e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR TECHNOLOGY, UIS
https://ejpi.uis.edu.my
e-ISSN 2289-7909

VOL. 12 NO. 5, SPECIAL EDITION, (DECEMBER, 2025): page 257-271
DOI: 10.53840/ejpi.v12i5.322

Received: October 27, 2025 | Accepted: December 8, 2025 | Published: December 31, 2025

Analysis of Oil Palm Plant Nutrient Prediction Using the Random Forest Regression Method

Analisis Prediksi Nutrisi Tanaman Kelapa Sawit Menggunakan Metode Random Forest Regression

Farhan Andika Putra[1], Erwin Hermawan[2], Sahid Agustian Hudjimartsu[3], Abdul Haris[4]
[1,2,3] Informatics Engineering, Faculty of Technology & Sains, Universitas Ibn Khaldun Bogor Jl. Sholeh Iskandar, RT.01/RW.10, Kedungbadak, Kec. Tanah Sereal, Bogor City, West Java 16162 Indonesia.
Email: [1]*farhanandikaputra0@gmail.com*, [2]*rwien.gfm@gmail.com*, [3]*shudjimartsu@uika-bogor.ac.id*

ABSTRACT

Palm oil is an important commodity for Indonesia, so monitoring plant nutrition is crucial to maintain productivity. This research aims to develop a prediction system for palm oil nutrition—including nitrogen (N), phosphorus (P), potassium (K), and magnesium (Mg)—using multispectral drone photo-based remote sensing. The Random Forest Regression (RFR) method is used to model the relationship between spectral data and plant nutrient levels. The results showed that RFR was able to predict nutrients with adequate accuracy, with R^2 values of 0.641 (N), 0.601 (P), 0.558 (K), and 0.765 (Mg), respectively. These findings confirm that the integration of multispectral data and RFR is an effective approach to monitor nutrient conditions broadly and support more efficient fertilization decisions.

Keywords: Palm Oil Nutrition; Remote Sensing; Random Forest Regression; Multispectral

ABSTRAK

Minyak sawit merupakan komoditas penting bagi Indonesia, sehingga pemantauan nutrisi tanaman menjadi krusial untuk menjaga produktivitas. Penelitian ini bertujuan mengembangkan sistem prediksi nutrisi kelapa sawit—meliputi nitrogen (N), fosfor (P), kalium (K), dan magnesium (Mg)—menggunakan penginderaan jauh berbasis foto drone multispektral. Metode Random Forest Regression (RFR) digunakan untuk memodelkan hubungan antara data spektral dan kadar nutrisi tanaman. Hasil penelitian menunjukkan bahwa RFR mampu memprediksi nutrisi dengan akurasi yang memadai, dengan nilai R^2 masing-masing 0,641 (N), 0,601 (P), 0,558 (K), dan 0,765 (Mg). Temuan ini menegaskan bahwa integrasi data multispektral dan RFR merupakan pendekatan efektif untuk memantau kondisi nutrisi secara luas dan mendukung keputusan pemupukan yang lebih efisien.

Kata kunci: Nutrisi Kelapa Sawit; Penginderaan Jauh; Random Forest Regression; Multispektral.

©The Authors 2025. Published by Penerbit UIS. This is an Open Access article under the CC BY 4.0 license.

INTRODUCTION

Palm oil is a strategic commodity for the Indonesian economy, with a major contribution to exports and labor absorption (Ariyanto et al., 2022). To maintain productivity, crop nutrition monitoring needs to be carried out efficiently, especially in a very large plantation area. Remote sensing technology through satellite imagery and multispectral drones offers a fast, accurate, and cost-effective method of monitoring vegetation conditions, including plant health through indices such as NDVI (Yuniasih & Adjie, 2022; Evrilia, 2020).

Previous research has shown that multispectral data is able to predict plant nutrients, such as N, P, and K, with good accuracy, and can be a more efficient alternative to laboratory analysis (Ariyanto et al., 2022). In addition, the Random Forest Regression (RFR) method has been proven to be effective in processing spectral variables to produce stable nutrient predictions through bagging techniques and random feature selection (Wahyudi, 2022; Xie et al., 2021).

Based on this foundation, this study aims to identify and predict nutrient deficiencies N, P, K, and Mg in oil palm plants using multispectral drone photographs and RFR modeling. This approach is expected to provide more accurate nutritional information and support appropriate and sustainable fertilization decision-making.

PROBLEM STATEMENT

1. How to predict the nutrient content (N, P, K, and Mg) of oil palm plants using the Random Forest Regression method?
2. How to evaluate the accuracy of the prediction results of these nutrients compared to the data from laboratory analysis?

LITERATURE REVIEW

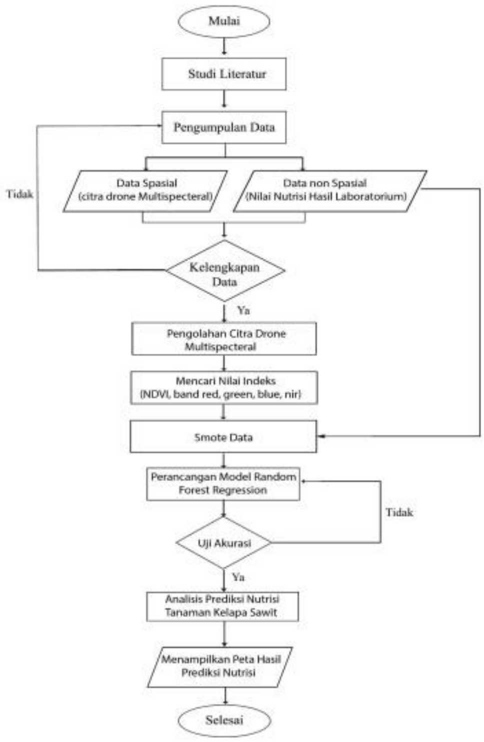
Table 1: Literature reviews

No.	Author, Title, Year of Research	Result	Relationship
1.	M. Hadisaputra Ariyanto, K. Rega Prilianti, H. Setiawan, and P. Mimboro. Method of Rapid Measurement of Nutrient Content of Oil Palm Plants with Remote Sensing and Artificial Intelligence Methods. (2022)	The Artificial Neural Network method has been proven to predict the nutrition of Nitrogen, Phosphorus and Potassium in oil palm plants using remote sensing. The ANN model will use an adam optimizer with Mean Absolute Error and Loss Mean Squared Error accuracy, while for hidden layers and nodes of hidden layers, trial and error will be performed to get the best results	Explain about analyzing nutrients in oil palm plants.
2.	Samuel, K. Rega Prilianti, H. Setiawan, and P. Mimboro. Automatic Tree Detection Method in Oil Palm Plantation Imagery Using Convolutional Neuteral Network (CNN) Model in Geographic Information System Software. (2022)	The area of oil palm plantations in Indonesia is around 14.33 million hectares and produced 42.9 million tons of oil palm in 2018. In 2019, the area of oil palm plantations in Indonesia grew by 1.88% to 14.60 million hectares and production increased by 12.92% to 48.42 million tons. In 2020, the area	Explaining the area of oil palm plantations in Indonesia

