BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

Prediction of Palm Oil Fresh Fruit Bunch Yield using Support Vector Machine (SVM)

Try Widyawanti*, M Fakhriza

Faculty of Science and Technology, Department of Computer Science, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1,*]trywidyawanti@gmail.com, [2]fakhriza@uinsu.ac.id
Correspondence Author Email: trywidyawanti@gmail.com

Abstract−Palm oil fresh fruit bunch (FFB) production plays a crucial role in plantation management and decision making. However, fluctuations in environmental conditions and plantation characteristics often make yield estimation difficult to perform accurately. This study aims to predict palm oil fresh fruit bunch yield using the Support Vector Machine (SVM) algorithm as a machine learning–based approach. The dataset used in this research consists of monthly production data from 2020 to 2024, including several influential variables such as plant age, land area, rainfall, and soil characteristics. The data were preprocessed through cleaning, transformation, and normalization using the min–max scaling method to ensure consistency and stability during model training. The SVM model was implemented using the Radial Basis Function (RBF) kernel, which is suitable for handling nonlinear data patterns. Model evaluation was conducted by dividing the dataset into training and testing data with a ratio of 80% and 20%, respectively. The performance of the proposed model was measured using Root Mean Square Error (RMSE) and accuracy metrics. Experimental results show that the SVM model achieved an RMSE value of 1.316561 and an accuracy rate of 56.6%, indicating that the model is able to capture the general pattern of palm oil FFB yield data with a relatively small prediction error. Although the accuracy obtained is moderate, the results demonstrate that SVM can be applied as an initial predictive tool for estimating palm oil yield. The findings of this study are expected to support plantation managers in planning harvest activities and optimizing resource allocation.

Keywords: Palm Oil; Fresh Fruit Bunch Yield; Support Vector Machine; Machine Learning; Agricultural Prediction

1. INTRODUCTION

Oil palm (Elaeis guineensis) is a strategic commodity in Indonesia, contributing substantially to national income and rural employment, and playing a critical role in global vegetable oil supply. Fluctuations in fresh fruit bunch (FFB) yield present major challenges for plantation management, as unpredictable production levels can impede effective planning and resource allocation[1]. Sustainable yield prediction models can help mitigate these challenges by enabling better forecasting of harvest outcomes based on historical and environmental factors. Accurate prediction systems are essential to support decision-making processes that optimize harvest schedules and improve economic outcomes for plantation stakeholders[2].

Several factors influence palm oil yield, including plant age, land area, rainfall patterns, and soil properties. Integrating these agronomic and environmental variables into a predictive model improves the capacity to estimate production trends accurately[3]. In data-driven agriculture, machine learning approaches have increasingly been adopted because they can capture complex, non-linear relationships in agricultural datasets that traditional statistical models often fail to detect. Previous parts of this study describe the collection and preprocessing of monthly production data from 2020 to 2024, including normalization and feature transformation to prepare the dataset for analysis[4].

Support Vector Machine (SVM) is a supervised learning algorithm widely applied in classification and regression tasks due to its robustness in high-dimensional spaces and ability to handle nonlinear patterns via kernel functions[5]. In the context of agricultural forecasting, SVM and its variants have shown promise in modeling relationships between predictor variables and crop yield[6]. The present research implements an SVM model with a radial basis function (RBF) kernel to estimate FFB yield, offering a balance between simplicity and predictive effectiveness within the available dataset[7].

Moreover, yield prediction is distinguished from simple trend analysis by its focus on learning predictive patterns that generalize to unseen data[8][9]. The present study adopts an 80:20 training–testing split to assess model performance, measuring accuracy and root mean square error (RMSE) to quantify predictive accuracy[10][11]. The empirical results demonstrate measurable success in capturing key patterns in the dataset, suggesting the model’s practical utility for preliminary forecasting tasks in plantation settings[12][13].

Despite advances in modeling approaches, there remains a need for context-specific research that reflects Indonesia’s agronomic conditions and data realities[13][14]. While general machine learning literature reveals strong potential for crop yield forecasting internationally, localized evidence grounded in Indonesian plantation data is comparatively sparse[15][16]. Research that reflects national agribusiness nuances is critical for improving operational relevance and adoption[17].

For instance, Ardian Pramana Putra and Harahap developed a palm oil productivity forecasting system using a Long Short-Term Memory (LSTM) model and web-based visualization, demonstrating its application to plantation data with multi-year patterns in productivity at PTPN IV Kebun Bah Birung Ulu[18]. Similarly, Sartika Ayunda applied a Double Moving Average method to forecast palm oil production in Riau Province, achieving high predictive accuracy with MAPE under 10% for 2023 estimates. These national studies validate the active interest in predictive modeling for palm oil yield and underscore the relevance of developing improved machine learning models tailored to Indonesian plantation contexts [19].

ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

However, existing research reveals limitations such as narrow focus on single algorithms, reliance on traditional forecasting techniques, or lack of comprehensive agronomic variables in modeling frameworks. Few studies systematically leverage nonlinear supervised learning models like SVM within full agronomic contexts, particularly using recent multi-year plantation datasets. This gap highlights the need for research that combines advanced machine learning techniques with comprehensive feature integration to enhance prediction accuracy and practical usability.

Therefore, this study aims to develop and evaluate an SVM-based model for predicting palm oil FFB yield using multiple agronomic and environmental features. By quantitatively assessing model performance through RMSE and accuracy metrics, the research seeks to contribute an effective predictive tool for operational planning and resource allocation in palm oil plantations..

2. RESEARCH METHODOLOGY

This study adopts a structured research methodology to develop a machine learning–based model for predicting palm oil fresh fruit bunch yield using the Support Vector Machine algorithm. The overall research process includes data input, preprocessing, model implementation, and performance evaluation. Agronomic and environmental variables are incorporated to represent plantation conditions. The complete sequence of system operations and data processing stages is illustrated in Figure 1, which presents the system flowchart of the proposed prediction model.

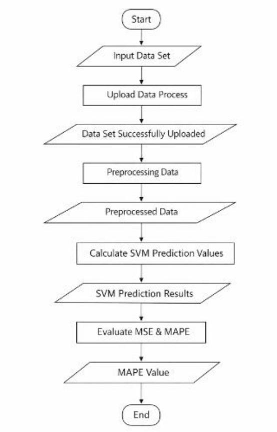

Figure 1. Framework of the Research

2.1 Dataset and Data Collection

This study uses a secondary dataset obtained from official plantation records to develop a prediction model for palm oil fresh fruit bunch yield. The dataset represents real production conditions and is structured to capture temporal variations in plantation output. Data collection focuses on integrating agronomic and environmental variables that influence yield performance, enabling the model to learn patterns based on historical plantation data.

Table 1. Palm Oil Plantation Production Data of PTPN IV Adolina (2020–2024)

No.	Year	Palm Oil Plantation Production (Tons)
1	2020	21,968.25
2	2021	24,826.63
3	2022	25,192.56
4	2023	24,159.69
5	2024	23,783.26

The complete dataset consists of monthly time-based records from 2020 to 2024, resulting in 60 observations. Each record contains several input variables that represent plantation and environmental conditions. The input variables include plant age, which reflects the maturity level of oil palm trees; land area, which indicates the size of the plantation under cultivation; rainfall, representing climatic conditions that directly affect crop productivity; and soil characteristics, which describe the physical and chemical properties of the soil. The target variable in this study is fresh fruit bunch yield, measured in tons, which serves as the output for model prediction [20].

To evaluate model performance objectively, the dataset is divided into training and testing subsets using an 80:20 ratio. The training data are used to build and train the Support Vector Machine model, while the testing data are reserved for performance evaluation. This data partitioning strategy allows the model to be assessed on unseen data, ensuring that the prediction results reflect the model's generalization capability rather than memorization of historical values.

2.2 Data Preprocessing

Data preprocessing is a crucial stage to ensure that the dataset is suitable for machine learning modeling and capable of producing accurate prediction results. Raw plantation data often contain inconsistencies, different measurement scales, and attributes that are not directly usable in computational models. Therefore, preprocessing is applied to improve data quality, standardize input values, and enhance the performance of the prediction model. The preprocessing steps implemented in this study are described as follows:

a. Data Cleaning
Incomplete records and missing values are identified and removed to ensure that each observation contains complete information for all variables. This step helps eliminate noise and prevents bias during model training.

b. Data Transformation
Categorical and temporal attributes are transformed into numerical values to enable mathematical processing. This transformation allows all variables to be represented in a format compatible with the Support Vector Machine algorithm.

c. Data Normalization
Normalization is conducted using the min–max scaling method to convert all variable values into a common range between 0 and 1. This process ensures that no single variable dominates the learning process due to scale differences and improves model stability.

d. Final Dataset Preparation
After normalization, the dataset is reviewed to confirm consistency and readiness for the training and testing stages. The resulting dataset serves as standardized input for model development.

These preprocessing procedures ensure that the dataset is clean, structured, and suitable for predictive modeling. As a result, the Support Vector Machine model can learn relationships among variables more effectively and generate reliable yield predictions.

