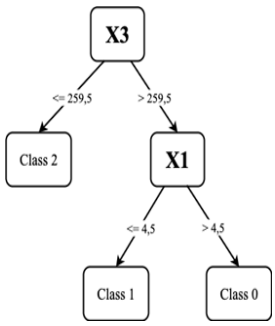

Figure 4. Formation of Tree No. 1

The image shows the visual form of a decision tree that has been constructed using the Decision Tree method. Each branch illustrates the results of attribute separation based on the threshold value obtained from the information gain calculation. Based on this structure, three decision rules (rule base) are obtained as follows:

R_1: $IF\ X_3 \leq 259{,}5\ THEN\ Class = 2$

R_2: $IF\ X_3 > 259{,}5\ AND\ X_1 \leq 4{,}5\ THEN\ Class = 1$

R_3: $IF\ X_3 > 259{,}5\ AND\ X_1 > 4{,}5\ THEN\ Class = 0$

Test Data Prediction

The test data prediction process is carried out to determine the extent to which the decision tree model is able to recognise patterns in new data that was not used during training. This test data has the same attribute structure as the training data so that it can be processed using the decision rules that have been formed. Each data point will go through tree branching based on attribute values until it reaches the predicted result class. The test data prediction results can be seen in Table 5.

Tabel 5. Test Data Prediction Results Using the Decision Tree Algorithm

No	X1	X2	X3	X4	X5	Actual	Predict
1	5	0	264	19,6	4	0	0
2	5	0	264	19,6	4	0	0
3	5	0	264	19,6	4	0	0
4	3	1	259	16	3	2	2
5	3	1	259	16	3	2	2
6	3	1	259	16	3	2	2
7	5	0	264	19,6	4	0	0
8	4	1	260	17,6	4	1	1
9	3	1	259	16	3	2	2
10	4	1	260	17,6	4	1	1

The table above shows the results of the test data prediction process using the decision tree model that was built in the previous stage. Each row in the table shows the input attribute values (X1 to X5) along with the actual class labels (Y) and the prediction results obtained. Based on the results shown, the model is able to classify data with high accuracy, where most of the prediction values match the actual classes. These results indicate that the decision tree model produced has good generalisation capabilities.

3.3. Random Forest Analysis

The Bootstrap Sampling stage is carried out through a process of random sampling from the training dataset using the bootstrap sampling method. This technique allows each tree in the Random Forest to have a different subset of data. The results of the Random Forest analysis process are presented in Table 6. [24][25]

Table 6. Prediction Results Data Using the Random Forest Algorithm

No	Class	Row	X1	X2	X3	X4	X5	Actual	Predict
1	0	1	5	0	264	19,6	4	0	0
		2	5	0	264	19,6	4	0	0
		3	5	0	264	19,6	4	0	0
		7	5	0	264	19,6	4	0	0
		18	5	0	264	19,6	4	0	0
		20	5	0	264	19,6	4	0	0
		21	5	0	264	19,6	4	0	0
		23	5	0	264	19,6	4	0	0
		31	5	0	264	19,6	4	0	0
		32	5	0	264	19,6	4	0	0
		41	5	0	264	19,6	4	0	0
		42	5	0	264	19,6	4	0	0
2	1	8	4	1	260	17,6	4	1	1
		10	4	1	260	17,6	4	1	1
		12	4	1	260	17,6	4	1	1
		13	4	1	260	17,6	4	1	1
		15	4	1	260	17,6	4	1	1
		35	4	1	260	17,6	4	1	1
		38	4	1	260	17,6	4	1	1
		43	4	1	260	17,6	4	1	1
		44	4	1	260	17,6	4	1	1
		47	4	1	260	17,6	4	1	1
3	2	4	3	1	259	16	3	2	2
		5	3	1	259	16	3	2	2
		6	3	1	259	16	3	2	2
		9	3	1	259	16	3	2	2
		11	3	1	259	16	3	2	2
		14	3	1	259	16	3	2	2
		16	3	1	259	16	3	2	2
		17	3	1	259	16	3	2	2
		19	3	1	259	16	3	2	2
		22	3	1	259	16	3	2	2
		24	3	1	259	16	3	2	2
		25	3	1	259	16	3	2	2
		26	3	1	259	16	3	2	2
		27	3	1	259	16	3	2	2
		28	3	1	259	16	3	2	2
		29	3	1	259	16	3	2	2
		30	3	1	259	16	3	2	2
		33	3	1	259	16	3	2	2
		34	3	1	259	16	3	2	2

36	3	1	259	16	3	2	2
37	3	1	259	16	3	2	2
39	3	1	259	16	3	2	2
40	3	1	259	16	3	2	2
45	3	1	259	16	3	2	2
46	3	1	259	16	3	2	2
48	3	1	259	16	3	2	2

The table above shows that the Random Forest model also produces perfect predictions with an error rate of 0%. Looking at the performance of both models in a more structured way. The evaluation matrix (Confusion Matrix) is shown in Table 7..

