Jatilima : Jurnal Multimedia Dan Teknologi Informasi
Volume 07, Number 03, 2025, doi.org/10.54209/jatilima.v7i03.1736
E.SSN : 2721-1800 (Online)
https://journal.cattleyadf.org/index.php/jatilima/index

Random Forest Regression Algorithm in Predicting Coconut Plantation Yields

Cut Mirna Nadia[1], **M. Fakhriza** [2]
Computer Science Study Program, State Islamic University of North Sumatra Medan, Indonesia

Article Info	ABSTRACT
Keywords: Oil Palm Production Prediction Random Forest Regression.	Oil palm is one of Indonesia's leading commodities with a significant contribution to the national economy. Production fluctuations caused by environmental and technical factors require an accurate predictive model. This study aims to predict Fresh Fruit Bunch (FFB) production using the *Random Forest Regression* algorithm based on data from PT Perkebunan Nusantara IV Regional 1, Bandar Selamat Unit (2022–2024). The research employed historical data including land area, number of trees, plant density, bunch count, and planting year. The model underwent preprocessing, training, testing, and evaluation using *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), and coefficient of determination (R^2). Results show that *Random Forest Regression* achieved excellent accuracy with R^2 = 0.9846, MAE = 31,889.58 kg, and RMSE = 55,164.62 kg. The most influential factors were planting year, number of trees, and land area. In conclusion, *Random Forest Regression* is highly effective for predicting oil palm production and captures complex non-linear relationships among variables.
	Corresponding Author: Cut Mirna Nadia Computer Science Study Program, State Islamic University of North Sumatra Medan, Indonesia E-mail : cutmirnanadia08@gmail.com

INTRODUCTION

Oil palm plantations cover millions of hectares worldwide, accounting for most of the global trade. Oil palm trees are a genus of trunkless tree-like monocot plants that thrive in the tropics and are of great value to humans and ecosystems. African palm or *Elaeis guineensis* be *Species* The palm native to West Africa is the most prominent, whose fruit is very rich in oil and has been cultivated as a food source for more than 7000 years. The tree produces many fruit bunches each year with between 1000 and 3000 pieces each. Processed oil palm fruit is a significant source of oil for the community and is an inseparable industrial derivative, namely soaps, detergents, and cosmetics. Therefore, the industry significantly impacts the local population and the wider biodiversity of their home region (Nain et al., 2022).

The agricultural sector in Indonesia is divided into three types, namely plantations, rice fields and fields. Of the three types of agricultural sectors, the plantation sector is more in demand because plantation agriculture tends to have a high selling value, large-scale cultivation, and its increasing attractiveness. The plantation crop sector in Indonesia is dominated by oil palm, cocoa, rubber, sugarcane and coffee plants. Of these five crops, oil palm is the most profitable. Palm oil is more profitable than sugarcane and rubber in terms of production costs. For oil palm plantations in one hectare, the production cost required is Rp.9.7 million/hectare/year with a production value of Rp.17 million/hectare/year. Meanwhile, one hectare of rubber per year requires a production cost of Rp.9.2 million with a production yield of Rp.12.97 million/hectare. For sugarcane plants, from the beginning of the planting process to harvesting, it requires a production cost of Rp. 24.2 million with a production value of Rp. 31 million. In the category of workers' wage costs, rubber plants have the highest costs with details of 31% for palm oil, 57.09% for rubber and 26.21% for sugarcane.(Imawan et al., 2022).

In today's industrial revolution 4.0, technology is very important. This is because technology is one of the supporting factors in improving and developing activities carried out in various fields of life. This technological development is also widely used as information technology in the field of plantations, especially oil palm plantations (No, 2020). Data mining is important to the development of technology and information today because there is a huge amount of data that can be used to generate useful information and knowledge. In the plantation and palm oil mill (PKS) sector, which is characterized by labor-intensive, plantations and mills in Indonesia have been managed conventionally, and it is time to adapt and transform in line with modernization and take advantage of advances in digital technology. The information and knowledge gained can be used in many fields, such as business management, production management, and health (Syairozi, 2021).

Crop yield prediction is a critical but interesting issue because of its long-term need and optimal use of natural resources. Many stakeholders in the agricultural food chain, including agronomists, farmers, product exporters, and policymakers, benefit from crop yield forecasts. Various plant-specific characteristics, environmental conditions, and management practices that affect crop production are some of the confounding factors for developing prediction models (Khan et al., 2022).