2.3 Support Vector Machine Model

Support Vector Machine is employed in this study as the core prediction algorithm to model the relationship between agronomic variables and palm oil fresh fruit bunch yield. SVM is well suited for handling nonlinear data patterns by transforming input data into a higher-dimensional feature space using kernel functions. This capability allows the model to identify an optimal hyperplane that minimizes prediction error while maintaining good generalization performance. In this research, SVM is implemented to process normalized input data and generate yield predictions based on learned patterns. The prediction function of the Support Vector Machine model is defined as :

$$f(x) = \sum i = 1^n (a_i - a_i) K(X_i, X) + b \tag{1}$$

where f(x) represents the predicted output, x is the new input data, x_i denotes the training data, $(\alpha_i - \alpha_i)^*$ are the Lagrange multipliers obtained during the training process, $K(x_i, x)$ is the kernel function, and b is the bias term. This study uses the Radial Basis Function kernel, which is expressed as :

$$K(X_i, X_j) = \exp\left(-y\left|\left|x_i - x_j\right|\right|^2\right) \tag{2}$$

where γ is the kernel parameter that controls the influence of each training sample, and $\|x_i - x_j\|^2$ represents the squared Euclidean distance between data points. The RBF kernel is selected because of its effectiveness in capturing complex nonlinear relationships between input variables and yield output. By adjusting kernel parameters, the SVM model can balance prediction accuracy and model flexibility. After training, the model applies the learned function to unseen data to estimate fresh fruit bunch yield. This approach enables the prediction model to represent nonlinear interactions among plant age, land area, rainfall, and soil characteristics in a structured and computationally efficient manner.

2.4 Model Evaluation

Model evaluation is conducted to measure the performance of the Support Vector Machine model in predicting palm oil fresh fruit bunch yield. Evaluation metrics are used to quantify the accuracy and reliability of the prediction results when applied to testing data. In this study, model performance is assessed using Root Mean Square Error (RMSE) and accuracy, which provide complementary perspectives on prediction quality.

RMSE is used to measure the average magnitude of prediction errors by calculating the square root of the mean squared difference between predicted and actual values. A lower RMSE value indicates that the model predictions are closer to the actual production values, reflecting better predictive accuracy. Accuracy is employed to evaluate the proportion of correct predictions generated by the model relative to the total number of test samples. The use of these evaluation metrics allows the model's performance to be objectively assessed and provides a clear basis for analyzing its effectiveness in capturing yield patterns and supporting decision making in palm oil plantation management.

3. RESULT AND DISCUSSION

This section presents the experimental results and discusses the performance of the proposed Support Vector Machine model in predicting palm oil fresh fruit bunch yield categories. The discussion focuses on the outcomes obtained from data preprocessing, model training, prediction, and evaluation stages. Each result is analyzed to assess the effectiveness of the model and to identify patterns, strengths, and limitations based on the dataset used. The presentation of results is organized systematically to ensure clarity and to support an objective interpretation of the model's predictive capability.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

3.1 Experimental Dataset and Setup

This study uses a monthly dataset collected from PTPN IV Adolina to evaluate the proposed Support Vector Machine model under real plantation conditions. The data span five years from 2020 to 2024 and capture temporal variations in fresh fruit bunch yield across multiple production cycles. Each observation includes plant age, land area, rainfall, and soil pH as input variables, while monthly yield serves as the target variable. These variables reflect key agronomic and environmental factors that influence productivity and are obtained from official plantation records to ensure reliability. The dataset structure and variable variation are presented in Table 2.

Table 2. Dataset of PTPN IV Adolina (Excerpt)

No	Month	Year	Plant Age (Year)	Land Area (Ha)	Rainfall (mm)	Soil Type (pH)	Yield Production (ton)
1	January	2020	7	6,156	155.4	4.5	8,374,810
2	February	2020	7	6,156	48.9	4.5	8,873,880
3	March	2020	7	6,156	32.7	4.5	9,447,950
4	April	2020	7	6,156	173.9	4.8	10,735,720
5	May	2020	7	6,156	207.7	5.6	11,088,240
6	June	2020	7	6,156	194.6	5.7	11,599,310
7	July	2020	7	6,156	155.6	5.8	12,749,830
8	August	2020	7	6,156	202.7	6.3	13,335,610
9	September	2020	7	6,156	345.3	5.8	12,342,910
10	October	2020	7	6,156	176.7	5.5	11,839,890
…	…	…	…	…	…	…	…
55	July	2024	11	5,879	265.8	6.3	15,829,130
56	August	2024	11	5,879	343.7	5.8	14,908,610
57	September	2024	11	5,879	235.6	5.5	12,818,020
58	October	2024	11	5,879	203.5	5.5	10,495,890
59	November	2024	11	5,879	356.5	5.5	10,101,330

The complete dataset consists of 60 monthly observations covering the period from 2020 to 2024. The input variables include plant age, land area, rainfall, and soil characteristics, while fresh fruit bunch yield serves as the target variable. For experimental evaluation, the dataset is divided into training and testing subsets to assess the model's ability to generalize to unseen data. This setup ensures that the reported results reflect the predictive performance of the model under realistic plantation conditions rather than memorization of historical values.