No.	Author, Title, Year of Research	Result	Relationship
		of oil palm plantations did not change significantly.	
3.	B. Yuniasih & A. R. Adjie. Evaluation of Oil Palm Plantation Conditions Using NDVI Index from Sentinel 2 Satellite Imagery. (2022)	The use of remote sensing technology using satellite imagery or drones is one way to know and monitor the condition of oil palm plantations effectively for a large area.	Explain the use of remote sensing technology
4.	Y. Huang, C. Zhong-xin, Y. Tao, H. Xiang-zhi, & G. Xing-fa. Big data remote sensing agriculture: Management and applications. (2018)	Remote sensing differentiates objects based on the differences in transmission, reflection, and absorption of electromagnetic waves by each object. Remote sensing works on visible, infrared, and microwave bands.	Explaining Remote Sensing
5.	W. Noviliasari, S. Dedy Kurnia, & M. A. Yulianandha The use of NDVI (Normalize Difference Vegetation Index) and SAVI (Soil Adjusted Vegetation Index) methods to determine the availability of green open space to meet oxygen needs. (2020)	The Normalize Difference Vegetation Index (NDVI) is an index of the photosynthetic activity of vegetation or vegetation, and one of the most commonly used vegetation indices.	Explaining the Normalization of Vegetation Index Differences (NDVI)

METHODOLOGY

Figure 1: Frame of mind



The research stages used in this study have three main components in the thesis preparation process consisting of literature study, data collection and data processing. In this research method, it can be seen from the flow diagram of the thinking framework in Figure 1 above.

Data Collection Methods

The first stage was a direct field survey at PT. Barito Putera Plantation, Barito Kuala Regency, South Kalimantan. Furthermore, real-time positioning is carried out using the Real Time Kinematic (RTK) method, which allows direct determination of coordinates even when the tool is in motion. The data taken was in the form of aerial photographs by taking through drones or drones with DJI Phantom 4 multispectral Drone type at PT. Barito Putera Plantation.

The image data used in this study includes five components, namely one NDVI (Normalized Difference Vegetation Index) and four red, green, blue, and near-infrared (NIR) spectral bands. These indices and bands are used to analyze vegetation conditions and other characteristics in the research area. Nutritional data was obtained from secondary data from the results of field surveys with a total of 345 sample points that will be processed and analyzed further.

Data Processing

At this stage, the goal is to find out how to process the data carried out in this study. In this study, the Random Forest Regression method was used. Data processing can be seen as follows:

1. UAV Image Processing
   At this stage, to find the NDVI index, the red, green, blue, and NIR bands use R-studio software to generate index images that incorporate UAV images. Then combine the results of the index that has been processed using Arcgis software. To find the NDVI index by combining NIR and red bands using the NDVI formula, for red, green, blue band indices, NIR there is no need to search in the R-studio software as it is already available from multispectral drone photos, just combine them using Arcgis software.
2. Index Value Extraction
   To obtain the index value, by extracting a raster based on the nutritional value of laboratory results which has 345 plot points of oil palm plants.
3. Oversampling Data
   At this stage, data balancing is carried out using the SMOTE (Synthetic Minority Oversampling Technique) technique. This SMOTE technique works by looking for the closest neighbours of the K data in each minority class. Then the synthetic data is duplicated in the minor class and the environment as many percentages as are created, which will later be randomly selected.

RFR Model Design

The analysis method used in this method is the Random Forest Regression (RFR) method. Random Forest Regression is a Machine Learning algorithm that can predict and forecast well. With the concept of randomly combining decision trees. This method is aided by a gridsearch cross-validation algorithm that will combine grid search and cross-validation to select the optimal parameters.

1. Select some data in a training set as many as K pieces (If the data is too little, for example less than 30, there is no need to divide it into training and test sets).
2. Create a Decision Tree from the pre-selected K data.
3. Select the number of N-trees you want to create. Next, repeat steps 1 and 2. The point is to keep regressing the decision tree as much as possible (generally as many as 200 times, 300, 500, etc.).
4. Determine the metric to use, which is the RMSE metric for regression analysis.
5. The process of normalizing the data as needed is then created to create a model.
6. Evaluation for the regression of mtry metrics, Root Mean Squared Error (RMSE), and Determination Coefficient (R-squared)

Prediction of Oil Palm Plant Nutrition with Laboratory Results

By obtaining parameter values from the gridseacrhCV algorithm, it was followed by analyzing the nutrient value prediction model using the Random Forest Regression method.

RFR Model Accuracy

This stage was carried out to measure the performance of the Random Forest Regression method that has been carried out in analyzing the nutrition prediction of oil palm plants using the random forest regression method.

In the results of the accuracy of the RFR model, there are several metrics in it, including:

1. Parameter Mtry Parameter
   Mtry is the sum of independent variables used to build a tree on each iteration. This parameter is necessary to create a model to accept the best model with the lowest possible error value (Hakim et al., 2023).
2. Mean-Squared Root (RMSE)
   RMSE is often used to measure the difference between the values predicted by the example and the values observed. The RMSE penalizes a greater absolute value by giving more weight to that value. The greater the RMSE disparity, the greater the individual fault variant (Li et al., 2018).
3. R-Square (R2)
   R-Squared represents the proportion of independent variable variants described by independent variables in the model. An R2 A score of 1 provides a perfect match between predicted and experimental outcomes, while a score of 0 provides no relationship between independent and dependent variables (Lillo-Bravo et al., 2023).

FINDINGS AND DISCUSSION

At this stage, each sample point is made into a polygon that follows the shape of the leaves of the oil palm plant. The purpose of creating these polygons is to ensure higher accuracy when extracting index and image values on each ribbon. With polygon shapes adapted to the leaf structure, data extraction can reflect vegetation conditions more representatively. This process is carried out using ArcGIS software, which will facilitate further analysis at a later stage, such as the calculation of vegetation indexes or the processing of other image data.

Extraction of Vegetation Index Values and Band Images

At this stage, ArcGIS software is used to find the index values on each component of the image. This process is done by calculating and visualizing index values using ArcGIS. The results of the calculation are shown in Table 1, which contains five image components, namely NDVI, Red, Green, Blue, and Near-Infrared (NIR).

The data in this table shows the rate of light reflection in each colour band (red, green, blue) as well as near infrared (NIR), which is the basis for calculating NDVI. NDVI values in the table range from 0.872 to 0.907, where higher values indicate areas with healthier vegetation and active photosynthesis. Each row of data in the table represents the value of a single point or pixel in the image, which shows the variation in vegetation conditions in the area. This analysis helps in understanding the distribution of vegetation and identifying land conditions based on the level of vegetation health in the mapped area. This NDVI data can be used for further evaluation in land use planning, crop health monitoring, and general environmental management, as shown in Table 2.