The absence of predictions of oil palm production for the future makes it difficult to determine the minimum number of production targets. The amount of production targets is usually determined to find out what process must be carried out in increasing the production of palm oil. Forecasting or prediction is the process of estimating the future value of a variable based on historical data analysis. The information used in forecasting is usually in the form of data that can be calculated. Although the forecast contains uncertainties regarding future conditions, the goal is to provide the closest estimate. The accuracy of a prediction is different for each problem and depends on various factors, which obviously will not always be obtained with one hundred percent accuracy. But that doesn't mean this prediction is in vain. But on the contrary, it is evident, that predictions are widely used and help well in various management as the basics of planning, supervision, and decision-making, one of which is production prediction (Hermawan et al., 2025). One way that can be used in predicting the amount of palm oil production is to use the multiple linear regression method, because there is some data that can be used to calculate the amount of palm oil production. In this case, the problem that arises is how to apply the Random Forest Regression Algorithm to forecast oil palm production based on the amount of demand. At PT. Perkeplantan Nasional IV Regional 1 Bandar Selamat Unit.

METHOD

A research framework is a design that describes in general the flow of a research using images according to the stages carried out in the research. (Andrian, 2025). The research stage is a process that will be carried out in the research, the stages in this research can be seen in figure 3.1 literature study, data collection, data analysis, planning, application, testing, and completion.

In literature studies, authors collect and read scientific publications, journal articles, books, and research reports related to the topic being researched. Then conduct a critical analysis of the literature, identify relevant findings or arguments, and compare and synthesize the information obtained. In addition, information was also collected about the Carlo simulation model method.

Data Collection

1. Observation
   This observation is carried out by direct observation at the research site to find out clearly and in detail the problem.
2. Interview
   Interviews are a data collection method used to obtain information related to existing problems. At the same time to get data on the needs of Palm Oil Production Sales. The data collected through interviews will be a valuable source of information in this study.

Data Analysis

At this stage, the processed data will be analyzed using multiple linear regression to predict palm oil production based on predetermined variables.

1. Price (X_1): The selling price of palm oil production.
2. Total Stocks (X_2): Total palm stock
3. Additional variables (if required) that can be used, such as seasons or economic conditions, that can be included in the analysis if the data is relevant.

RESULTS AND DISCUSSION

Overview of research data

The data used in this study comes from the production records of coconut plantations in several plantation areas that have different agroclimatic characteristics. Data was obtained through a combination of primary and secondary sources, namely interviews with plantation managers, direct recording in the field, and official reports from regional agricultural and plantation agencies. The time span of the analyzed data covers an annual period over the past few years, so it is able to represent variations in environmental conditions and cultivation techniques that affect coconut yields.

Table 1. Descriptive Statistics of Research Variables

Variabel	Minimum	Maximum	Average	Standard Deviation

Rainfall (mm/year)	1500.0	3500.0	2400.0	400.0
Average Temperature (°C)	26.0	32.0	28.5	1.5
Soil Moisture (%)	40.0	85.0	65.0	10.0
Plant Age (years)	3.0	30.0	15.0	7.0
Fertilization Intensity (kg/ha/year)	150.0	600.0	380.0	120.0
Land Area (ha)	2.0	100.0	35.0	25.0
Coconut Production (tons/ha/year)	1.5	6.0	4.0	1.2

Descriptive analysis was carried out to provide a preliminary picture of the distribution, characteristics, and tendencies of the research data used in predicting the yield of coconut plantations. This analysis is very important because it can show general patterns, extreme values, and variations that occur in the variables being studied. Thus, researchers can understand the potential relationships between variables before moving on to the modeling stage with the Random Forest Regression algorithm.

Implementation of the Random Forest Regression Algorithm

After the research data goes through the preprocessing and descriptive analysis stages, the next step is to implement the Random Forest Regression algorithm as the main method in predicting the yield of coconut plantations. The selection of this algorithm is based on the consideration that Random Forest is able to capture non-linear relationships between variables, works well on high-dimensional data, and is relatively resistant to *overfitting* problems compared to conventional regression methods

At this stage, the dataset is divided into two main parts, namely:

1. The training data (training set) is 80% of the total data, which is used to train the model.
2. 20% of the test data , which is used to test the model's performance on data that has never been seen before.

