Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 10 No. 1, April 2025, pp. 41-51
DOI: 10.62411/jais.v10i1.13037

# Classification of Oil Loss Levels in Palm Oil Processing Using Near-Infrared Spectroscopy with Machine Learning

**Muhamad Ilham Fauzan*[1], Jaka Adi Baskara[2], Wahyuningdiah Trisari Harsanti Putri[3], Gatot Tri Pranoto[4]**
Paramadina University, Jl. Raya Mabes Hankam No.Kav 9, Setu, Kec. Cipayung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13880, Telp. (021) 79181188
*Email : muhamad.fauzan@students.paramadina.ac.id[1], jaka.baskara@students.paramadina.ac.id[2], wahyuningdiah.trisari@paramadina.ac.id[3], gatot.pranoto@trilogi.ac.id[4]*
***Corresponding Author**

**Received: 10 June 2025, Revised: 8 July 2025, Accepted: 22 August 2025**

---

**Abstract** - Oil losses in palm oil processing materials, such as Final Effluent, Empty Fruit Bunches, Kernels, Pressed Fiber, and Decanter Solids, pose significant challenges in ensuring production efficiency. FOSS-NIRS technology has been proven capable of quickly and efficiently detecting oil content, but its detection accuracy requires further analytical support. This study aims to develop a machine learning model that can accurately classify FOSS-NIRS data to detect oil losses that are either above the standard (red category) or below the standard (green category). By utilizing FOSS-NIRS data across five material categories, the proposed model is expected to provide precise predictions and support decision-making in palm oil production processes. The results of the study indicate that applying machine learning methods to FOSS-NIRS data can enhance the accuracy of oil loss classification, making it a potential solution for broader implementation in the palm oil processing industry to optimize production efficiency.

**Keywords** - Oil, Palm Oil, Losses, FOSS-NIRS.

## 1. INTRODUCTION

The palm oil processing industry plays a strategic role in the global economy, particularly in major producing countries such as Indonesia. Production efficiency is a key focus, especially in reducing oil losses in processed materials such as final effluent, empty fruit bunches, kernels, pressed fibre, and decanter solids. These oil losses have significant economic implications and affect the sustainability of factory operations [1]. Furthermore, such inefficiencies not only lead to financial waste but also exacerbate the environmental footprint of palm oil processing, contributing to deforestation and habitat loss [2].

FOSS-NIRS (Near-Infrared Spectroscopy) technology has been widely used to detect oil content quickly and non-destructively. However, manual interpretation of FOSS-NIRS data is often inefficient for fast and accurate decision-making. A recent study concludes that while NIRS is capable of measuring oil loss effectively, integrating it with additional data from machine learning algorithms could improve accuracy [1]. This hybrid approach allows for a more nuanced understanding of the variables influencing oil extraction rates, leading to better-informed decisions in plantation management [3].

Machine learning has been proven capable of handling large and complex datasets in the agriculture and food sectors, as well as identifying patterns to produce accurate predictions [3], [4]. Research indicates that by utilizing machine learning frameworks for disease detection, timely interventions can be made that prevent extensive crop damage [5], [6]. For example, combining machine vision with convolutional neural networks has shown considerable promise in detecting early infections of basal stem rot, which is crucial for maintaining yield and overall plantation health [6], [7].

To ensure sustainable palm oil production, there is an increasing need for sophisticated technological solutions that include machine learning for better land use classification and resource allocation. Innovations such as these improve the efficiency and profitability of oil palm plantations and assist in minimizing their environmental impacts by refining plantation management practices through targeted intervention strategies [8], [9].

This study aims to develop a machine learning model to classify the level of oil loss based on FOSS-NIRS data into two categories: above the standard (red) and below the standard (green). This research offers two main contributions: (1) expanding the utilization of FOSS-NIRS technology supported by machine learning in the palm oil industry, and (2) providing a reference for factories to minimize economic losses from oil waste, thereby supporting production efficiency and sustainability.

## 2. RESEARCH METHOD

This study employs an analytical approach using machine learning algorithms to accurately classify data obtained from the FOSS-NIRS machine in detecting oil loss levels that are either above (red category) or below (green category) the standard. Python software is used to process the data and build the classification model, utilizing machine learning libraries such as scikit-learn. The model’s validity is evaluated through accuracy, precision, and recall tests, while the reliability of the instrument is tested using cross-validation by splitting the dataset into training and testing data [5], [10], [11], [12].

The research steps include: literature review, data collection, data determination, data pre-processing, model development, and classification result analysis. The results present data correlation, the developed model, model evaluation, and model implementation. The research flow and stages conducted are illustrated in Figure 1.

