Classification Of Nutrient Deficiency In Lettuce Plants (*Lactuca Sativa*) Using Machine Learning Algorithm

Zuriati [1,a)] , Dewi Kania Widyawati [2,b)] , Kurniawan Saputra [3,c)] , Oki Arifin [4,d)]

[1,2,3,4] Department of Information Technology, State Polytechnic of Lampung, Bandar Lampung, Indonesia

[a)] Corresponding author: zuriati_mi@polinela.ac.id
[b)] dewi_mi@polinela.ac.id
[c)] kurniawan_mi@polinela.ac.id
[d)] okiarifin@polinela.ac.id

Abstract. Plants require appropriate nutrients or nutrients for their growth and development. Inappropriate nutrient levels can interfere with the plant growth process, resulting in less-than-optimal harvest results. Therefore, it is very important for farmers to know the nutrient levels of their plants, neither excessive nor lacking. Identification of nutrient deficiencies in plants such as Lettuce (Lactuca Sativa) traditionally requires careful observation of the physical characteristics of the plant, which is often long-drawn out and stand in need of a high level of accuracy. Leaf color is often used as an indication, for example if it is pale or yellow it can indicate a lack of nitrogen or iron. This requires expertise and experience in cultivation for lettuce cultivators. So, a tool is needed that can identify nutrient deficiencies accurately, quickly, and easily. This study aims to overcome this challenge, namely identifying nutrient deficiencies in lettuce plants. This approach utilizes machine learning technology to distinguish four main classes of deficiencies, namely: nitrogen (N), phosphorus (P), and potassium (K), as well as normal or healthy lettuce leaf conditions. The proposed research method consists of the following stages: 1). Lettuce leaf image dataset collection, 2). Preprocessing dataset, 3). Implementation of machine learning using the Support Vector Machine (SVM) algorithm. In the implementation of SVM, experiments were carried out by applying various SVM kernel spesifically: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid, 4). Evaluation of model performance. Model performance was evaluated by measuring its level of accuracy in classifying nutrient deficiencies in Lettuce leaf image data. The results of the experiment showed that SVM with the RBF kernel had the best accuracy, namely: 92%. The findings of this study provide valuable insights into the effectiveness of machine learning approaches in classifying nutrient deficiencies in Lettuce plants. This study can help farmers to optimize their crop production more efficiently and accurately.

Keywords: Algorithm, SVM Kernel, Machine Learning, Lettuce, Support Vector Machine

INTRODUCTION

Nutrients or nutrients are chemical substances that are very vital for plant growth and reproduction. Plants demand nutrients in balanced proportions during their growth and development [1] . Specifically, plants require the nutrients Nitrogen (N), Phosphorus (P), and Potassium (K) because they play a major role in the vegetative phase for optimal growth [2] . The availability of the nutrients N, P, and K greatly affects the dimensions and number of leaves, which in turn produces healthy vegetable leaves. [1] . The need for nutrients is an important prerequisite in the early stages of plant growth. Therefore, continuous monitoring of nutrient content in plants is crucial.

In agricultural practice, fertilizer application is one of the effective methods to ensure that plants get an optimal supply of nutrients. The availability of sufficient nutrients is very necessary and is generally met through routine fertilization, considering that natural nutrient sources in the soil are limited [3] . To determine the right dose of fertilizer and estimate the cost of its application, information is needed regarding the nutrient content in the soil and the nutritional condition of the plant. That way, the nutritional needs of the plant can be met optimally [4] , [5] . It is important to remember that plants are very susceptible to nutritional problems, both in the form of nutrient deficiencies (lack) and toxicity (excess) of nutrients. Therefore, careful monitoring of the nutritional condition of plants is very crucial. Detection of nutrient deficiencies or toxicity in plants can be done through visual diagnosis, which is a simple and easy-to-apply method [6] . Nutrient deficiency is a condition of lack of minerals or essential nutrients in plants during the growth period [7] . The impact of this deficiency is related to the role of nutrients in plant cell metabolism [5] . Nutrient deficiencies can cause disease in plants, leading to growth disorders and physiological and metabolic changes [5] . Plants that experience nutrient deficiencies will experience growth inhibition and affect the yield of leaves or fruit produced.