Table 2: Value Extraction Results from Each Image Component

No	NDVI	RED	GREEN	BLUE	NIR
1	0,904	0,027	0,057	0,020	0,546
2	0,897	0,024	0,047	0,017	0,454
3	0,903	0,025	0,050	0,019	0,511
4	0,894	0,027	0,054	0,018	0,486
5	0,893	0,028	0,061	0,019	0,511
6	0,899	0,026	0,055	0,019	0,514
7	0,880	0,030	0,067	0,019	0,485
8	0,904	0,028	0,060	0,020	0,567
9	0,902	0,024	0,052	0,017	0,479

10	0,907	0,026	0,061	0,018	0,552
...	...	...	...	...	...
340	0,886	0,032	0,073	0,019	0,520
341	0,890	0,029	0,069	0,017	0,510
342	0,893	0,028	0,064	0,018	0,506
343	0,876	0,029	0,062	0,018	0,452
344	0,893	0,029	0,061	0,019	0,521
345	0,894	0,031	0,070	0,019	0,548
346	0,872	0,031	0,066	0,019	0,468

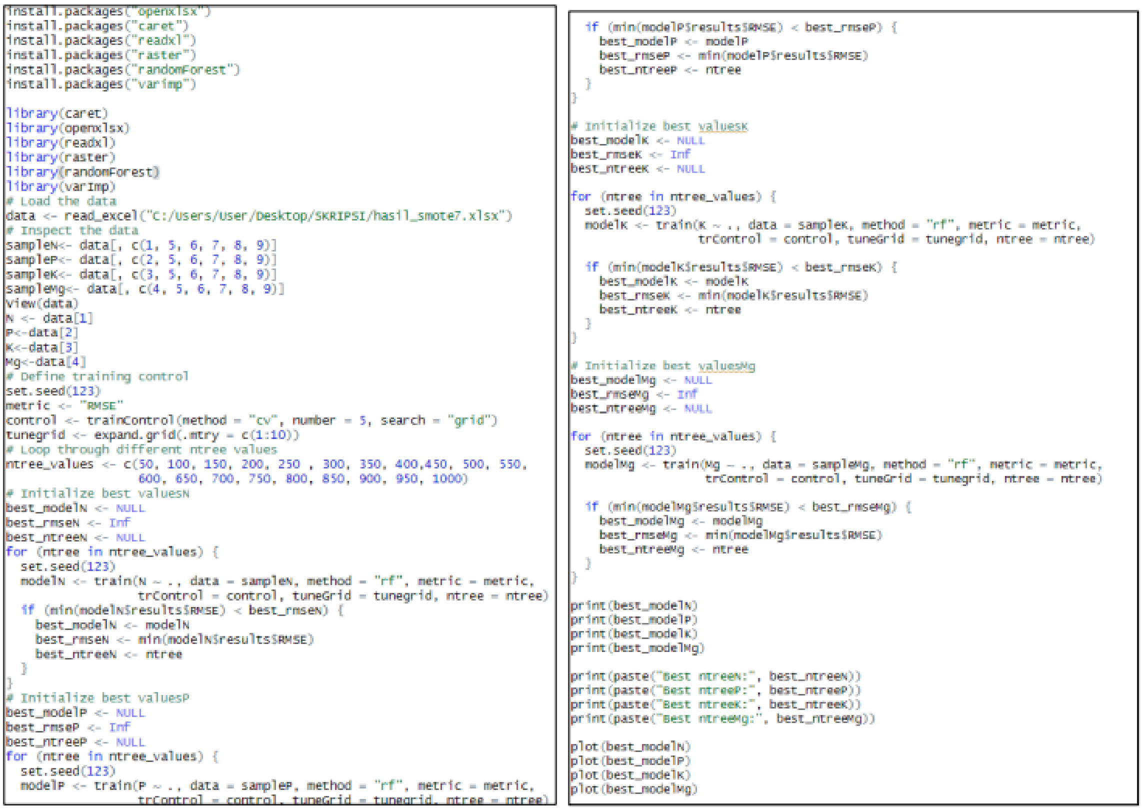
Random Forest Regression

The Random Forest Regression method was used to predict the nutrients of oil palm and produce an accuracy (R^2) of N = 0.641; P = 0.601; K = 0.558; and Mg = 0.765. Before modelling, the optimal parameters are determined using GridSearchCV to obtain the best combination of hyperparameters.

Figure 2 shows a predictive modelling flow using Random Forest to predict four nutrient variables (N, P, K, and Mg). The initial data was processed using the SMOTE technique to overcome class imbalances, then read and processed using the openxlsx and readxl libraries. The randomForest library is used to build the model, while varImp is utilized to assess the importance of variables.

Each target variable is separated into a separate subset of data. The model was trained using a 5-fold cross-validation scheme to improve generalization capabilities. Hyperparameters, specifically mtry, are set through a grid search, while ntree is tested in some variation through a loop. The final model is selected based on the best performance in predicting each nutrient.

Figure 2: Random Forest Regression Source Code



```
install.packages("openxlsx")
install.packages("caret")
install.packages("readxl")
install.packages("raster")
install.packages("randomForest")
install.packages("varimp")

library(caret)
library(openxlsx)
library(readxl)
library(raster)
library(randomForest)
library(varImp)
# Load the data
data <- read_excel("C:/Users/User/Desktop/SKRIPSI/hasil_smote7.xlsx")
# Inspect the data
sampleN<- data[, c(1, 5, 6, 7, 8, 9)]
sampleP<- data[, c(2, 5, 6, 7, 8, 9)]
sampleK<- data[, c(3, 5, 6, 7, 8, 9)]
sampleMg<- data[, c(4, 5, 6, 7, 8, 9)]
View(data)
N <- data[1]
P<-data[2]
K<-data[3]
Mg<-data[4]
# Define training control
set.seed(123)
metric <- "RMSE"
control <- trainControl(method = "cv", number = 5, search = "grid")
tunegrid <- expand.grid(.mtry = c(1:10))
# Loop through different ntree values
ntree_values <- c(50, 100, 150, 200, 250 , 300, 350, 400,450, 500, 550,
                  600, 650, 700, 750, 800, 850, 900, 950, 1000)
# Initialize best valuesN
best_modelN <- NULL
best_rmseN <- Inf
best_ntreeN <- NULL
for (ntree in ntree_values) {
  set.seed(123)
  modelN <- train(N ~ ., data = sampleN, method = "rf", metric = metric,
                  trControl = control, tuneGrid = tunegrid, ntree = ntree)
  if (min(modelN$results$RMSE) < best_rmseN) {
    best_modelN <- modelN
    best_rmseN <- min(modelN$results$RMSE)
    best_ntreeN <- ntree
  }
}
# Initialize best valuesP
best_modelP <- NULL
best_rmseP <- Inf
best_ntreeP <- NULL
for (ntree in ntree_values) {
  set.seed(123)
  modelP <- train(P ~ ., data = sampleP, method = "rf", metric = metric,
                  trControl = control, tuneGrid = tunegrid, ntree = ntree)
  if (min(modelP$results$RMSE) < best_rmseP) {
    best_modelP <- modelP
    best_rmseP <- min(modelP$results$RMSE)
    best_ntreeP <- ntree
  }
}

# Initialize best valuesK
best_modelK <- NULL
best_rmseK <- Inf
best_ntreeK <- NULL

for (ntree in ntree_values) {
  set.seed(123)
  modelK <- train(K ~ ., data = sampleK, method = "rf", metric = metric,
                  trControl = control, tuneGrid = tunegrid, ntree = ntree)

  if (min(modelK$results$RMSE) < best_rmseK) {
    best_modelK <- modelK
    best_rmseK <- min(modelK$results$RMSE)
    best_ntreeK <- ntree
  }
}

# Initialize best valuesMg
best_modelMg <- NULL
best_rmseMg <- Inf
best_ntreeMg <- NULL

for (ntree in ntree_values) {
  set.seed(123)
  modelMg <- train(Mg ~ ., data = sampleMg, method = "rf", metric = metric,
                   trControl = control, tuneGrid = tunegrid, ntree = ntree)

  if (min(modelMg$results$RMSE) < best_rmseMg) {
    best_modelMg <- modelMg
    best_rmseMg <- min(modelMg$results$RMSE)
    best_ntreeMg <- ntree
  }
}

print(best_modelN)
print(best_modelP)
print(best_modelK)
print(best_modelMg)

print(paste("Best ntreeN:", best_ntreeN))
print(paste("Best ntreeP:", best_ntreeP))
print(paste("Best ntreeK:", best_ntreeK))
print(paste("Best ntreeMg:", best_ntreeMg))

plot(best_modelN)
plot(best_modelP)
plot(best_modelK)
plot(best_modelMg)
```