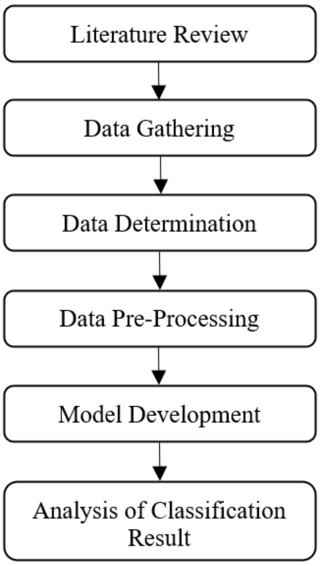


Figure 1. Flow and research stages

### 2.1. Literature Review

Oil losses in the palm oil processing industry represent a significant challenge to production efficiency, particularly concerning materials such as Final Effluent, Empty Fruit Bunches (EFB), Kernels, Pressed Fiber, and Decanter Solids. These losses negatively impact the economic viability of palm oil production and worsen the environmental footprint of the industry. Effective monitoring and reduction of oil losses can be supported through advanced detection technologies combined with analytical machine learning methods.

The palm oil industry has been criticized for its substantial land-use changes, particularly the clearing of peatlands for oil palm cultivation. This transformation contributes to significant greenhouse gas emissions and biodiversity loss [13]. Processing stages susceptible to oil loss include high-pressure steam treatments, where oil may be retained in biomass fractions like EFB, which is known to contain a quantifiable amount of recoverable oil[1].

FOSS-NIRS technology provides a non-destructive means to measure oil content efficiently. This method leverages near-infrared light to determine the composition of materials based on the absorption spectra produced. Research has shown that NIRS can effectively measure oil content across various palm oil processing materials, thus facilitating the monitoring of oil losses in real-time [14]. While FOSS-NIRS shows promise in detecting oil compositions, its accuracy is contingent on the quality of analysis, necessitating further calibration and support via analytical methods [1].

Integrating machine learning with FOSS-NIRS technology can enhance the classification accuracy of oil content in palm oil processing materials. Support Vector Machines (SVM), Random Forest, and Convolutional Neural Networks (CNN) are powerful machine learning algorithms that can uncover patterns in complex data—patterns that might be missed using traditional approaches [5], [15]. Employing machine learning methods could lead to improved classification accuracy of features extracted from FOSS-NIRS data, optimizing the identification of oil loss categories and supporting effective decision-making in palm oil production [16].

The proposed research aims to develop a machine learning model that accurately classifies FOSS-NIRS data into categories indicating acceptable or excessive oil losses. By integrating this technology with machine learning analytics, it is anticipated that the model will provide a more precise and actionable basis for operational decisions in palm oil processing. Studies suggest that intelligent systems driven by machine learning can significantly enhance predictive capabilities and operational efficiency [4], [17].

### 2.2. Data Gathering

The raw data used in this study were obtained through measurements using FOSS-NIRS technology. Data collection was carried out by taking samples from palm oil processing waste or residues such as final effluent, empty fruit bunches, kernels, pressed fiber, and decanter solids [14]. Before being analyzed by FOSS-NIRS, samples were dried if overly wet and then crushed to uniform size to ensure more accurate spectral readings. The samples were then analyzed using FOSS-NIRS technology, producing near-infrared spectra that reflect oil content in a fast and non-destructive manner [1].

### 2.3. Data Determination

The dataset used in this research was collected from 51 palm oil mills on May 23, 2025. The total amount of data was limited to 8,000 entries covering five main categories of palm oil processing materials: final effluent, empty fruit bunches, kernels, pressed fiber, and decanter solids. This data was then processed in the pre-processing stage and subsequently used for the development of the classification model.

### 2.4. *Data Pre-Processing*

The data pre-processing stage aims to improve data quality to ensure optimal performance in the model development process. This stage consists of the following steps:

a. Data Cleaning: Delete rows containing NULL values so as not to interfere with the quality of the analysis[15], [18].
b. Data Transformation: Perform label encoding on category data, and rename columns to make them easier for the algorithm to understand [16], [19].
c. Data Analysis: Perform exploratory data analysis to understand the relationships between features using a correlation matrix visualized through a heatmap.

### 2.5. *Model Development*

At this stage, the classification model is developed using two machine learning algorithms, namely Random Forest and Decision Tree. The processed data is divided into two subsets, namely training data of 60% and testing data of 40%. Both models are trained on the training dataset and then tested on separate test data to evaluate their performance using metrics such as accuracy, precision, recall, and F1-score [5], [10], [11], [12].

### 2.6. *Analysis of Classification Data*

Once the model is built, the classification results are analyzed further to gain insights into the data characteristics and assess which features are most relevant in determining the extent of oil loss. This analysis also includes decision tree visualization to clarify important features that affect the classification results [20].

## 3. RESULTS AND DISCUSSION

### 3.1. Data *Set*

The dataset used in this study was collected from 51 palm oil mills with a limited testing period on Friday, May 23, 2025. Data collection was carried out by taking the latest data first, which was then limited to 8,000 data covering five main categories of processing materials, namely Final Drab, Empty Bunches, Seeds, Press Pulp, and Solid Decanter.