If plant care is done traditionally, plants are usually observed visually one by one every day to detect nutritional disorders or nutrients. Identification of nutrient problems in plants can be done by observing the symptoms that appear. In modern agricultural management, nutrient provision must be based on the diagnosis of plant nutrient conditions, either through visible symptoms or through plant analysis [8] . For example, nutrient deficiencies can be identified from symptoms visible on plant parts such as leaves. Leaf color is often an important indicator of whether a plant is healthy or deficient. When lacking essential elements, plants will show typical symptoms of deficiency. These symptoms can be stunted growth in the roots, stems or leaves, as well as the appearance of chlorosis or necrosis in various parts of the plant. Chlorosis is characterized by leaves that lose their green color due to lack of chlorophyll formation, so that the leaves turn yellow or pale white. While necrosis is characterized by damage or death to organ cells such as stems. These symptoms help identify the function of nutrients in plants and provide a guide for farmers to know the right way and time to fertilize their plants [8] .

Visual diagnosis is a useful method for detecting nutrient deficiencies or toxicity in plants by observing visible symptoms. However, this method has several disadvantages, such as: the need for quite large costs, time, and labor. In addition, adequate expertise, knowledge, and experience are required to be able to identify nutrient deficiencies correctly, which are not always possessed by all farmers or cultivators. Therefore, a tool is needed that can overcome these limitations. One solution that can be applied is a classification model to identify nutrient deficiencies, which utilizes machine learning technology from computer science.

Machine learning is a branch of computer science that develops algorithms and statistical models to solve specific tasks by utilizing data patterns and performing inferences [9] . In this study, machine learning technology was used to develop a classification model that is able to detect macronutrient deficiencies, namely: Nitrogen (N), Phosphorus (P), and Potassium (K) in Lettuce plants (*Lactuca Sativa*). The data utilized in digital format images of Lettuce plants were collected and processed to be classified into 4 categories, namely: healthy plants, N deficiency, P deficiency, and K deficiency. In **FIGURE 1** presents image data of healthy lettuce plants.

FIGURE 1. Healthy Lettuce Plants

On **FIGURE 2** shows image data of lettuce leaves that are deficient in the nutrients nitrogen (N), phosphorus (P), and potassium (K).

(a) N deficiency (b) P deficiency (c) K deficiency

FIGURE 2. Lettuce Plants Deficiency of Nutrients (a) K (b) N (c) P

The machine learning algorithm utilized in this research is the SVM. The SVM algorithm has been confirmed to be able to perceived and point out images effectively with a high level of accuracy [9] , [10] , [11] , [12] , [13] . SVM is an algorithm that is well-known for its ability to produce optimal solutions in classification. Vapnik [14] first introduced SVM as a kernel-based machine learning model, which can be applied to both classification and regression tasks. One of the advantages of SVM is its ability to generalize well, even when the available training data is small [9] . SVM also works effectively on high-dimensional data, although this is more optimal if the data is linear, for non-linear data, the use of a kernel function is required so that SVM can still classify well [14]

In the case of non-linear data, the resulting generalization may not be as good as in linear data. Therefore, SVM transforms the primary input set in a high-dimensional space, called the feature space, to maximize the separation between classes. Kernels are then used to transform the data into this higher-dimensional space, called the kernel space, which allows for linear separation of the data. In addition, SVM is very suitable for handling high-dimensional data because the kernel only uses data that is relevant for the classification process. However, until now there has been no consensus on which kernel is best used in a particular application, so this study will compare the performance of four commonly used kernels, namely: Linear, Polynomial, RBF, and Sigmoid. Several studies that compare the performance of SVM kernels include: [15] , [16] , [17] , [13] . Determination of the best SVM model by calculating the best accuracy value and ROC curve. The best kernel is the model with the highest accuracy value.

METHODS

FIGURE **3** shows an illustration of the applied research method. The tool used is Orange Data Mining.