3.2 Data Preprocessing Results

Data preprocessing is a critical stage to ensure that all variables used in the model have comparable scales and are suitable for classification using the Support Vector Machine algorithm. This stage focuses on transforming the dataset into a standardized numerical range to reduce bias caused by differences in measurement units and value magnitudes. Min–Max normalization is applied to all input variables to scale the data into a uniform range between 0 and 1.

Table 3. Minimum and Maximum Values

Parameter	C1	C2	C3	C4	C5
Min	11	6,156	596.7	6.3	15,950,370
Max	7	5,561	20.7	4.5	4,859,220

Table 3 shows the minimum and maximum values for each variable prior to normalization. The plant age variable ranges from 7 to 11 years, reflecting differences in plantation maturity across the observation period. Land area values vary between 5,561 and 6,156 hectares, indicating changes in plantation coverage over time. Rainfall exhibits a wide range, from 20.7 mm to 596.7 mm, highlighting significant variability in climatic conditions. Soil pH values range between 4.5 and 6.3, representing acidic to moderately acidic soil conditions. Yield production shows the largest numerical range, from 4,859,220 tons to 15,950,370 tons, which justifies the need for normalization to reduce scale imbalance among variables.

Table 4. Normalized Dataset

No	Plant Age	Land Area	Rainfall	Soil Type (pH)	Yield Production
1	0	1	0.233854167	0	0.316972541
2	0	1	0.048958333	0	0.361969679
3	0	1	0.020833333	0	0.413728964
4	0	1	0.265972222	0.166666667	0.529836852
5	0	1	0.324652778	0.611111111	0.561620752
6	0	1	0.301909722	0.666666667	0.607699833
7	0	1	0.234201389	0.722222222	0.711432989

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

No	Plant Age	Land Area	Rainfall	Soil Type (pH)	Yield Production
8	0	1	0.315972222	1	0.764248072
9	0	1	0.563541667	0.722222222	0.674744278
10	0	1	0.270833333	0.555555556	0.629391001
…	…	…	…	…	…
55	1	0.4	0.424652778	1	0.951712812
56	1	0.4	0.563541667	0.722222222	0.894132237
57	1	0.4	0.369097222	0.555555556	0.612707265
58	1	0.4	0.347222222	0.555555556	0.41896493
59	1	0.4	0.608680556	0.555555556	0.396486596
60	1	0.4	0.221527778	0.555555556	0.361147084

Table 4 presents representative samples of the normalized dataset after applying Min–Max scaling. The normalization process successfully maps all variables into the range of 0 to 1. For example, plant age values progress from 0 to 1, indicating increasing plantation maturity across years. Rainfall values vary continuously within the normalized range, reflecting preserved variability after scaling. Yield production values also span from low normalized values near 0 to high values approaching 1, ensuring that production differences remain distinguishable without dominating other features. This transformation confirms that the dataset is numerically stable and suitable for SVM-based classification.

3.3 Labeling and SVM Training Results

This subsection presents the results of the labeling process and the training phase of the Support Vector Machine model. Labeling is applied to convert the normalized yield values into discrete classes based on a predefined threshold, enabling the formulation of a classification problem. After labeling, the SVM model is trained using a Gaussian kernel to capture nonlinear relationships among the input variables. The results presented in this section demonstrate how the dataset is categorized and how similarity between data points is computed during the training process.

Table 5. Dataset Classification (Excerpt).

No	Plant Age	Land Area	Rainfall	Soil Type (pH)	Yield Production	Yield Label
1	0	1	0.233854167	0	0.316972541	-1
2	0	1	0.048958333	0	0.361969679	-1
3	0	1	0.020833333	0	0.413728964	-1
4	0	1	0.265972222	0.166666667	0.529836852	-1
5	0	1	0.324652778	0.611111111	0.561620752	-1
6	0	1	0.301909722	0.666666667	0.607699833	-1
7	0	1	0.234201389	0.722222222	0.711432989	1
8	0	1	0.315972222	1	0.764248072	1
9	0	1	0.563541667	0.722222222	0.674744278	1
10	0	1	0.270833333	0.555555556	0.629391001	-1
11	0	1	0.173611111	0.555555556	0.658688233	1
12	0	1	0.2046875	0.555555556	0.70546066	1
13	0.25	0.114285714	0.567708333	0	0.307482993	-1
14	0.25	0.114285714	0	0	0.412177276	-1
15	0.25	0.114285714	0.031597222	0	0.720605167	1
16	0.25	0.114285714	0.092361111	0.166666667	0.74436284	1
17	0.25	0.114285714	0.090277778	0.611111111	0.56580402	-1
18	0.25	0.114285714	0.321006944	0.666666667	0.68537888	1
19	0.25	0.114285714	0.292708333	0.722222222	0.74346213	1
20	0.25	0.114285714	0.564583333	1	0.77637305	1
21	0.25	0.114285714	0.538194444	0.722222222	0.8679499	1
22	0.25	0.114285714	0.513715278	0.555555556	0.59374997	-1
23	0.25	0.114285714	0.650694444	0.555555556	0.5023635	-1
24	0.25	0.114285714	0.321527778	0.555555556	0.42334203	-1
25	0.5	0	0.377777778	0	0	-1
26	0.5	0	0.386805556	0	0.21034789	-1
27	0.5	0	0.317534722	0	0.44076854	-1
28	0.5	0	0.414756944	0.166666667	0.5687435	-1
29	0.5	0	0.279861111	0.611111111	0.5311368	-1
30	0.5	0	0.463888889	0.666666667	0.74977437	1
31	0.5	0	0.195486111	0.722222222	0.7214374	1
32	0.5	0	0.550173611	1	0.93810741	1