### 3.2. *Data Determination*

The dataset will be processed through a data preprocessing stage. Each entry in the dataset is the result of analysis using FOSS-NIRS, which is then used as input for the classification process using machine learning methods. However, before that, it is necessary to examine and identify the contents of the dataset.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | PKS_CODE | PRODUCT_NAME | OIL_WM | VM | OIL_DM | NOS | RILL_LOSIS_MS | INC_LOSIS_MS |
| 2 | CP03 | Drab Akhir | 0,62 | 97,62 | 26,06 | 1,76 | 0,39 | red |
| 3 | DP13 | Tandan Kosong | 0,76 | 55,14 | 1,68 | 44,1 | 0,17 | green |
| 4 | DP18 | Drab Akhir | 0,47 | 98,13 | 25,3 | 1,39 | 0,27 | green |
| 5 | EP01 | Biji | 0,32 | 12,8 | 0,37 | 86,87 | 0,04 | green |
| 6 | CP09 | Tandan Kosong | 1,11 | 69,86 | 3,68 | 29,03 | 0,16 | orange |
| 7 | CP08 | Ampas Press | 3,57 | 41,77 | 6,12 | 54,67 | 0,46 | green |
| 8 | CP10 | Ampas Press | 5,04 | 34,63 | 7,72 | 60,33 | 0,66 | green |
| 9 | CP09 | Tandan Kosong | 0,5 | 54,13 | 1,09 | 45,37 | 0,07 | orange |
| 10 | CP08 | Ampas Press | 3,98 | 39,52 | 6,59 | 56,5 | 0,52 | green |
| 11 | DP16 | Solid Decanter | 4,16 | 74,92 | 16,57 | 20,93 | 0,11 | red |
| 12 | EP05 | Solid Decanter | 2,4 | 79,31 | 11,59 | 18,29 | 0,07 | green |
| 13 | CP09 | Ampas Press | 3,94 | 37,78 | 6,33 | 58,28 | 0,51 | green |
| 14 | CP01 | Biji | 1,09 | 8,26 | 1,19 | 90,65 | 0,13 | red |
| 15 | AP06 | Ampas Press | 4,44 | 51,91 | 9,23 | 43,65 | 0,62 | green |
| 16 | EP03 | Drab Akhir | 0,52 | 99,31 | 74,67 | 0,18 | 0,31 | green |
| 17 | MP22 | Ampas Press | 4,13 | 31,25 | 6,01 | 64,62 | [NULL] | orange |
| 18 | CP02 | Tandan Kosong | 1,48 | 38,73 | 2,41 | 59,79 | 0,21 | orange |
| 19 | EP12 | Drab Akhir | 0,95 | 95,27 | 19,99 | 3,78 | 0,57 | red |
| 20 | DP13 | Drab Akhir | 1,04 | 91,77 | 12,66 | 7,19 | 0,53 | red |
| 21 | AP06 | Solid Decanter | 3,35 | 78,14 | 15,32 | 18,51 | 0,09 | red |
| 22 | FP04 | Biji | 1,34 | 19,84 | 1,68 | 78,81 | 0,17 | red |
| 23 | EP07 | Solid Decanter | 2,11 | 81,6 | 11,46 | 16,29 | 0,06 | green |
| 24 | EP04 | Biji | 0,28 | 13,85 | 0,33 | 85,87 | 0,04 | green |
| 25 | EP03 | Biji | 0,4 | 13,77 | 0,46 | 85,84 | 0,05 | green |

Figure 2. Dataset of FOSS-NIRS

Based on the data table in Figure 2, at least eight columns or attributes will be used as input for the machine learning classification model. These attributes include:

a. PKS_CODE : The origin factory code indicating where the product was obtained.
b. PRODUCT_NAME : Product name or the name of the palm oil processing material.
c. OIL_WM : Oil content in the material under wet conditions.
d. VM : Volatile content.
e. OIL_DM : Oil content in the material under dry conditions.
f. NOS : Non-oil substance content.
g. RILL_LOSIS_MS : Percentage of oil loss.
h. INC_LISIS_MS : Oil loss category.

### *3.3. Data Pre-Processing*

This section covers the data preparation stage, which consists of data cleaning, data transformation, and data analysis.

*a. Data cleaning*

In this dataset, [NULL] values are present. Therefore, data cleaning is necessary to handle these [NULL] values to improve the overall results. Columns containing [NULL] values will be removed from the dataset. In Jupyter Notebook, a lambda function is used to identify specific columns that contain [NULL] values. Once the corresponding rows are identified, they are deleted. After the process is complete, the number of rows removed will be displayed. The results are shown in Figure 3.