Table 7. Confusion Matrix

	CLASS	Prediction: Many (0)	Sufficient Prediction (1)	Under prediction (2)
Algoritma Decision Tree	Current Many (0)	12	0	0
	Current Sufficient (1)	0	10	0
	Current Insufficient (2)	0	0	26
	CLASS	**Prediction: Many (0)**	**Sufficient Prediction (1)**	**Under prediction (2)**
Algoritma Random Forest	Current Many (0)	12	0	0
	Current Sufficient (1)	0	10	0
	Current Insufficient (2)	0	0	26

The table above shows that both the Decision Tree and Random Forest algorithms are able to classify all data correctly without classification errors. This can be seen from the values in each class, which are all on the main diagonal of the matrix, indicating that the prediction results are the same as the actual data. Both algorithms have the accuracy levels presented in Table 8.

Table 8. Model Performance Results

No	Algorithm	Accuracy	Percision	Recall	F1-Score	Avg	Presentase
1	Decision Tree	1	1	1	1	1	99%
2	Random Forest	1	1	1	1	1	99%

The table above shows that both the Decision Tree and Random Forest algorithms perform very well, with accuracy, precision, recall, and F1-score values reaching 99%. This indicates that the model is capable of classifying all data correctly. Based on this, both algorithms have the same high level of accuracy and reliability.

4. Conclusions

This study shows that the Decision Tree and Random Forest algorithms are capable of accurately predicting oil palm productivity after rejuvenation at KUD Tirta Kencana in Kuantan Singingi Regency. The performance comparison results show that the Random Forest algorithm has superior performance compared to Decision Tree with an accuracy rate of 99%. Therefore, Random Forest is recommended as a more optimal method in supporting data-based decision making for managing oil palm productivity after rejuvenation..

References

[1] A. Fricticarani, A. Hayati, I. Hoirunisa, and G. M. Rosdalina, "Strategi pendidikan untuk sukses di era teknologi 5.0," *J. Inov. Pendidik. Dan Teknol. Inf.*, vol. 4, no. 1, pp. 56–68, 2023.

[2] V. Maria, S. D. Rizky, and A. M. Akram, "Mengamati perkembangan teknologi dan bisnis digital dalam transisi menuju era industri 5.0," *Wawasan J. Ilmu Manajemen, Ekon. Dan Kewirausahaan*, vol. 2, no. 3, pp. 175–187, 2024.

[3] M. Hudori, "Perkembangan Industri Kelapa Sawit Indonesia sejak Era Kemerdekaan hingga Saat Ini," *J. Citra Widya Edukasi*, vol. 16, no. 3, pp. 227–240, 2024.

[4] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024.

[5] M. Rizaludin, "Prediksi Prediksi Perilaku Pelanggan Pada Produk UMKM Batik Dengan Menggunakan Algoritma Decision Tree," *Teknomatika*, vol. 13, no. 02, pp. 8–16, 2023.

[6] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024.

[7] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan algoritma decision tree dalam klasifikasi data prediksi kelulusan mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023.

[8] "Prediksi Penyakit Hipertensi Menggunakan Metode Decison Tree dan Random Forest," *J. Ilm. Komputasi*, vol. 23, no. 1, Mar. 2024, doi: 10.32409/jikstik.23.1.3503.

[9] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023.

[10] P. R. Rosmila, R. Risqiati, and A. S. Darmawan, "KOMPARASI ALGORITMA NAÏVE BAYES DAN DECISION TREE UNTUK MENGANALISIS KEPUASAN MASYARAKAT DPMPTSP KABUPATEN