No	Plant Age	Land Area	Rainfall	Soil Type (pH)	Yield Production	Yield Label
33	0.5	0	0.375173611	0.722222222	0.85917636	1
34	0.5	0	0.570573611	0.555555556	0.85976386	1
35	0.5	0	0.609027778	0.555555556	0.65685343	1
36	0.5	0	0.788020833	0.555555556	0.65792720	1
37	0.75	0.17142857	0.153993056	0	0.14773400	-1
38	0.75	0.17142857	0.008854167	0	0.35311577	-1
39	0.75	0.17142857	0.186805556	0.166666667	0.56308317	-1
40	0.75	0.17142857	0.052430556	0.611111111	0.55128600	-1
41	0.75	0.17142857	0.513888889	0.611111111	0.65462463	-1
42	0.75	0.17142857	0.177951389	0.722222222	0.86915410	1
43	0.75	0.17142857	0.426388889	1	0.74303798	1
44	0.75	0.17142857	0.564236111	0.722222222	0.88635529	1
45	0.75	0.17142857	0.369444444	0.555555556	0.52913160	-1
46	0.75	0.17142857	0.347569444	0.555555556	0.57629891	-1
47	0.75	0.17142857	0.611631944	0.555555556	0.082382996	-1
48	0.75	0.17142857	0.182291667	0	0.18789919	-1
49	1	0.534453782	0.188715278	0	0.258828886	-1
50	1	0.534453782	0.057465278	0	0.268549249	-1
51	1	0.534453782	0.360763889	0	0.481601998	-1
52	1	0.534453782	0.078472222	0.166666667	0.435598653	-1
53	1	0.534453782	0.531076389	0.611111111	0.708173634	1
54	1	0.534453782	0.159201389	0.666666667	0.72334519	1
55	1	0.534453782	0.193055556	0.722222222	0.873852576	1
56	1	0.534453782	0.425520833	1	0.989068762	1
57	1	0.534453782	0.560763889	0.722222222	0.90607286	1
58	1	0.534453782	0.373090278	0.555555556	0.717581135	1
59	1	0.534453782	0.317361111	0.555555556	0.508213305	-1
60	1	0.534453782	0.582986111	0.555555556	0.472638996	-1

Table 5 shows the results of the dataset classification based on normalized yield values. A binary labeling scheme is applied, where data points with yield values above the median threshold are assigned to class 1, while those below the threshold are assigned to class −1. The excerpt illustrates that the labeling process successfully separates the dataset into two distinct classes, forming the basis for SVM training. This transformation allows the model to focus on distinguishing high-yield and low-yield production patterns rather than predicting continuous values.

Table 6. Gaussian Kernel (RBF) Computation Results

	A1	A2	A3	A4	A5	A6	A7	A8	…	A47	A48
A1	1	0.973921	0.953152	0.918763	0.310515	0.370241	0.632284	0.622775	…	0.030045	0.026018499
A2	0.973921	1	0.996814	0.8589	0.248263	0.301014	0.620523	0.57334	…	0	0
A3	0.953152	0.996814	1	0.869927	0.258181	0.312155	0.634839	0.587648	…	0.02224	0.01074
A4	0.918763	0.8589	0.869927	1	0.535435	0.607724	0.830181	0.822167	…	0.08303	0.02504
A5	0.310515	0.248263	0.258181	0.535435	1	0.827769	0.786876	0.827939	…	0.19536	0.06243
A6	0.370241	0.301014	0.312155	0.607724	0.827769	1	0.848771	0.856672	…	0.184217813	0.046216028
A7	0.632284	0.620523	0.634839	0.830181	0.786876	0.848771	1	0.996974	…	0.177683505	0.025653552
A8	0.622775	0.57334	0.587648	0.822167	0.827939	0.856672	0.996974	1	…	0.11762429	0.012682733
A47	0.030045	0	0.02224	0.08303	0.19536	0.184217813	0.177683505	0.11762429	…	1	0.675028762
A48	0.026018499	0	0.01074	0.02504	0.06243	0.046216028	0.025653552	0.012682733	…	0.675028762	1