Rows Affected: 649

| | PKS_CODE | PRODUCT_NAME | OIL_WM | VM | OIL_DM | NOS | RILL_LOSIS_MS | INC_LOSIS_MS |
|---|---|---|---|---|---|---|---|---|
| 0 | CP03 | Drab Akhir | 0.62 | 97.62 | 26.06 | 1.76 | 0.39 | red |
| 1 | DP13 | Tandan Kosong | 0.76 | 55.14 | 1.68 | 44.10 | 0.17 | green |
| 2 | DP18 | Drab Akhir | 0.47 | 98.13 | 25.30 | 1.39 | 0.27 | green |
| 3 | EP01 | Biji | 0.32 | 12.80 | 0.37 | 86.87 | 0.04 | green |
| 4 | CP09 | Tandan Kosong | 1.11 | 69.86 | 3.68 | 29.03 | 0.16 | orange |
| 5 | CP08 | Ampas Press | 3.57 | 41.77 | 6.12 | 54.67 | 0.46 | green |
| 6 | CP10 | Ampas Press | 5.04 | 34.63 | 7.72 | 60.33 | 0.66 | green |
| 7 | CP09 | Tandan Kosong | 0.50 | 54.13 | 1.09 | 45.37 | 0.07 | orange |
| 8 | CP08 | Ampas Press | 3.98 | 39.52 | 6.59 | 56.50 | 0.52 | green |
| 9 | DP16 | Solid Decanter | 4.16 | 74.92 | 16.57 | 20.93 | 0.11 | red |
| 10 | EP05 | Solid Decanter | 2.40 | 79.31 | 11.59 | 18.29 | 0.07 | green |
| 11 | CP09 | Ampas Press | 3.94 | 37.78 | 6.33 | 58.28 | 0.51 | green |
| 12 | CP01 | Biji | 1.09 | 8.26 | 1.19 | 90.65 | 0.13 | red |
| 13 | AP06 | Ampas Press | 4.44 | 51.91 | 9.23 | 43.65 | 0.62 | green |
| 14 | EP03 | Drab Akhir | 0.52 | 99.31 | 74.67 | 0.18 | 0.31 | green |
| 16 | CP02 | Tandan Kosong | 1.48 | 38.73 | 2.41 | 59.79 | 0.21 | orange |
| 17 | EP12 | Drab Akhir | 0.95 | 95.27 | 19.99 | 3.78 | 0.57 | red |
| 18 | DP13 | Drab Akhir | 1.04 | 91.77 | 12.66 | 7.19 | 0.53 | red |
| 19 | AP06 | Solid Decanter | 3.35 | 78.14 | 15.32 | 18.51 | 0.09 | red |
| 20 | FP04 | Biji | 1.34 | 19.84 | 1.68 | 78.81 | 0.17 | red |
| 21 | EP07 | Solid Decanter | 2.11 | 81.60 | 11.46 | 16.29 | 0.06 | green |
| 22 | EP04 | Biji | 0.28 | 13.85 | 0.33 | 85.87 | 0.04 | green |
| 23 | EP03 | Biji | 0.40 | 13.77 | 0.46 | 85.84 | 0.05 | green |
| 24 | DP16 | Solid Decanter | 4.07 | 75.58 | 16.68 | 20.34 | 0.11 | red |
| 25 | CP08 | Solid Decanter | 2.41 | 80.48 | 12.33 | 17.12 | 0.06 | green |

Figure 3. Results of Data Cleaning

*b. Data Transformation*

The next stage involves renaming the columns to make them more relevant to the problem context and easier for the machine learning model to interpret. Label encoding is then applied to the categorical columns, converting them into numerical values so they can be used in further analysis. The results can be seen in Figure 4.

```python
#data transformations
categorical_columns = [
    'PKS_CODE',
    'PRODUCT_NAME',
    'INC_LOSIS_MS'
]

for col in categorical_columns:
    df[col+'_LABELED'] = label_encoder.fit_transform(df[col])

df.head()
```

✓ 0.0s Python

| | PKS_CODE | PRODUCT_NAME | OIL_WM | VM | OIL_DM | NOS | RILL_LOSIS_MS | INC_LOSIS_MS | PKS_CODE_LABELED | PRODUCT_NAME_LABELED | INC_LOSIS_MS_LABELED |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | CP03 | Drab Akhir | 0.62 | 97.62 | 26.06 | 1.76 | 0.39 | red | 5 | 2 | 2 |
| 1 | DP13 | Tandan Kosong | 0.76 | 55.14 | 1.68 | 44.10 | 0.17 | green | 19 | 4 | 0 |
| 2 | DP18 | Drab Akhir | 0.47 | 98.13 | 25.30 | 1.39 | 0.27 | green | 24 | 2 | 0 |
| 3 | EP01 | Biji | 0.32 | 12.80 | 0.37 | 86.87 | 0.04 | green | 25 | 1 | 0 |
| 4 | CP09 | Tandan Kosong | 1.11 | 69.86 | 3.68 | 29.03 | 0.16 | orange | 10 | 4 | 1 |

Figure 4. Results of Label Encoding

After performing label encoding, the next data transformation step is to rename the machine-generated column names to more readable and user-friendly labels.