In the modelling process, the number of trees (*ntrees*) is varied between 50–1000 to find the best configuration. Each model is evaluated using RMSE, where the lowest RMSE value is selected as the optimal model. At each iteration, if a new model produces a better RMSE, then the best models and metrics are updated.

Once all variables (N, P, K, Mg) have been trained, a summary of the optimal performance and number of trees is displayed. Visualizations are then created to see the model's performance more clearly. The best model is then used to generate predictions, which are then combined with the original data using *cbind* to facilitate comparative analysis.

The final result—in the form of actual data and predicted values—is stored in a new Excel file through *write.xlsx*, making it easier to document and distribute the analysis results. Figure 3 shows an example of the best model calculation results for the prediction of the N element.

Figure 3: Results of Random Forest Regression Model N

```
> print(best_modelP)
Random Forest

1313 samples
   5 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 1050, 1051, 1050, 1051, 1050
Resampling results across tuning parameters:

  mtry  RMSE         Rsquared   MAE
   1    0.003108990  0.5886272  0.002341049
   2    0.003021006  0.6017272  0.002148824
   3    0.003013761  0.5997391  0.002094637
   4    0.003026473  0.5938092  0.002070526
   5    0.003049114  0.5859519  0.002067409
   6    0.003044086  0.5873160  0.002064140
   7    0.003046726  0.5868685  0.002068415
   8    0.003046600  0.5867207  0.002064189
   9    0.003049296  0.5859497  0.002068175
  10    0.003044402  0.5872799  0.002068638

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 3.
```

Figure 3: Output print(best_modelN) shows a summary of the best Random Forest model for the target variable N, which was trained with 1313 samples and 5 predictors without additional preprocessing. This model uses 5-fold cross-validation, with an even sample size in each fold. The results table shows the evaluation for mtry values from 1 to 10 based on three metrics: RMSE (Root Mean Square Error), R-squared, and MAE (Mean Absolute Error).

The lowest RMSE value was achieved at mtry = 3, which is 0.06759118, which indicates the model has a good performance in predicting the N. R-squared for this model is 0.6419371, indicating that about 64.19% of the variance in the data can be explained by the model. Thus, the best model uses mtry = 3 because it provides a good balance between complexity and prediction accuracy.

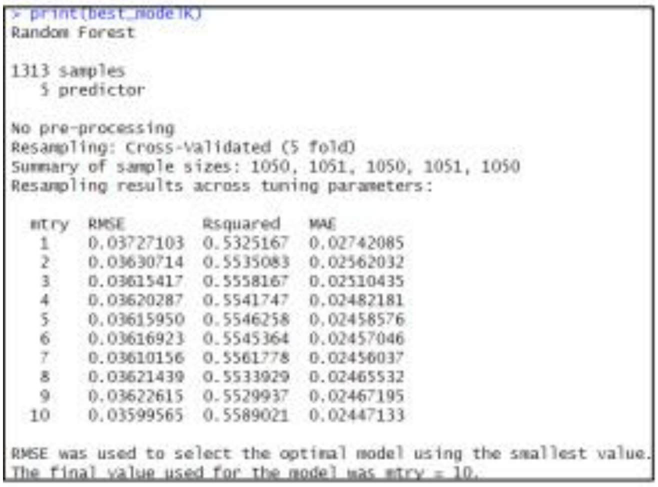
Figure 4: Results of the Random Forest Regression Model P

```
> print(best_modelN)
Random Forest

1313 samples
   5 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 1050, 1050, 1049, 1052, 1051
Resampling results across tuning parameters:

  mtry  RMSE        Rsquared   MAE
   1    0.07010942  0.6267493  0.05612385
   2    0.06814555  0.6392167  0.05183070
   3    0.06759118  0.6419371  0.05018186
   4    0.06793952  0.6366011  0.04979644
   5    0.06885430  0.6245402  0.05004231
   6    0.06869534  0.6262748  0.04978602
   7    0.06853517  0.6285335  0.04980628
   8    0.06874407  0.6256216  0.04991788
   9    0.06858043  0.6278917  0.04975714
  10    0.06860363  0.6274675  0.04982011

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 3.
```

Figure 4: The print(best_modelP) output summarizes the best Random Forest model for the target variable P, which was trained with 1313 samples and 5 predictors without additional preprocessing, and used 5x cross-validation with consistent sample sizes. The evaluation results table shows mtry values from 1 to 10, with the lowest RMSE recorded at mtry = 3, which is 0.003013761, which indicates excellent predictive ability. The R-squared value for this model is 0.5997391, indicating that about 59.97% of the variance in the data can be explained by the model, while the MAE of

0.002094637 indicates a very low prediction error. Thus, the best model uses mtry=3 because it provides an optimal balance between complexity and prediction accuracy.