Table 6 presents the results of the Gaussian kernel computation used during the SVM training process. Each value represents the similarity between pairs of data points in the transformed feature space. High kernel values, such as those close to 1, indicate strong similarity, while lower values indicate weaker relationships. The presence of high similarity

ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

values among certain data points suggests clustering behavior within the same class, which supports the effectiveness of the RBF kernel in capturing nonlinear patterns in the dataset. These kernel computations form the foundation for constructing the optimal decision boundary in the SVM classification model.

3.4 Prediction and Evaluation Results

This subsection presents the prediction and evaluation outcomes of the Support Vector Machine model using the testing dataset. The evaluation focuses on the computation of decision function values, classification results, comparison between actual and predicted labels, and overall model performance. These results provide a clear assessment of the model's ability to classify palm oil fresh fruit bunch yield accurately.

Table 7. Results of f(x) Computation and Classification.

Testing No.	f(x) Value	Classification Result (Sign f(x))
1	0.00085883	1
2	0.001137083	1
3	0.000546	1
4	0.000493113	1

Table 7 shows the output values of the SVM decision function f(x) for each testing instance. All f(x) values are positive, resulting in a predicted class label of 1 for all testing data. The magnitude of f(x) indicates the distance of each testing sample from the decision boundary. Testing data 2 produced the highest f(x) value (0.001137083), indicating stronger confidence in its classification compared to other samples.

Table 8. Comparison Between Actual Labels and System Classification

Testing No.	Actual Label	Classification Result	Remark
1	1	1	Correct
2	1	1	Correct
3	-1	1	Incorrect
4	-1	1	Incorrect

As shown in Table 8., two out of four testing samples were classified correctly, while the remaining two samples were misclassified. Misclassification occurred when the model predicted class 1 for samples whose actual labels were -1. This result indicates that the model tends to favor the positive class in the given testing scenario. The evaluation results show that the proposed SVM model achieved an accuracy of 50%, indicating that two out of four testing samples were classified correctly. This result reflects a moderate classification performance and suggests that the model is able to capture general yield patterns, although misclassification still occurs for certain samples.

3.5 Results Discussion

This section discusses the overall performance of the proposed Support Vector Machine model based on prediction and evaluation results. The discussion focuses on the model's classification behavior, prediction accuracy, and error magnitude as reflected by the evaluation metrics. These results provide insight into the strengths and limitations of the model when applied to palm oil fresh fruit bunch yield classification.

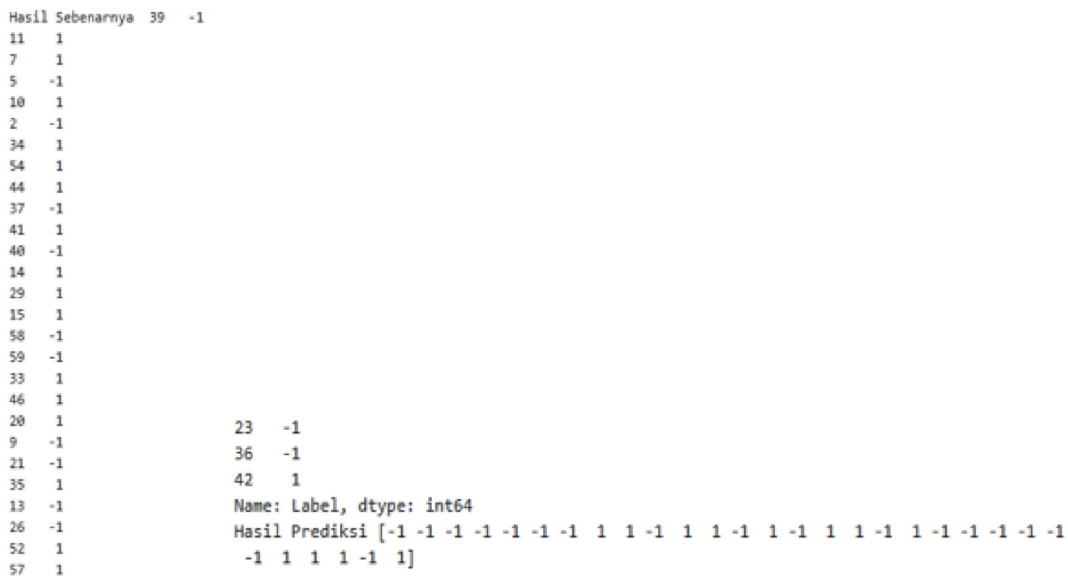

Figure 2. Prediction Results of the Model

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

Figure 2 illustrates the comparison between actual labels and predicted labels generated by the SVM model on the testing dataset. The figure shows that the model is able to correctly classify a substantial portion of the testing data, while several misclassifications are also observed. Based on the evaluation results, the model correctly predicted 17 out of 30 testing samples, resulting in an accuracy of 56.67%. The misclassified samples indicate that the decision boundary formed by the model still overlaps between classes, which may be influenced by data distribution, limited training samples, and overlapping feature characteristics between high-yield and low-yield classes.

```
Prediksi Benar :  17.0  data uji dari :  30  data uji
Persentase :  56.666666666666664  %
RMSE :  1.3165611772087666
```