- PKS_CODE → FactoryCode
- PRODUCT_NAME → ProductName
- OIL_WM → OilWetMatter
- VM → Volatile Matter
- OIL_DM → OilDryMatter
- NOS → NonOilSolids
- RILL_LOSIS_MS → RillLosis
- INC_LOSIS_MS → Category

Refer to Figure 5 to see the results of the label encoding process.

```python
#Mengambil kolom tertentu
selected_columns = [
    'PKS_CODE_LABELED',
    'PRODUCT_NAME_LABELED',
    'OIL_WM',
    'VM',
    'OIL_DM',
    'NOS',
    'RILL_LOSIS_MS',
    'INC_LOSIS_MS_LABELED',
]

data = df[selected_columns].copy()

#perubahan alias nama, nama asal dari mesin
new_column_names = {
    'PKS_CODE_LABELED' : 'FactoryCode',
    'PRODUCT_NAME_LABELED':'ProductName',
    'OIL_WM': 'OilWetMatter',
    'VM': 'VolatileMatter',
    'OIL_DM': 'OilDryMatter',
    'NOS': 'NonOilSolids',
    'RILL_LOSIS_MS': 'RillLosis',
    'INC_LOSIS_MS_LABELED': 'Category'
}

data = data.rename(columns=new_column_names)
data.head()
```

✓ 0.0s Python

| | FactoryCode | ProductName | OilWetMatter | VolatileMatter | OilDryMatter | NonOilSolids | RillLosis | Category |
|---|---|---|---|---|---|---|---|---|
| 0 | 5 | 2 | 0.62 | 97.62 | 26.06 | 1.76 | 0.39 | 2 |
| 1 | 19 | 4 | 0.76 | 55.14 | 1.68 | 44.10 | 0.17 | 0 |
| 2 | 24 | 2 | 0.47 | 98.13 | 25.30 | 1.39 | 0.27 | 0 |
| 3 | 25 | 1 | 0.32 | 12.80 | 0.37 | 86.87 | 0.04 | 0 |
| 4 | 10 | 4 | 1.11 | 69.86 | 3.68 | 29.03 | 0.16 | 1 |

Figure 5. Results of Rename Labels

## c. *Data Analysis*

The data is analyzed to understand the relationships between each pair of columns and to identify which columns have strong or weak correlations. A correlation matrix is computed and displayed from the DataFrame, and visualized using a heatmap. This is done to gain insight into the relationships among the variables in the DataFrame by calculating and visualizing the correlation matrix [4]. Using the matplotlib and seaborn libraries available in the Python programming language, the correlation matrix for each pair of columns is generated. This matrix indicates the strength of the relationships between column pairs and is presented in the form of a heatmap, as shown in Figure 6.

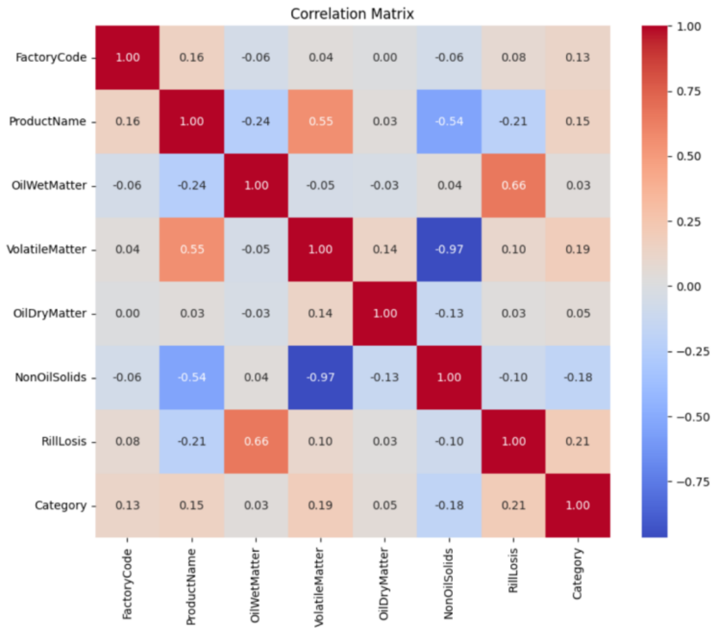


Figure 6. Correlation Matrix of DataFrame

The correlation matrix reveals significant relationships between several features, such as free oil content and the level of oil loss. Features that show a high correlation with the target variable are retained for the model training process.

### *3.4. Model Development*

At this stage, the study employs two primary algorithms—Random Forest and Decision Tree—to build a classification model for determining the level of palm oil loss using data obtained from FOSS-NIRS analysis. The model development process is carried out using Python programming language with the support of the scikit-learn machine learning library.