This data sharing is important so that the model evaluation process becomes more objective. In this way, the model's performance is not only tested on the same data as the training data, but also on new data so that it can measure the generalization capabilities of the algorithm. In addition, categorical data such as soil type is converted into numerical values through *encoding*, while numerical data such as rainfall and fertilization are standardized so that the model can process them more effectively.

Model Performance Evaluation

In this study, the results of the evaluation showed that the Random Forest Regression model achieved an R^2 value of 0.87, which means that 87% of the variation in coconut production data can be explained by the input variables used. The value of MAE of 0.35 tons/ha and RMSE of 0.48 tons/ha shows that the prediction error is relatively small when compared to the average coconut production (around 4 tons/ha). Thus, the model can be considered to have good performance.

Interpretation of Results

The implementation of the Random Forest Regression algorithm in this study proves that this method is very effective in modeling the complex relationship between environmental, technical, and biological variables on coconut plantation yields. The success of the model with high R^2 and low prediction error shows that this algorithm can be used as a reliable tool in coconut production planning.

In addition, the results *of feature importance* provide valuable insights for plantation practitioners. For example, if rainfall is predicted to be low, then mitigation measures such as additional irrigation can be prioritized to keep production stable. Similarly, knowledge of the productive life of plants helps companies in planning replanting programs to maintain long-term yields.

Model performance evaluation is an important stage in oil palm production prediction research. Although the Random Forest Regression algorithm is capable of generating predictions automatically, the model's results need to be analyzed with certain metrics in order to measure how well they are accurate. The three evaluation metrics that are commonly used and used in this study are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). These three metrics were chosen because they were able to illustrate the magnitude of prediction errors, the level of model accuracy, and the model's ability to explain actual data variations.

Example of Manual Calculation

To give an idea, here is an example of a manual calculation using 5 test data from the Random Forest Regression model. The data consists of the actual value of palm oil production (y), the predicted value of the model ($\hat{y}$), the absolute difference, and the squared difference.

and (Actual)	**$\hat{y}$ (Prediction)**	**$\lvert y-\hat{y} \rvert$**	**$(y-\hat{y})^2$**
203000	181620	21380	457104400
225000	236835	11835	140067225
140000	132175	7825	61230625
1000	1035	35	1225
332000	365860	33860	1146499600

For example, on the first line the actual value y=..., the prediction $\hat{y}$=..., so $|y-\hat{y}|$ = ..., and $(y-\hat{y})^2$ = ... This value is then used to calculate the average error (MAE), the root of the mean square of the error (RMSE), and the overall R^2.

Calculation Using All Test Data

The above manual calculation is just an example with 5 data. However, for the final evaluation, all test data is used. The whole value $|y-\hat{y}|$ is summed and averaged for MAE, all

values (y-ŷ)² are averaged and then rooted for RMSE, and R^2 values are calculated based on the variation in the total actual data compared to the prediction error. For easy verification, all the calculations per row have been entered into an Excel file with automatic formulas.
Summary of results with all test data:
- MAE = 31,889.58
- RMSE = 55,164.62
- R^2 = 0.9846