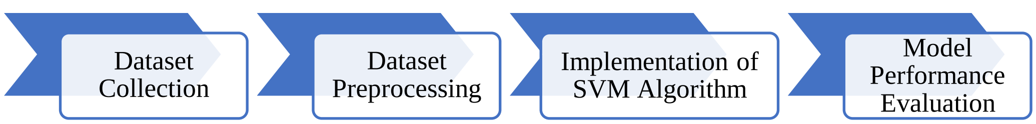

FIGURE 3 . Research Methods

1. Dataset Collection

Problem identification in this study is to develop a nutrient deficiency identification model in lettuce plants using the SVM algorithm and measure model performance using a confusion matrix. Plant data was obtained from experimental data on lettuce plant treatments that were conditioned to lack nutrients N, P, K and healthy plants. The data collected were in the form of image data or photos of lettuce leaves. The collected data were grouped into 4 criteria, namely: healthy plant criteria, Nitrogen (N) nutrient deficiency criteria, Phosphorus (P) nutrient deficiency criteria, and Potassium (K) nutrient deficiency criteria. The total data collected was 211 data. Healthy lettuce plant image data were 12 data, N nutrient deficiency image data were 59, P deficiency was 68, and K deficiency was 72 data.

2. Dataset Preprocessing

Data preprocessing is a crucial step that is carried out before the data is further processed by machine learning, to ensure that the data is in an efficient format and ready to be processed. One of the preprocessing techniques that is often used is embedding, which is the process of transforming data into a form of numerical representation that is more easily understood by machine learning algorithms. Embedding is often used to transform complex data, such as text or images, into vectors of numbers that represent relationships or patterns in the data. This technique not only structures the data into a more structured format, but also helps in dimensionality reduction and extraction of key features, which are critical to improving the efficiency and accuracy of the model.

In image data, embedding is performed before the image is processed by the SVM algorithm. Image data in raw format consists of pixels that indicate color intensity, which cannot be directly interpreted by machine learning algorithms. Therefore, embedding is used to convert an image from a collection of pixels into a more meaningful numeric representation (vector). This vector contains important information about the image, such as patterns, shapes, or key features, which will later be processed by the algorithm. Image data has a very large dimension, especially for high-resolution images, which require a lot of computing resources to process. With embedding, the image dimension can be reduced without losing important information, resulting in a more efficient and compact representation. This process greatly helps in accelerating data processing by the SVM algorithm. In addition, embedding allows the extraction of more relevant features from the image, such as edges, textures, colors, and shapes, which are very important for tasks such as classification or object detection. By extracting only the most significant features, embedding makes it easier for the algorithm to focus on the most useful information. Embedding plays an important role in transforming images into more compact, informative, and ready-to-be-processed representations by machine learning algorithms. This process not only helps the algorithm understand the image content better, but also improves the computational efficiency and accuracy of the results produced.

3. Implementation of SVM Algorithm

Next, after the data is ready to use, the development of a classification model using the SVM algorithm is carried out. The SVM method uses support vectors to separate data into different classes based on the characteristics of each data [14] . SVM applies the kernel transformation technique to convert data into higher feature dimensions, which aims to facilitate class separation. The kernel function works by changing data into a more complex form, so that data patterns or structures become easier to recognize and separate [14] , [9] . There are 4 types of SVM kernel functions that are commonly used in SVM, namely: Linear, Polynomial or Gauss kernel, RBF, and Sigmoid. The formula for Linear SVM is:

$$f(x) = sign(w \cdot x + b) \quad (1)$$

Where f(x) is a prediction function that determines the class of the input data, w is the normal vector of the hyperplane, which represents the weights or coefficients of the features, x is the input feature vector, and b is the bias or interception that determines the position of the hyperplane. This function uses a hyperplane to separate data into two different classes. The input x is projected onto the hyperplane, and the sign of the projection determines the output class of the data.

In this study, 4 types of SVM kernels are used to handle classification problems, namely: Linear, Polynomial RBF, and Sigmoid. Here is the formula for each SVM kernel function:

a. Linear Kernel

Linear kernel is used when the data can be separated linearly, by utilizing the dot product between two input vectors x and y. Here is the formula for Linear kernel.

$$K(x_i, y_j) = x_I \cdot y_j \quad (2)$$

b. Kernel Polynomial

Kernel polynomials allow separating non-linear data by extending features to higher dimensions.

$$K(x_i, y_j) = (\gamma x_{I} \cdot y_j + r)^d \quad (3)$$

γ is a scale parameter (usually $\gamma > 0$), r is a free constant, and d is the degree of the polynomial.

c. Radial Basis Function (RBF) Kernel

RBF uses Gaussian functions (known as radial basis functions) to measure the similarity between two input vectors in the feature space.

$$K(x_i, y_j) = exp(-\gamma||x_i - y_j||^2) \quad (4)$$

Where γ is a parameter that controls how far the influence of one data sample. RBF kernel is very popular because it can handle non-linear classification problems well.

d. Sigmoid Kernel

$$K(x_i, y_j) = \tanh(\gamma x_I . y_j + r) \quad (5)$$

γ is a scale parameter, and r is a constant.