First, the preprocessed data is divided into two subsets: 60% is used as training data and 40% as testing data.

The Random Forest model is developed by utilizing multiple decision trees simultaneously. Each decision tree within the Random Forest is built independently using different subsets of the data, and their predictions are then combined to produce the final result.

The use of multiple decision trees aims to reduce the risk of overfitting, improve stability, and produce more accurate predictions. Meanwhile, the Decision Tree model is constructed using a single decision tree that explicitly maps classification rules based on the most relevant features. One of the key advantages of the Decision Tree model is its interpretability, which allows for a deeper analysis of the most influential features in the classification of oil loss.

```python
features = ['FactoryCode','ProductName','OilWetMatter','VolatileMatter','OilDryMatter','NonOilSolids', 'RillLosis']
target = 'Category'

# Splitting the data into training and testing sets

X = data[features]
Y = data[target]

localOptima = None
localOptimaScore = 0.0
optModel = None

trainX, testX, trainY, testY = train_test_split(X, Y, random_state=42, test_size=0.4)
for modelName in models.keys():
    accuracy, report, model = classifier(trainX, trainY, testX, testY, modelName)
    models[modelName] = model
    if(accuracy > localOptimaScore):
        localOptimaScore = accuracy
        localOptima = modelName
        optModel = model

print('Most Optimal ' + localOptima)
print('With Score ' + str(localOptimaScore))
```

```
✓ 0.1s                                                                    Python
Result model decision_tree 0.9887793267596056
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      1999
           1       1.00      1.00      1.00        98
           2       0.98      0.98      0.98       844

    accuracy                           0.99      2941
   macro avg       0.99      0.99      0.99      2941
weighted avg       0.99      0.99      0.99      2941


Result model random_forest 0.9901394083645019
              precision    recall  f1-score   support

           0       0.99      1.00      0.99      1999
           1       0.99      0.97      0.98        98
           2       0.99      0.98      0.98       844

    accuracy                           0.99      2941
   macro avg       0.99      0.98      0.99      2941
weighted avg       0.99      0.99      0.99      2941


Most Optimal random_forest
With Score 0.9901394083645019
```

Figure 7. Results of Training and Testing data

The two trained models are then evaluated using the testing data subset, with evaluation metrics including accuracy, precision, recall, and F1-score. The evaluation results are presented in Table 1 below:

Tabel 1. Table of Evaluation Model

| Algoritma | Akurasi (%) | Presisi (%) | Recall (%) | F1-Score (%) |
| --- | --- | --- | --- | --- |
| Decision Tree | 0.99 | 0.99 | 0.99 | 0.988 |
| Random Forest | 0.99 | 0.99 | 0.98 | 0.989 |

From the table above, it can be observed that the Random Forest model delivers the best performance, achieving an accuracy of 98.9%, while the Decision Tree model yields an accuracy of 98.8%.

### *3.5. Analysis of Classification Data*

After understanding the model development results, it is important to conduct a more in-depth analysis of the obtained classification outcomes [16], [20], [21]. This aims to identify the key characteristics of the data that influence classification predictions and to understand the practical implications of the results.

In the classification results analysis stage, the decision trees generated by each model are visualized. From these visualizations, it can be seen that both models are able to clearly classify the levels of oil loss into distinct categories, namely high loss (above the standard) and low loss (below the standard).

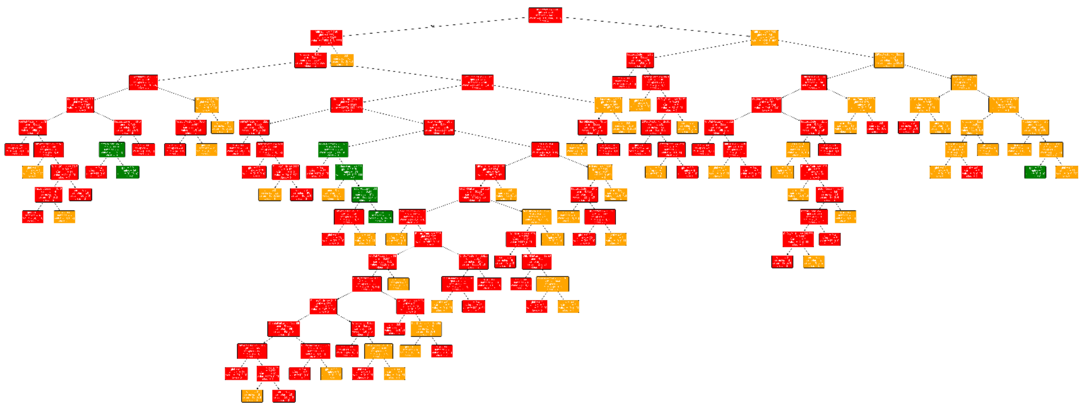

Figure 8. Plot Tree of Decision Tree (URL)