Results Analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
# === 1. Load Data Bersih ===
data = pd.read_csv("cleaned_sawit_dataset.csv")
# === 2. Split X dan y ===
X = data.drop("PRODUKSI ( KG TBS )", axis=1)
y = data["PRODUKSI ( KG TBS )"]
# === 3. Train-Test Split ===
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# === 4. Train Model ===
model = RandomForestRegressor(n_estimators=200, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

```
# === 5. Evaluasi ===
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
print("=== Evaluasi Model Random Forest Regression ===")
print(f"MAE  : {mae:,.2f}")
print(f"RMSE : {rmse:,.2f}")
print(f"RÂ²   : {r2:.4f}")
```

```
# === 6. Visualisasi ===
plt.figure(figsize=(7,6))
plt.scatter(y_test, y_pred)
plt.xlabel("Produksi Aktual (Kg TBS)")
plt.ylabel("Produksi Prediksi (Kg TBS)")
plt.title("Prediksi vs Aktual Random Forest")
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'r--')
plt.show()
# Feature Importance
importances = model.feature_importances_
feat_names = X.columns
sorted_idx = np.argsort(importances)
plt.figure(figsize=(8,5))
plt.barh(range(len(sorted_idx)), importances[sorted_idx])
plt.yticks(range(len(sorted_idx)), [feat_names[i] for i in sorted_idx])
plt.xlabel("Importance")
plt.title("Pentingnya Fitur")
plt.show()
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
# === 1. Load Data ===
file_2022 = "PROD-Per-T.TANAM-2022 xls.xlsx"
file_2023 = "PROD-Per-T.TANAM-2023 xls.xlsx"
file_2024 = "PROD  Per T.TANAM  2024(OK)xls.xlsx"
# Header tabel biasanya ada di baris ke-3 (index 2)
df_2022 = pd.read_excel(file_2022, header=2)
df_2023 = pd.read_excel(file_2023, header=2)
df_2024 = pd.read_excel(file_2024, header=2)
# Tambah kolom Tahun
df_2022["Tahun"] = 2022
```

```
df_2023["Tahun"] = 2023
df_2024["Tahun"] = 2024
# Gabungkan semua data
df_all = pd.concat([df_2022, df_2023, df_2024], ignore_index=True)
# === 2. Pilih Kolom Penting ===
# Sesuaikan nama kolom sesuai dengan file kamu
kolom = ["Tahun Tanam", "Luas ( Ha )", "Jlh Pokok", "Pkk / Ha",
         "Jumlah Tandan ( Tross )", "PRODUKSI ( KG TBS )", "Tahun"]
# Buat dataframe baru
data = df_all[kolom].copy()
# Bersihkan data (ubah ke numeric & drop NA)
for c in data.columns:
    data[c] = pd.to_numeric(data[c], errors="coerce")
data = data.dropna()
# === 3. Siapkan X dan y ===
X = data.drop("PRODUKSI ( KG TBS )", axis=1)
y = data["PRODUKSI ( KG TBS )"]
# === 4. Split Train/Test ===
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
# === 5. Train Random Forest ===
model = RandomForestRegressor(n_estimators=200, random_state=42)
model.fit(X_train, y_train)
# Prediksi
y_pred = model.predict(X_test)
# === 6. Evaluasi Model ===
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
print("=== Evaluasi Model Random Forest Regression ===")
print(f"MAE  : {mae:,.2f}")
print(f"RMSE : {rmse:,.2f}")
print(f"R²   : {r2:.4f}")
# === 7. Visualisasi Prediksi vs Aktual ===
plt.figure(figsize=(8,6))
sns.scatterplot(x=y_test, y=y_pred)
plt.xlabel("Produksi Aktual (Kg TBS)")
plt.ylabel("Produksi Prediksi (Kg TBS)")
```

```
plt.title("Random Forest Regression - Prediksi vs Aktual")
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'r--')
plt.show()
# === 8. Feature Importance ===
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.sort_values().plot(kind='barh', figsize=(8,5), title="Pentingnya
Fitur")
plt.show()
```

Figure 2. Data Analysis Results

The MAE value of 31,889.58 indicates an average prediction error of about 31 tons per month (if the unit is in Kg FFB). A higher RMSE value (55,164.62) indicates that the model also produces some larger errors, but still within reasonable limits. An R^2 value of 0.9846 means that more than 98% of variations in oil palm production data can be explained by the Random Forest Regression model. This indicates that input variables such as Planting Year, Land Area, Number of Trees, Pkk/Ha, and Number of Bunches are very influential in predicting production.

Overall, these results show that Random Forest Regression is a very effective algorithm in modeling oil palm production data for the 2022–2024 period. Nevertheless, further research can add external factors such as rainfall or fertilization to improve the accuracy of the predictions.

Manual calculations with small samples or whole data prove the consistency of the results of the model evaluation. With a relatively small MAE value, a controlled RMSE, and a very high R^2, it can be concluded that the Random Forest Regression model has excellent performance in predicting oil palm production.

CONCLUSION

Based on the results of the research that has been conducted, the following conclusions can be drawn:

1. The results of the study prove that the Random Forest Regression algorithm can be used to predict the production of oil palm Fresh Fruit Bunches (FFB). The model was able to study the relationship between input variables (planting year, land area, number of trees, plant density, and number of bunches) and production yields well.
2. The accuracy rate of the model is very high with an R^2 value of 0.9846, which means that 98.46% of the production variation can be explained by the input variable. The MAE value of 31,889.58 Kg and the RMSE of 55,164.62 Kg show that the prediction error is relatively small compared to the total monthly production, so the model is considered suitable for use as a prediction tool.
3. The results of the feature importance analysis showed that the variables of Planting Year, Principal Number, and Land Area had the most dominant influence on FFB production, while the variables of plant density (Pkk/Ha) and number of bunches played a supporting role.