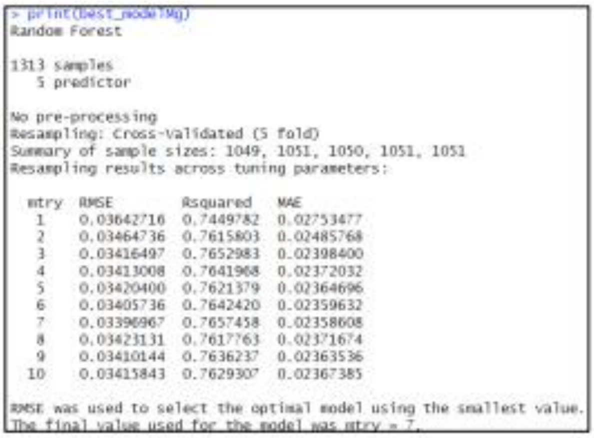
Figure 5: Results of the Random Forest Regression Model K



```
> print(best_modelK)
Random Forest

1313 samples
   5 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 1050, 1051, 1050, 1051, 1050
Resampling results across tuning parameters:

  mtry  RMSE        Rsquared   MAE
   1    0.03727103  0.5325167  0.02742085
   2    0.03630714  0.5535083  0.02562032
   3    0.03615417  0.5558167  0.02510435
   4    0.03620287  0.5541747  0.02482181
   5    0.03615950  0.5546258  0.02458576
   6    0.03616923  0.5545364  0.02457046
   7    0.03610156  0.5561778  0.02456037
   8    0.03621439  0.5533929  0.02465532
   9    0.03622615  0.5529937  0.02467195
  10    0.03599565  0.5589021  0.02447133

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 10.
```

Figure 5: Output print(best_modelK) provides a summary of the best Random Forest model for the target variable K, which was trained with 1313 samples and 5 predictors without additional preprocessing. This model uses 5-fold cross-validation, with fairly consistent sample sizes, namely 1050, 1051, 1050, 1051, and 1050. The evaluation results table displays mtry values from 1 to 10, where RMSE (Root Mean Square Error), R-squared (Rsquared), and MAE (Mean Absolute Error) are recorded for each mtry value. The lowest RMSE value was found at mtry = 10, which is 0.03599565, indicating that the model has good predictive capabilities for the K variable.

Figure 6: Results of Random Forest Regression Model Mg



```
> print(best_modelMg)
Random Forest

1313 samples
   5 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 1049, 1051, 1050, 1051, 1051
Resampling results across tuning parameters:

  mtry  RMSE        Rsquared   MAE
   1    0.03642716  0.7449782  0.02753477
   2    0.03464736  0.7615803  0.02485768
   3    0.03416497  0.7652983  0.02398400
   4    0.03413008  0.7641968  0.02372032
   5    0.03420400  0.7621379  0.02364696
   6    0.03405736  0.7642420  0.02359632
   7    0.03396967  0.7657458  0.02358608
   8    0.03423131  0.7617763  0.02371674
   9    0.03410144  0.7636237  0.02363536
  10    0.03415843  0.7629307  0.02367385

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 7.
```

The R-square for this model is 0.5589021, indicating that about 55.89% of the variance in the data can be explained by the model, while the MAE is recorded as 0.02447133, indicating a relatively low prediction error. Thus, the best model for the prediction of the K variable is the one that uses mtry = 10, as it provides the lowest RMSE value and shows a good balance between the complexity of the model and the accuracy of the prediction.

Figure 6: Output print(best_modelMg) provides a summary of the best Random Forest model for the target variable Mg, which was trained with 1313 samples and 5 predictors without additional preprocessing, and used 5-fold cross-validation with relatively consistent sample sizes, i.e. 1049, 1051, 1050, 1051, and 1051. The evaluation table displays mtry values from 1 to 10, where the lowest RMSE value is found at mtry = 7, which is 0.03396967, indicating excellent predictive ability. The R-squared value for this model reaches 0.7657458, indicating that about 76.57% of the variance in the data can be

explained by the model, while the MAE is recorded as 0.02358608, indicating a relatively small prediction error.

Thus, the best model for predicting the Mg variable is the one that uses mtry = 7, as it produces the lowest RMSE value and shows a good balance between model complexity and prediction accuracy.

Prediction of Oil Palm Plant Nutrition

After getting the best parameter values, the next step is to look for the prediction of the nutrition of oil palm plants. The prediction of oil palm plant nutrition uses five image components, namely NDVI, red band, green, blue and NIR. The following source code is used in finding the prediction of oil palm plant nutrition shown in Figure 7.

Figure 7: Nutrient Prediction Source Code N, P, K, MG

```
# Prediksi menggunakan model terbaik
prediksiN <- predict(best_modelN, data)
prediksiP <- predict(best_modelP, data)
prediksiK <- predict(best_modelK, data)
prediksiMg <- predict(best_modelMg, data)
hasil <- cbind(data, prediksiN, prediksiP, prediksiK, prediksiMg)
view(hasil)

# Save file
write.xlsx(hasil, file="C:/Users/User/Desktop/SKRIPSI/PREDIK_UNSURHARA2.xlsx")
```

In Figure 7, this source code is used to make predictions using the best pre-trained Random Forest model, which is applied to the preloaded dataset. First, four prediction variables are created using the predict() function for each of the best models (best_modelN, best_modelP, best_modelK, and best_modelMg). The results of this prediction function are stored in the prediction variables N, prediction P, prediction K, and prediction Mg, so that each of these variables stores the prediction value for the target variables N, P, K, and Mg.

Table 3: Prediction of Oil Palm Plant Nutrition

No	Prediction N	P Prediction	Prediction K	Mg Prediction
1	1,96	0,12	0,62	0,27
2	1,97	0,12	0,54	0,34
3	1,93	0,12	0,55	0,35
4	1,88	0,12	0,66	0,29
5	2,03	0,12	0,55	0,28
6	2,02	0,12	0,57	0,29
7	1,86	0,12	0,66	0,28
8	1,87	0,12	0,65	0,28
9	1,94	0,12	0,54	0,30
10	2,01	0,12	0,55	0,29
...	...	...	...	...
1308	1,81	0,12	0,53	0,38
1309	1,79	0,12	0,53	0,39
1310	1,79	0,12	0,53	0,39
1311	1,79	0,12	0,53	0,39
1312	1,79	0,12	0,53	0,39
1313	1,79	0,12	0,53	0,39

Next, the prediction results are combined with the original data using the cbind() function, resulting in a new data framework named results, which includes the original data along with the predicted results. After the merge, these dataframes are displayed using the view, so users can easily check and analyze the results of the predictions that have been made.

Thus, this code facilitates users to get a thorough overview of the results of the prediction model and helps in further interpretation and analysis. The resulting dataframe is then saved into an Excel file using write.xlsx(), with a specified storage location. The predicted results of nutrition in oil palm plants, including nitrogen (N), phosphorus (P), potassium (K), and magnesium (Mg), are presented in the form of predictive values for each nutrient variable based on the trained Random Forest model, as shown in Table 3.