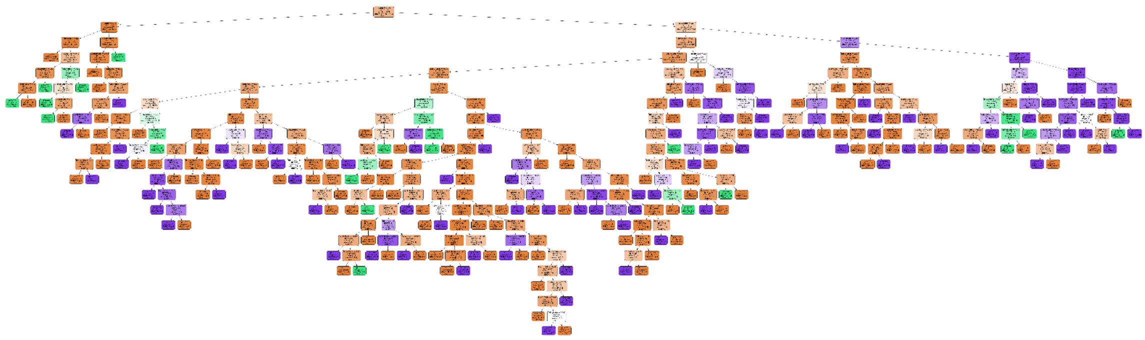

Figure 9. Plot Tree of Random Forest (URL)

The decision tree visualization from the Decision Tree model displays simple and easily interpretable classification rules, allowing for the identification of key features such as **RillLosis**, **OilDryMatter**, and **Volatile Matter** as the main determinants of classification. In contrast, the decision tree visualization in the Random Forest model reveals a more complex structure due to the involvement of multiple decision trees. This complexity contributes to greater stability and higher prediction accuracy compared to a single tree.

This analysis also reveals that the high loss category (red) frequently appears in specific parts of the production process, indicating the need for targeted actions to reduce such losses. Therefore, the classification results are not only useful for prediction but can also serve as a guideline for optimizing the production process.

## 4. CONCLUSION

This study successfully developed machine learning models based on the Random Forest and Decision Tree algorithms to classify the level of oil loss in the palm oil processing process using FOSS-NIRS data. The main findings indicate that the Random Forest algorithm achieved the best performance, with an accuracy of 98.9%, slightly higher than the Decision Tree algorithm, which achieved an accuracy of 98.8%. These results highlight the effectiveness of Random Forest in handling complex datasets, particularly those with numerical features that show significant correlation with oil loss levels.

This study addresses the main objective stated in the Introduction, namely how the application of machine learning algorithms can improve the accuracy of FOSS-NIRS data classification [3], [10], [22], [23]. The correlation analysis conducted revealed significant relationships among several features in the dataset, supporting the selection of relevant features to enhance model performance. These results are crucial for supporting production efficiency in the palm oil processing industry through faster and more accurate detection of oil loss.

This study makes a significant contribution to the application of machine learning technology in the palm oil processing industry, particularly in leveraging FOSS-NIRS data for production process optimization. The findings of this research may also serve as a reference for the development of automated systems for data classification in similar industrial contexts.

Suggestions for Future Research:

a. Dataset Expansion: Future research is expected to expand the dataset by increasing the number and diversity of samples to improve the generalizability of the model.
b. Exploration of Other Algorithms: Other algorithms such as Support Vector Machine (SVM), Gradient Boosting, or Neural Networks can be explored to gain additional insights into model performance.

***REFERENCES***


[1] I. Irfan, I. F. Aji, and D. Yunita, "APPLICATION OF NEAR INFRARED SPECTROSCOPY (NIRS) IN THE MEASUREMENT OF OIL LOSS AND ITS ACCURACY COMPARED TO SOXHLET METHOD," *Jurnal Agroindustri*, 2023, doi: https://doi.org/10.31186/jagroindustri.13.2.107-120.

[2] I. Numata, A. J. Elmore, M. A. Cochrane, and ..., "Deforestation, plantation-related land cover dynamics and oil palm age-structure change during 1990–2020 in Riau Province, Indonesia," *Environmental …*, 2022, doi: 10.1088/1748-9326/ac8a61.

[3] N. A. Suharyanti, E. Sinaga, F. Ersavan, and ..., "The usage of machine learning approach in predicting oil extraction rate based on yield making, yield taking and mill processing factors on oil palm plantations in East …," *IOP Conference Series …*, 2025, doi: 10.1088/1755-1315/1477/1/012055.

[4] I. Ahmed and P. K. Yadav, "Plant disease detection using machine learning approaches," *Expert Syst*, 2023, doi: 10.1111/exsy.13136.