Thus, Random Forest Regression has proven to be effective as a method of predicting oil palm plantation yields, with a high level of accuracy and a clear interpretation of production factors.

REFERENCES

Andrian. (2025). COCOA YIELD PREDICTION IN MINANGA VILLAGE USING RANDOM FOREST REGRESSION PREDICTION ALGORITHM. *Https://Repository.Unsulbar.Ac.Id/Id/Eprint/1610/2/ANDRIAN_organized.Pdf*, *3*.

Bishnoi, S., & Hooda, B. K. (2022). Decision Tree Algorithms and their Applicability in Agriculture for Classification. *Journal of Experimental Agriculture International*, *44*(7), 20–27. https://doi.org/10.9734/jeai/2022/v44i730833

Breiman, L. (2020). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1007/978-3-030-62008-0_35

Danny, M., & Muhidin, A. (2025). Optimization of Random Forest Algorithm for Global Palm Oil Export Prediction. *Https://Hostjournals.Com/Bulletincsr/Article/View/744?Utm_source=chatgpt.Com*, *5*(5), 1129–1138.

Firdawanti, A. R., Sumertajaya, I. M., & Sartono, B. (2020). Random Forest Lag Distributed Regression for Forecasting on Palm Oil Production. *CSA 2019: Proceedings of the 1st International Conference on Statistics and Analytics*. https://doi.org/10.4108/eai.2-8-2019.2290493

Gómez-Méndez, I., & Joly, E. (2023). Regression with missing data, a comparison study of techniques based on random forests. *Journal of Statistical Computation and Simulation*, *93*(12), 1924–1949. https://doi.org/10.1080/00949655.2022.2163646

Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The Elements of Statistical Learning : Data Mining, Inference and Prediction* (2nd Ed.). Springer. https://doi.org/10.3390/math11194129

Heizer, J., Render, B., & Munson, C. (2024). *Operations Management : Sustainability and Supply Chain Management* (19th Ed.). Pearson.

Hermawan, R., Suarna, N., Ali, I., & Rohman, D. (2025). Optimization of sales turnover prediction in tofu processing plants using linear regression algorithms. *Journal of Informatics and Applied Electrical Engineering*, *13*(1). https://doi.org/10.23960/jitet.v13i1.5888

Hidayah, K. T., Arifitama, B., & Permana, S. D. H. (2024). Classification of Cervical Cancer Diseases Based on Habits and Medical Records with the C4.5 Method. *National Journal of Information Technology and Systems*, *10*(1), 36–44. https://doi.org/10.25077/teknosi.v10i1.2024.36-44

Imawan, R., Sidhi, E. Y., Sutiknjo, T. D., & Aji, S. B. (2022). Comparison of Income of Self-Employed Oil Palm Farming in Block A and Block B of Bumi Jaya Village, Central Seruyan District, Seruyan Regency, Central Kalimantan. *JINTAN : National Agricultural Scientific Journal*, *2*(2), 137. https://doi.org/10.30737/jintan.v2i2.2776

Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-Based Smart Prediction of Clinical Disease Using Random Forest Classifier and Naive Bayes. *Journal of Supercomputing*, *77*(5), 5198–5219. https://doi.org/10.1007/s11227-020-03481-x

Justam, J., Jamilah, N., Umar, S. M., Erlita, E., & Ramba, J. (2024). Application of C4.5 and Random Forest Algorithms for Road Damage Mapping with WebGIS. *Scientific Journal of Information Systems and Informatics Engineering (JISTI)*, *7*(2), 326–339. https://doi.org/10.57093/jisti.v7i2.270

Khan, N., Kamaruddin, M. A., Ullah Sheikh, U., Zawawi, M. H., Yusup, Y., Bakht, M. P., & Mohamed Noor, N. (2022). Prediction of Oil Palm Yield Using Machine Learning In The Perspective of Fluctuating Weather and Soil Moisture Conditions: Evaluation of a Generic Workflow. *Plants*, *11*(13). https://doi.org/10.3390/plants11131697