Figure 3. RMSE Result of the Model

Figure 3 presents the Root Mean Square Error value obtained from the prediction results. The model achieved an RMSE of 1.316561, indicating the average magnitude of error between the predicted class values and the actual class labels. This RMSE value suggests that the model's prediction errors are still relatively significant for a binary classification task, reflecting moderate predictive performance. The RMSE result, together with the accuracy value, confirms that while the SVM model demonstrates the ability to capture general patterns in the dataset, further optimization such as parameter tuning, feature selection, or increased data volume is required to improve classification reliability..

4. CONCLUSION

This study successfully developed and evaluated a Support Vector Machine–based classification model to predict palm oil fresh fruit bunch yield using agronomic and environmental variables. The research addressed the problem of yield classification by integrating data preprocessing, labeling, kernel-based learning, and systematic evaluation. The results demonstrate that the SVM model with a Gaussian Radial Basis Function kernel is capable of capturing nonlinear relationships within the dataset. Based on the testing results, the model achieved an accuracy of 56.67%, correctly classifying 17 out of 30 testing samples, while producing an RMSE value of 1.316561. These findings indicate that the model is able to identify general production patterns, although misclassification still occurs in several cases. The evaluation results suggest that the performance of the model is influenced by factors such as limited dataset size, overlapping feature characteristics between classes, and the simplicity of the binary labeling scheme. Despite these limitations, the model provides a foundational framework for yield classification and demonstrates the feasibility of applying SVM for agricultural production analysis. This study contributes to the application of machine learning techniques in palm oil plantation management by offering a structured and reproducible approach to yield prediction. For future research, improvements can be made by increasing the volume and diversity of training data, incorporating additional influential variables, applying feature selection techniques, and optimizing kernel parameters to enhance model accuracy and robustness. The integration of alternative machine learning algorithms or ensemble methods may also be explored to achieve more reliable and precise prediction results.

REFERENCES

[1] A. Rifky dan S. Anugrah, "Sinkronisasi Kebijakan Kelapa Sawit Berkelanjutan," *J. Kebijakan Publik*., vol. 14, no. 3, hal. 319–328, 2023, doi: 10.31258/jkp.v14i3.8310.

[2] W. Fadhillah dan R. Susanti, "Kerapatan Dominansi Gulma Pada Tanaman Kelapa Sawit Pasca Aplikasi Tandan Kosong Kelapa Sawit (TKKS) dan Kompos Tandan Kosong Kelapa Sawit," *Agrium J. Ilmu Pertanian.* vol. 26, no. 2, hal. 143–148, 2023, doi: 10.30596/agrium.v26i2.16570.

[3] M. F. Praevia, "Analisis Pemanfaatan Limbah Tandan Kosong Kelapa Sawit Sebagai Cofiring pada PLTU Batubara," *J. Energi Baru dan Ter.,* vol. 3, no. 1, hal. 28–37, 2025, doi: 10.14710/jebt.2022.13367.

[4] E. Lubis, K. Tampubolon, M. Alqamari, dan R. Cemda, "Keanekaragaman Gulma Pada Integrasi Kelapa Sawit Dengan Padi Sawah," *J. Agrotek Tropika,* vol. 10, no. 2, hal. 321–328, 2022, doi: 10.23960/jat.v10i2.5786

[5] F. B. Kurniawan dan L. Farokhah, "Aplikasi Cerdas Prediksi Kelulusan Mahasiswa Berbasis Website Menggunakan Metode Support Vector Machine (SVM)," *J. Fasilkom Tekno. Inf. dan Ilmu Komp.,* vol. 15, no. 1, hal. 155–162, 2025, doi: 10.37859/jf.v15i1.8767

[6] A. K. Darmawan, I. Yudhisari, A. Anwari, dan M. Makruf, "Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest," *J. Minfo Polgan Pen. Manaj. Inf.,* vol. 12, hal. 387–400, 2023, doi: 10.33395/jmp.v12i1.12388.