[5] L. Z. Yong, S. Khairunniza-Bejo, M. Jahari, and ..., "Automatic detection of an early stage of basal stem rot disease infection using VGG16 and mask R-CNN," *… Series: Earth and …*, 2023, doi: 10.1088/1755-1315/1133/1/012076.

[6] S. H. Wan, J. C. E. Yong, E. H. Y. Leong, and J. Y. Chan, "Development of an Oil Palm Basal Stem Rot Disease Detection Model Via Machine Vision with Optimized Inception-Based Convolutional Neural Network," *… : Select Proceedings of …*, 2022, doi: 10.1007/978-981-19-3053-9_7.

[7] A. Y. Khaled, S. A. Aziz, S. K. Bejo, N. M. Nawi, and ..., "Artificial intelligence for spectral classification to identify the basal stem rot disease in oil palm using dielectric spectroscopy measurements," *Tropical Plant …*, 2022, doi: 10.1007/s40858-021-00445-1.

[8] M. S. Ozigis, S. Wich, A. Descals, and ..., "Mapping oil palm plantations and their implications on forest and great ape habitat loss in Central Africa," *Remote Sensing in …*, 2024, doi: 10.1002/rse2.428.

[9] M. Wagner, E. A. Wentz, and ..., "Quantifying oil palm expansion in Southeast Asia from 2000 to 2015: A data fusion approach," *… of Land Use Science*, 2022, doi: 10.1080/1747423X.2021.2020918.

[10] Z. H. Feng, L. Y. Wang, Z. Q. Yang, Y. Y. Zhang, X. Li, and ..., "Hyperspectral monitoring of powdery mildew disease severity in wheat based on machine learning," *Frontiers in Plant …*, 2022, doi: 10.3389/fpls.2022.828454.

[11] X. Cheng, A. Doosthosseini, and J. Kunkel, “Improve the deep learning models in forestry based on explanations and expertise,” *Front Plant Sci*, 2022, doi: 10.3389/fpls.2022.902105.
[12] J. Pišl, M. Rußwurm, L. H. Hughes, and ..., “Mapping drivers of tropical forest loss with satellite image time series and machine learning,” *Environmental …*, 2024, doi: 10.1088/1748-9326/ad44b2.
[13] D. Fiantis, Rudiyanto, F. I. Ginting, and ..., “Mapping peat thickness and carbon stock of a degraded peatland in West Sumatra, Indonesia,” *Soil Use and …*, 2024, doi: 10.1111/sum.12954.
[14] P. U. Thandapani, Z. Harahap, and ..., “Usage of near infrared spectrometer as an analyzing tool for nutrients in leaf, fertilizer and soil in oil palm industry,” *IOP Conference Series …*, 2024, doi: 10.1088/1755-1315/1308/1/012036.
[15] Y. P. Wang, N. H. Idris, F. M. Muharam, N. Asib, and ..., “Comparison of different variable selection methods for predicting the occurrence of Metisa Plana in oil palm plantation using machine learning,” *IOP Conference Series …*, 2023, doi: 10.1088/1755-1315/1274/1/012008.
[16] M. Arakeri, M. P. Dhatvik, A. V Kavan, and ..., “Intelligent pesticide recommendation system for cocoa plant using computer vision and deep learning techniques,” *Environmental …*, 2024, doi: 10.1088/2515-7620/ad58ae.
[17] P. Baltezar, P. J. Murillo-Sandoval, and ..., “A regional map of mangrove extent for Myanmar, Thailand, and Cambodia shows losses of 44% by 1996,” *Frontiers in Marine …*, 2023, doi: 10.3389/fmars.2023.1127720.
[18] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, and ..., “Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset,” *SN computer …*, 2021, doi: 10.1007/s42979-020-00394-7.
[19] R. Ashmore, R. Calinescu, and C. Paterson, “Assuring the machine learning lifecycle: Desiderata, methods, and challenges,” *ACM Computing Surveys (CSUR)*, 2021, doi: 10.1145/3453444.
[20] A. Bouguettaya, H. Zarzour, A. Kechida, and ..., “Deep learning techniques to classify agricultural crops through UAV imagery: A review,” *Neural computing and …*, 2022, doi: 10.1007/s00521-022-07104-9.
[21] A. Mirelman, M. B. O. Frank, M. Melamed, and ..., “Detecting sensitive mobility features for Parkinson’s disease stages via machine learning,” *Movement …*, 2021, doi: 10.1002/mds.28631.
[22] C. Chathurya, D. Sachdeva, and ..., “Chili Leaf Classification Using Deep Learning Techniques,” *Deep Learning Techniques …*, 2024, doi: 10.1002/9781394234271.ch2.
[23] C. Subbarayudu and M. Kubendiran, “A Comprehensive Survey on Machine Learning and Deep Learning Techniques for Crop Disease Prediction in Smart Agriculture.,” *Nature Environment \&Pollution …*, 2024, [Online]. Available: https://neptjournal.com/upload-images/(3)B-4091.pdf