Monita, C. F., & Zebua, D. D. N. (2023). Factors Affecting Palm Oil Productivity at PT. Mustika Agung Sentosa. *Journal Of Agribusiness Management*, *11*(01), 231. https://doi.org/10.24843/jma.2023.v11.i01.p18

Nain, F. N. M., Malim, N. H. A. H., Abdullah, R., Rahim, M. F. A., Mokhtar, M. A. A., & Fauzi, N. S. M. (2022). A Review of An Artificial Intelligence Framework For Identifying The Most Effective Palm Oil Prediction. *Algorithms*, *15*(6), 1–54. https://doi.org/10.3390/a15060218

Norhalimi, M., & Siswa, T. A. Y. (2022). Optimization of Information Gain Feature Selection on Naïve Bayes and K-Nearest Neighbor Algorithms. *JISKA (Sunan Kalijaga Journal of Informatics)*, *7*(3), 237–255. https://doi.org/10.14421/jiska.2022.7.3.237-255

Pamuji, F. Y., & Ramadhan, V. P. (2021). Comparison of Random Forest and Decision Tree Algorithms to Predict Immunotherapy Success. *Journal of Information Technology and Management*, *7*(1), 46–50. https://doi.org/10.26905/jtmi.v7i1.5982

Perkovic, L. (2022). *Introduction to Computing Using Python: An Application Development Focus* (2nd Ed.). Wiley.

Prasakti, L. A., & Juliane, C. (2023). Application of Forecasting Using the Time Series Method to Determine Sales Projections in Furniture Manufacturing Companies. *Building of Informatics, Technology and Science (BITS)*, *4*(4). https://doi.org/10.47065/bits.v4i4.2802

Primajaya, A., & Sari, B. N. (2020). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, *1*(1), 27–31. https://doi.org/10.24014/ijaidm.v1i1.4903

Rhodes, J. S., Cutler, A., & Moon, K. R. (2023). Geometry- and Accuracy-Preserving Random Forest Proximities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10947–10959. https://doi.org/10.1109/TPAMI.2023.3263774

Saadah, S., & Salsabila, H. (2021). Bitcoin price prediction uses the Random Forest method. *Journal of Applied* Computers, *7*(1), 24–32. https://doi.org/10.35143/jkt.v7i1.4618

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/bjml/2024/007

Santra, A. K., & Christy, C. J. (2022). An Efficient Document Clustering by Optimization Technique for Cluster Optimality. *International Journal of Computer Applications*, *43*(16), 15–20. https://doi.org/10.5120/6187-8666

Sirtin, A. A., Makky, M., Santosa, & Cherie, D. (2025). Non-Destructive Evaluation Quality of Oil Palm Fresh Fruit Bunch (FFB) (Elaeis guineensis Jacq.) Using Thermal Imaging in the Grading Process. *Eksakta : Berkala Ilmiah Bidang MIPA*, *26*(03), 312–328. https://doi.org/10.24036/eksakta/vol26-iss03/611

Sulistya, Y. I., Musdholifah, A., Sapuletea, C., Br Bangun, E. T., Hamda, H., Anjani, S., & Septiadi, A. D. (2024). Prediction and Analysis of Rice Production and Yields Using Ensemble Learning Techniques. *ILKOM Jurnal Ilmiah*, *16*(2), 115–124. https://doi.org/10.33096/ilkom.v16i2.1948.115-124

Sumartini, S. H., & Purnam, S. W. (2022). The Use of the Classification and Regression Trees (CART) Method for the Classification of Recurrence of Cervical Cancer Patients at Dr. Soetomo Surabaya Hospital. *ITS Journal of Science and Arts*, *4*(2), 211–216. https://doi.org/10.12962/j23373520.v4i2.10673

Syairozi, M. I. (2021). ANALYSIS OF POVERTY IN THE AGRICULTURAL SECTOR (Case Study of Rice Commodity in Malang Regency). *Media Economics*, *28*(2), 113–128. https://doi.org/10.25105/me.v28i2.7169

Tjandra, W., Ginting, C., & Gunawan, S. (2023). Determination of Fertilizer Dosage Based on Fresh Fruit Bunch (FFB) Tonnage Data on Oil Palm Plantations. *AGROISTA: Journal of Agrotechnology*, *7*(1), 8–16. https://doi.org/10.55180/agi.v7i1.341