BATANG," *Inf. Syst. J.*, vol. 8, no. 02, pp. 109–118, 2025.
[11] Y. L. Fatma and N. Rochmawati, "Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4. 5," *J. Informatics Comput. Sci.*, vol. 5, no. 04, pp. 486–493, 2024.
[12] Y. Gunawan, A. S. Firmansyah, A. Mubarok, B. Subaeki, K. Manaf, and S. W. Pitara, "Implementasi Algoritma Decision Tree untuk Klasifikasi Kemampuan Membaca Al-Qur'an dengan Metode Tahsin," *J. Informatics*, vol. 10, no. 1, 2025.
[13] A. R. Raharja, A. Pramudianto, and Y. Muchsam, "Penerapan algoritma decision tree dalam klasifikasi data 'Framingham'untuk menunjukkan risiko seseorang terkena penyakit jantung dalam 10 tahun mendatang," *Technol. J*, vol. 1, no. 1, 2024.
[14] I. N. Sari, L. F. Lidimilah, M. Kom, A. Lutfi, and M. Kom, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naive Bayes dalam Klasifikasi Gangguan Tidur," in *Proceeding National Conference of Research and Community Service Sisi Indonesia*, 2025, pp. 192–209.
[15] E. Helmud, F. Fitriyani, and P. Romadiana, "Classification comparison performance of supervised machine learning random forest and decision tree algorithms using confusion matrix," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 92–97, 2024.
[16] R. Harahap, M. Irpan, M. A. Dinata, and L. Efrizoni, "Perbandingan Algoritma Random Forest Dan Xgboost Untuk Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien," *BETRIK*, vol. 15, no. 02, pp. 130–141, 2024.
[17] A. P. Siregar, D. P. Purba, J. P. Pasaribu, and K. R. Bakara, "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, pp. 155–164, 2023.
[18] H. Santoso, R. A. Putri, and S. Sahbandi, "Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest," *J. Manaj. Inform.*, vol. 13, no. 1, pp. 62–72, 2023.
[19] X. Fu, Y. Chen, J. Yan, Y. Chen, and F. Xu, "BGRF: A Broad Granular Random Forest Algorithm," *J. Intell. Fuzzy Syst.*, vol. 44, no. 5, pp. 8103–8117, 2023, doi: 10.3233/jifs-223960.
[20] A. Khoirunnisa and N. G. Ramadhan, "Improving malaria prediction with ensemble learning and robust scaler: An integrated approach for enhanced accuracy," *J. Infotel*, vol. 15, no. 4, pp. 326–334, 2023.
[21] N. Saputra, R. Antoni, A. Widodo, E. M. Solissa, and I. Arief, "Improving foreign language proficiency in society by decision tree classification," in *AIP Conference Proceedings*, AIP Publishing LLC, 2024, p. 110005.
[22] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, "Meningkatkan model prediksi kelulusan santri tahfidz di Pondok Pesantren Al-Kautsar menggunakan algoritma Random Forest," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, 2025.
[23] A. Ristyawan, A. Nugroho, and T. K. Amarya, "Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke," *JATISI*, vol. 12, no. 1, 2025.
[24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[25] E. R. B. Sebayang, Y. H. Chrisnanto, and M. Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG (International J. Eng. Econ. Soc. Polit. Gov.*, vol. 1, no. 3, pp. 237–253, 2023.

Biographies of Authors

Sukardi, A.Md, S.Kom was born in Magelang on 15 November 1982. He completed his Diploma III (D3) in Information Systems at STT Unggulan Swarnadwipa, Riau, and obtained his Associate Degree in Computer Science (A.Md) in 2006. He then continued his studies in Information Technology at Universitas Islam Kuantan Singingi (UNIKS), Riau, and graduated with a Bachelor of Computer Science (S.Kom) in 2016. He is currently pursuing a Master of Information Technology (M.Kom) at Universitas Putra Indonesia YPTK Padang, which he began in 2024. In his professional career, he serves as an Information Technology Specialist at KUD Tirta Kencana and is actively involved in developing educational institutions in the Kuantan Singingi Regency. He can be contacted via email at ardysukardi03@gmail.com.

Prof. Dr. Yuhandi, S.Kom.,M. Kom was born in Tanjung Alam on May 15, 1973. He is an assistant professor in Faculty of Computer Science, Universitas Putra Indonesia YPTK. He received the bachelor's degree in informatics management and master's degree in information technology in 1992 and 2006 from Universitas Putra Indonesia YPTK. Moreover, he completed his Doctor of Information Technology as informatics medical image expertise from Gunadarma University in April 2017. He is a lecturer at the Faculty of Computer Science, Universitas Putra Indonesia YPTK. He can be contacted at email: yuyu@upiyptk.ac.id.

Prof. Dr. H. Sarjon Defit, S.Kom., M. Sc

Was bornPadangSibusuk / 07August 1970. He is the Chancellor of PutraIndonesia University YPTK Padang. Currently active as a lecturer in Computer Science. The educational history of Siatthe College of Informatics and Computer Management (STMIK "YPTK" Padang) with a graduation in 1993. An educational history of S2 at Universiti Teknologi Malaysia, Johor Bahru, graduated in 1998. Then a Doctoral Education History at Universiti Teknologi Malaysia, Johor Bahru, graduated in 2003. The field of science consists ofdatamining, artificia lintelligence, decisionsupportsystems, and others. He can be contacted at email: sarjon_defit@upiyptk.ac.id