Accuracy Results

The accuracy results obtained from the analysis of oil palm plant nutrition predictions used laboratory nutritional values, vegetation index and ribbon image index which resulted in R-squared (R2) values N = 0.641, P = 0.601, K = 0.558, Mg = 0.765. The value of R 2 is used to measure how much an independent variable affects a dependent variable. And there are results of the calculation accuracy of RMSE Error (Root Mean Squared Error) which is N = 0.067, P = 0.003, K = 0.035 and Mg = 0.033.

Oil Palm Plant Nutrition Prediction Map

This study used five indices, namely NDVI, red band, green, blue, and near-infrared (NIR), and combined data processing to analyze the results of the nutrient distribution of oil palm plants, shown in Figure 8.

Figure 8: Source Code Stacking Raster

```
# Prediksi menggunakan model terbaik
prediksiN <- predict(best_modelN, data)
prediksiP <- predict(best_modelP, data)
prediksiK <- predict(best_modelK, data)
prediksiMg <- predict(best_modelMg, data)
hasil <- cbind(data, prediksiN, prediksiP, prediksiK, prediksiMg)
view(hasil)

# Save file
write.xlsx(hasil, file="C:/Users/User/Desktop/SKRIPSI/PREDIK_UNSURHARA2.xlsx")
```

The source code in Figure 7 is used to modify the raster data by adding new attributes and saving them back into TIFF format. First, the dysmo library is enabled to provide the functionality necessary for spatial analysis.

Then, all raster files with .tif extension are retrieved from the specified directory using list.files() and stored in predictor_files variable. Next, the stack() function is used to combine those raster files into a single raster object called a predictor. To ensure that the raster name matches the model variable, the name in the predictor object is changed using names(), with corresponding names such as "rata_rata_ndvi", "rata_rata_red", "rata_rata_green", "rata_rata_blue", and "rata_rata_nir". Once the raster object is ready, the next step is to create a prediction for each variable. For N, predict() is used to apply the best model (best_modelN) to the predictor object, and the results are stored in predN. Then, the results of this prediction are visualized using plot() with the title "Prediction N". The results of these predictions are also saved into a new raster file at the specified location by using writeRaster().

The same steps are performed for the P, K, and Mg variables. The prediction results are stored in the predP, predK, and predMg variables, then mapped and stored in a separate raster file with the corresponding name in the specified directory. With these steps, this code facilitates the process of predicting and storing the prediction results for further analysis.

Figure 9: Nitrogen Prediction Results Map (N)

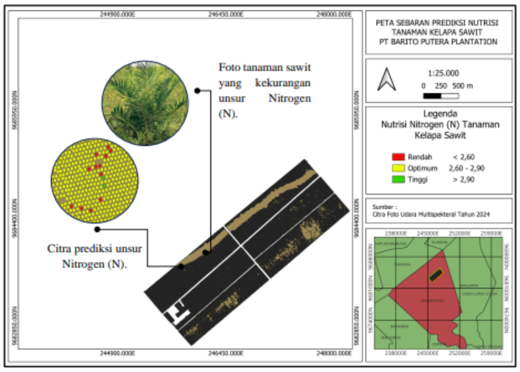


Figure 9 shows the Map of Nitrogen (N) Nutrient Prediction Results in Oil Palm Plants with additional details, including:

1. Prediction of Nitrogen (N) Nutrient Distribution: This map provides information on the status of nitrogen nutrients in oil palm plantation land. The colour on the map indicates nitrogen nutrient levels: Red indicates areas with low nitrogen nutrients (2.90). Yellow indicates areas with nitrogen nutrients at optimal levels (2.60 - 2.90). • Green indicates areas with high nitrogen nutrients (>2.90).
2. Illustration of Nitrogen Deficiency:
   a. There are photos of oil palm plants that show symptoms of nitrogen deficiency. It is usually characterized by yellowing of the leaves, which is one of the signs of nitrogen deficiency in plants.
   b. A small circle image with red dots shows a predictive image of the nitrogen element in a given area. It may use technologies such as multispectral imaging to detect nitrogen deficiencies in specific crops.
3. Data Source: This data is based on multispectral aerial photograph imagery taken in 2024.
4. Location Index Map: At the bottom right, there is an additional map showing the location of oil palm plantations in the context of a larger area (in red).
5. Map Scale and Direction: The map scale uses 1:25,000 with north-directional arrows, which helps in determining the orientation and estimation of distances in the map. The conclusion of this map is that there are certain parts of the land that are deficient in nitrogen, which may require further fertilization measures to achieve nutrient balance and improve the health of oil palm plants.

Figure 10: Phosphorus Prediction Results Map (P)

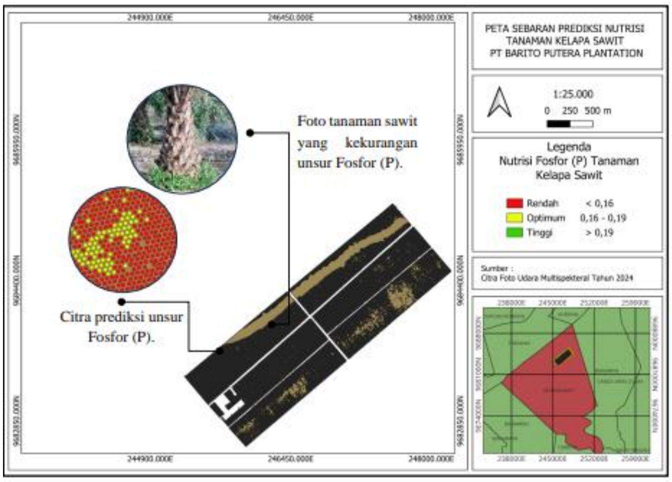


Figure 10 shows the Phosphorus (P) Nutrient Prediction Results Map on Oil Palm Plants with additional details including:

1. Predicted Phosphorus Nutrient Distribution (P): This map provides information on the status of phosphorus nutrients in oil palm plantations. The colour on the map shows phosphorus nutrient levels Red indicates areas with low phosphorus nutrients (0.19). Green indicates areas with phosphorus nutrients at optimal levels (0.16 - 0.19). Yellow indicates areas with high phosphorus nutrients (>0.19).
2. Illustration of Phosphorus Deficiency:
   a. There are photos of oil palm plants that show symptoms of phosphorus deficiency.
   b. A small circle image with red dots shows a picture of the predicted phosphorus levels in a given area.
3. Data Source: This data is based on multispectral aerial photograph imagery taken in 2024.
4. Location Index Map: At the bottom right there is an additional map showing the location of oil palm plantations in the context of a larger area (in red).

5. Map Scale and Direction: The map scale uses 1:25,000 with a northbound point. The map shows that some parts of the land are deficient in phosphorus, which may require further fertilization measures to achieve nutrient balance and improve the health of oil palm plants.