[7] R. Safira dan I. P. Sari, "Model Prediksi Perkembangan Tumbuh Kembang Anak untuk Pencegahan Risiko Gizi Buruk Menggunakan Support Vector Machine (SVM) dengan Feature Selection Backward Elimination," *Blend Sains J. Teknik,* vol. 4, no. 1, hal. 1–11, 2025, doi: 10.56211/blendsains.v4i1.812.

[8] R. S. Sinambela, M. Ula, dan A. F. Ulva, "Prediksi Harga Emas Menggunakan Algoritma Regresi Linear Berganda Dan Support Vector Machine (SVM) Prediction Gold Price Prediction Using Multiple Linear Regression Algorithm and Support Vector Machine (SVM)," *Justin J. Sistem dan Tek. Inf.,* vol. 12, no. 2, hal. 253–258, 2024, doi: 10.26418/justin.v12i2.73386.

[9] R. Hidayad, R. A. Ronaldo, dan R. A. Prasetiyo, "Optimasi Parameter Support Vector Machine Menggunakan Algoritma Genetika untuk Meningkatkan Prediksi Pergerakan Harga Saham," *Coreai J. Kecerdasan Buatan, Komputasi dan Teknol. Inf.*, vol. 3, no. 1, 2022, doi: 10.33650/coreai.v3i1.3859.

[10] H. S. Wafa *et al.*, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Index Infor. nd*

ISSN 2774-3659 (Media Online)
Vol 6, No 2, February 2026 | Hal 744-752
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i2.1002

digital exp. vol. 4, no. 1, hal. 40–45, 2022, doi: 10.36423/index.v4i1.895.

[11] L. Lukman dan H. Herlinda, "Prediksi Kelulusan Siswa dengan Metode Support Vector Machine (SVM) di SMK Adiluhur," *J. STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 9, no. 1, 2024, doi: 10.30998/string.v9i1.23355.

[12] D. Sari, "Prediksi Gangguan Tidur pada Sleep Health and Lifestyle Menggunakan Support Vector Machine dan Neural Network," *J. Vokasi Inf.,* vol. 4, no. 1, hal. 36–42, 2024, doi: 10.24036/javit.v4i1.168.

[13] Sherly dan D. Sitanggang, "Model Prediksi Obesitas Dengan Menggunakan Support Vector Machine," *J. Sist. Infromasi Dan Ilmu Komput. Prima*, vol. 5, no. 2, hal. 172–175, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2443.

[14] M. R. Pradana, W. Witanti, dan A. Komarudin, "Prediksi Tingkat Keparahan Diabetes Melitus Menggunakan Support Vector Machine (SVM) dengan Kernel Polinomial dan RBF," *J. Locus Penelit. dan Pengabdi.*, vol. 4, no. 8, hal. 7521–7533, 2025, doi: 10.58344/locus.v4i8.4357.

[15] Sudriyanto, F. Syahro, dan N. Fitrian, "Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, Dan K-Nearest Neighbors Dalam Prediksi Harga Saham," *Jars (J. of Advanced Research in Inf.),* vol. 2, no. 1, hal. 13–21, 2023, doi: 10.24929/jars.v2i1.2983.

[16] B. I. Nugroho, N. A. Santoso, A. A. Murtopo, "Prediksi Kemampuan Akademik Mahasiswa dengan Metode Support Vector Machine," *REMIK Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 7, hal. 177–188, 2023, doi: 10.33395/remik.v7i1.12010.

[17] N. S. Syam, V. Yoga, P. Ardhana, E. A. M. Sampetoding, dan M. Sulthan, "Model Support Vector Machine untuk Prediksi pada Penggunaan Energi Listrik di Rumah Hemat Energi," *J. Informatika,* vol. 1, no. 2, hal. 56–59, 2022, doi: 10.57094/ji.v1i2.360.

[18] A. P. Putra dan A. M. Harahap, "Implementasi Time Series Forecasting dengan Algoritma LSTM untuk Pemantauan dan Prediksi Produktivitas Kelapa Sawit Berdasarkan Hasil Panen," *Jurikom (J. Riset Komp.),* vol. 12, no. 2, hal. 46–56, 2025, doi: 10.30865/jurikom.v12i2.8495.

[19] S. Ayunda, "Prediksi Jumlah Produksi Kelapa Sawit di Provinsi Riau Menggunakan Metode Double Moving Average," *ICoPSS: Indonesian C. of Premier Stal Sci.* vol. 3, no. 2, hal. 62–70, 2024, doi: 10.24014/icopss.v3i2.32269.

[20] M. F. Muttaqin, T. Bukhori, N. Agustina, dan M. Naseer, "Sistem Prediksi Berita Palsu Tentang Virus Covid-19 Menggunakan Algoritma Support Vector Machine (SVM)," *Naratif J. Nas. Ris. Apl. dan Tek. Inform.*, vol. 05, no. 01, hal. 26–33, 2023, doi: 10.53580/naratif.v5i1.187.