Figure 11: Potassium (K) Prediction Result Map

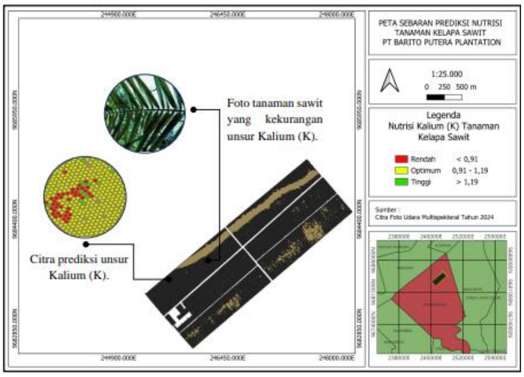
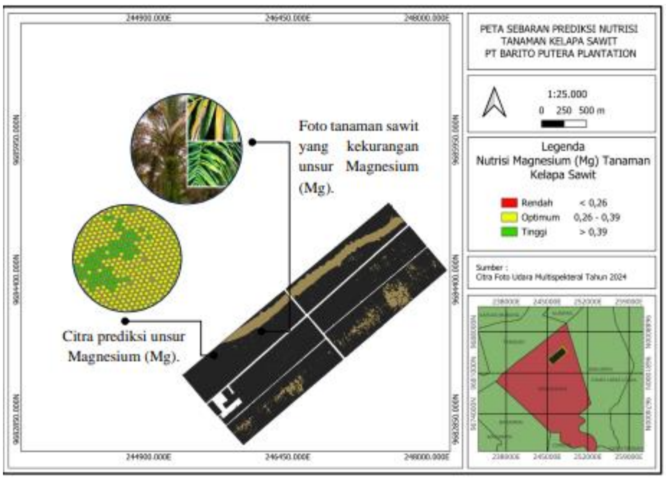


Figure 11 is the Map of Potassium (K) Prediction Results in Oil Palm Plants in PT Barito Putera's Plantation, which shows the distribution of potassium nutrients in the plantation area. Here are the details:

1. Predicted Potassium Nutrient Distribution (K): This map identifies potassium levels in the soil using three colors: Red indicates low potassium levels (1.19), which means the area has excessive or very adequate potassium nutrients. Yellow indicates optimal potassium levels (0.91 - 1.19), indicating sufficient nutrients for plant growth. Green: indicates high potassium levels (>1,19), which means the area has excessive or very adequate potassium nutrients.
2. Illustration of potassium deficiency:
   a. The photo of the palm leaves in the top left shows symptoms of potassium deficiency, which is usually seen on leaves that have yellowing or necrotic spots.
   b. A small circle image with red dots shows a predictive image of the potassium element, showing specific areas with varying levels of potassium.
3. Data Source: Data from multispectral aerial imagery in 2024, allows for a detailed analysis of potassium nutrients in a large area of plantations.
4. Location Index Map: The additional map at the bottom right shows the location of the plantation on a larger scale (in red), providing context for the area being analyzed.
5. Map Scale and North Direction: The map scale is 1:25,000, equipped with north arrows for orientation and distance measurement. Overall, the map is used to identify areas of plantations that require additional potassium nutrient interventions to maintain the health of oil palm crops and increase productivity.

Figure 12: Magnesium (Mg) Prediction Results Map



In Figure 12 there is a map of the Magnesium (Mg) Prediction Results in Oil Palm Plants at PT Perkebunan Barito Putera, which shows the distribution of magnesium content in oil palm plantation areas. Here are the details:

1. Predicted Magnesium Nutrient Distribution (Mg): This map uses three colours to show magnesium levels: Red indicates low magnesium levels (0.39), which means the area has sufficient or excess magnesium nutrients. Yellow indicates optimal magnesium levels (0.26 - 0.39), indicating areas with adequate nutritional conditions. Green indicates high magnesium levels (>0.39), which means the area has sufficient or excessive magnesium nutrients
2. Magnesium Deficiency Illustration:
   a. There is a photo of palm leaves in the upper left that shows symptoms of magnesium deficiency, usually characterized by yellowing of the leaves, especially on the leaf bones and the area in between.
   b. A small circle image with green dots is a predictive image of the element magnesium.
3. Data Source: The data on this map is derived from multispectral aerial imagery in 2024, which allows for more accurate analysis of magnesium nutrients in plantation areas.
4. Location Index Map: There is an index map that shows the location of oil palm plantations in the context of a larger area (in red), providing a geographical context for the research site.
5. Map Scale and North Direction: The map scale is 1:25.000, equipped with a north arrow that aids with orientation. The map provides information to identify areas that are deficient in magnesium and require interventions, such as additional fertilization, to maintain the nutritional balance of oil palm plants to support crop productivity and health.

The higher prediction accuracy of the element magnesium (Mg) is influenced by the physiological and spectral characteristics of Mg which are more consistent with the colour change and health of the leaves. Mg plays an important role in the formation of chlorophyll, so its deficiency causes a clear and stable change in leaf colour, especially in red and NIR canals. The relationship between Mg and vegetation index (e.g. NDVI, GNDVI) is also stronger than that of other elements such as N, P, and K, which tend to be more influenced by environmental conditions and soil dynamics. In addition, the variation of Mg data in the sample is generally more structured so that the Random Forest model is easier to recognize the prediction pattern. A more pronounced leaf response to Mg deficiency results in clearer spectral signals, thereby improving the model's ability to produce accurate predictions.

The difference in prediction accuracy between nutrients is due to variations in nutrient sensitivity to spectral characteristics recorded by multispectral sensors. Magnesium (Mg) has the highest accuracy due to the following factors:

1. More Consistent Mg Effect on Mg Leaves plays an important role in the formation of chlorophyll. A deficiency or excess of Mg often gives rise to a pronounced discoloration of the leaves, making them easier to recognize through spectral reflections, especially in red and NIR channels. This consistency makes the spectral pattern more stable and easy for the model to learn.
2. Stronger Spectral – Physiological Relationships Mg variations typically show a high correlation with vegetation indices such as NDVI or GNDVI. This strong correlation is what enhances the ability of the Random Forest model to find predictive patterns.
3. In many cases, the spread (variance) of Mg data is more uniform than that of N, P, or K. When data variability is more structured, machine learning models can learn patterns more stably, resulting in increased accuracy.
4. Influence of Lower External Factors Nutrients such as N and P tend to be more influenced by environmental conditions, leaf age, rainfall, or soil dynamics that are difficult to record by drone sensors. In contrast, Mg is relatively more stable so its spectral signals are not much "disrupted" by external factors.
5. Leaf Response Is More Pronounced to Mg Imbalance Symptoms of Mg deficiency (e.g., chlorosis between leaf bones) produce strong spectral contrast. This provides a clearer differentiating feature for the Random Forest algorithm, making predictions more accurate.

CONCLUSION

Based on the results of the research on the design of the Random Forest Regression model for oil palm plant nutrition prediction, it can be concluded that the results of the oil palm plant nutrition prediction model use the Random Forest Regression method with the amount of Nitrogen (N) ± 1.78-2.06, Phosphorus (P) ± 0.11-0.12, Potassium (K) ± 0.53-0.67, and Magnesium (Mg) ± 0.21-0.39, which is the result of predicting the nutritional value of oil palm plants.

In addition, in this study, using 5 indices, including NDVI, red, green, blue, and NIR bands, accuracy results were obtained with the values of R-Squared (R2) Nitrogen (N) = 0.641, Phosphorus (P) = 0.601, Potassium (K) = 0.558, Magnesium (Mg) = 0.765, with a Root Mean Squared Error (RMSE) of N = 0.067, P = 0.003, K = 0.035 and Mg = 0.033.

ACKNOWLEDGEMENT

The author would like to express his deepest gratitude for his assistance in the preparation of writing a thesis proposal, both moral and material so that the writing of this thesis proposal can be completed properly, for his prayers, enthusiasm, motivation and support so far, the author would like to thank the following parents: Both parents for their prayers, enthusiasm and motivation and support so far, Dr. Ir. Muhammad Nanang Prayudianto, M.Sc. as the Dean of the Faculty of Engineering & Science, Ibnu Khaldun University Bogor, Fitrah Satrya Fajar Kusumah S.Komp., M.Kom., as the Head of the Informatics Engineering Study Program, Faculty of Engineering, Ibnu Khaldun University Bogor. Dr. Erwin Hermawan, S.Si., M.Sc as the Main Supervisor who has helped provide direction, encouragement, and guidance to the writer. Sahid Agustian Hudjimartsu S.T., M.Kom as a Companion Supervisor who never tires of helping to provide direction, encouragement, and guidance to writers. All Lecturers and Staff at the Faculty of Engineering & Science, Ibnu Khaldun University Bogor, especially Lecturers and Staff in the Informatics Engineering Study Program. Friends of Geospatial Information Technology 2020 who have helped and motivated the author in completing the final project. Myself so far in completing the final project. The author realizes that in compiling this final project report, it is still far from perfect, considering the limited knowledge and skills possessed. Therefore, what is presented in this report is inseparable from its shortcomings.

REFERENCES

Arini, D., Sari, SM, & Driptufany, D.M. (2022). The use of Landsat 8 imagery to detect the health level of oil palm plants using the Normalized Difference Vegetation Index (NDVI) method in Bengkalis Regency, Mandau District. *El-Jughrafiyah*, *2*(2), 50. https://doi.org/10.24014/jej.v2i2.18744

Ariyanto, M. H., Prilianti, K. R., & ... (2022). Method of Rapid Measurement of Nutrient Content of Oil Palm Plants with Remote Sensing and Artificial Intelligence Methods. *National Seminar on the Environment...*, *September*, 3218–3228.

Armita, D., Wahdaniyah, W., Hafsan, H., & Al Amanah, H. (2022). Visual diagnosis of important nutritional problems in different types of plants. *Technoscience: Science and Technology Information Media*, *16*(1), 139–150. https://doi.org/10.24252/teknosains.v16i1.28639

Evrilia, N. (2020). *Analysis of the level of productivity and health of oil palm using multi-species aerial photo data and lidar (Case Study: Batin XXIV Regency, Jambi Province).*

Hakim, A. R., Atmaja, D. M. U., Basri, A., & Syafii, M. (2023). Implementation of Random Forest Algorithm on Palm Oil Price Data. *Journal Of Tech- E*, *6*(2), 34–42. http://bsti.ubd.ac.id/e-jurnal

Kurnia, D., & Yulianandha, A. (2019). *The use of NDVI (Normalized Difference Vegetation Index) and SAVI (Soil Adjusted Vegetation Index) methods to determine the availability of green open space to meet oxygen needs (Case Study: Yogyakarta City).*

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J.C., van den Bossche, P., Van Mierlo, J., & Omar, N. (2018). Random forest regression for the estimation of the online capacity of lithium-ion batteries. *Applied Energy*, *232* (September), 197–210. https://doi.org/10.1016/j.apenergy.2018.09.182

Lillo-Bravo, I., Vera-Medina, J., Fernandez-Peruchena, C., Perez-Aparicio, E., Lopez-Alvarez, J.A., & Delgado-Sanchez, J.M. (2023). Random Forest Model to predict the performance of a solar water heating system. *Renewable Energy*, *216* (July), 119086 https://doi.org/10.1016/j.renene.2023.119086

Mustajab, R. (2022). *The area of Indonesia's oil palm plantations is almost 15 million hectares in 2022*. https://dataindonesia.id/agribisnis-kehutanan/detail/luas-kebun%02sawit-indonesia-hampir-15-juta-hektare-pada-2022

Nasution, S. H., Hanum, C., & Ginting, J. (2014). Growth of Oil Palm Seedlings (Elaeis guineensis Jacq.) In various comparisons of dense decanter planting media and empty oil palm bunches in a single-stage system. *Journal of Agrotechnology Online*, *2*(2), 691–701.

Pahan, I. (2021). *Guide to Oil Palm Cultivation for Plantations*. Independent Spreader.

Pangestu, NHA, & Banowati, G. (2023). Mapping the health of oil palm plantations based on the Normalized Difference Vegetation Index (NDVI) value using Landsat-8 imagery on PT. Potential Utility. *Agriprima: Journal of Applied Agricultural Sciences*, *7*(1), 40–49. https://doi.org/10.25047/agriprima.v7i1.513

Samuel, Prilianti, K. R., Setiawan, H., Mimboro, P., & Correspondence, P. (2022). Oil Palm Plantations Using Convolutional Neural Network (CNN) Model in Information System Software Automatic Tree Detection Method on Oil Palm Plantation Imagery Using Convolutional Neural Network (CNN) Model. *Journal of Information Technology and Computer Science (JTIIK), 9*(7), 1689–1698. https://doi.org/10.25126/jtiik.202296772

Wahyudi, B. (2022). Predict the ranking of apps on Google Play using the Random Forest method. *National Journal of Computer Technology*, *2*(1), 38–47. https://doi.org/10.61306/jnastek.v2i1.25

Wijaya, N. A., & Silalahi, D. D. (2018). Identification of Malnutrition in Oil Palm Plantations of PT. Smart Tbk., using remote sensing. *LET'S, 5*(2).

Wulandari, N. (2020). The use of ndvi (Normalized Difference Vegetation Index) and Savi (Soil Adjusted Vegetation Index) methods to determine the availability of open space.... *ITM Civil Engineering Journal*. http://eprints.itn.ac.id/4597/

Xie, X., Wu, T., Zhu, M., Jiang, G., Xu, Y., Wang, X., & Pu, L. (2021). Comparison of random and multiple linear regression models for estimation of soil extracellular enzyme activity in agricultural reclaimed coastal saltlands. *Ecological Indicators*, *120*, 106925. https://doi.org/10.1016/j.ecolind.2020.106925

Yuniasih, B., & Adjie, ARP (2022). Evaluation of Oil Palm Plantation Conditions Using NDVI Index from Sentinel 2 Satellite Imagery. *Journal of Technocracy*, *16*(2), 127. https://doi.org/10.241.