The sigmoid kernel behaves similarly to activation functions in neural networks and is suitable for certain cases that require non-linearity.

4. Model Performance Evaluation

Model performance is measured and evaluated using cross validation and confusion matrix to obtain model accuracy. The relationship between whether data is grouped can be seen from the confusion matrix. This matrix consists of True positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). The confusion matrix is presented in **TABLE 1** .

TAB LE 1. Confusion Matrix

Classification	Positive	Negative
Positive Prediction	TP	FP
Negative Prediction	FN	TN

From the confusion matrix, various parameters can be obtained to attempt the performance of the algorithm, one of which is accuracy. Accuracy measures how well the algorithm can forecast accurately, which is calculated as the ratio between the number of valid forecast and the total amount of data in the dataset. Accuracy provides a general idea of how well the model is performing, but it needs to be considered together with other metrics, especially when dealing with imbalanced data. The formula for calculating accuracy is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

RESULTS AND DISCUSSION

The following are the outcomes and discours of this research.

DATASET COLLECTION

Data were collected and grouped into 4 criteria, namely: healthy plant criteria, Nitrogen (N) nutrient deficiency criteria, Phosphorus (P) nutrient deficiency criteria, and Potassium (K) nutrient deficiency criteria. Data were collected from https://kaggle.com. **FIGURE 4** presents healthy Lettuce plant image data. The total number of healthy Lettuce plant image data is 4 data.

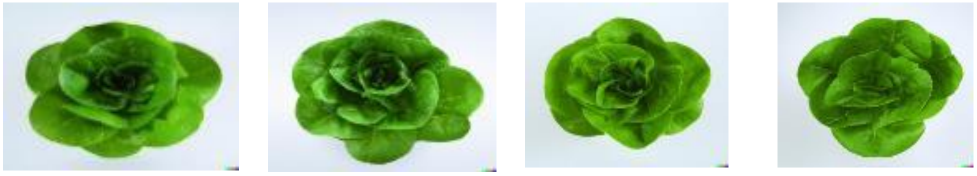

FIGURE 4. Healthy Lettuce Plant Image Data

The total data for lettuce plants lacking N is 59 data. In **FIGURE 5** An example of image data of lettuce plants lacking element N is presented.

FIGURE 5 . Image Data Lettuce Plants Lack N Elements

FIGURE **6** presents an example of image data of lettuce plants lacking P. Overall, 68 data of image data of lettuce plants lacking P were successfully collected.

FIGURE 6. Image Data: Lettuce Plants Lack Element P

In **FIGURE 7,** examples of image data for lettuce plants that lack the element K are presented. There are 72 data on lettuce plants that lack the element K.

FIGURE 7 . Image Data Lettuce Plants Lack N Elements

By using Orange Data Mining tools, the dimensions of each image data can be known. **FIGURE 8** shows a screenshot of the dimensions of the Lettuce image data, data 1 to data 26.

	category	image name	image	size	width	height
1	K	k_1	K\k_1.png	1856983	1024	1001
2	K	k_10	K\k_10.png	1711796	1020	1002
3	K	k_11	K\k_11.png	1716927	1024	1002
4	K	k_12	K\k_12.png	1847055	1020	1002
5	K	k_13	K\k_13.png	1684971	1024	1003
6	K	k_14	K\k_14.png	1683014	1023	1003
7	K	k_15	K\k_15.png	1634295	1023	1006
8	K	k_16	K\k_16.png	1564633	1015	1006
9	K	k_17	K\k_17.png	1572429	1023	1005
10	K	k_18	K\k_18.png	1539969	1024	1003
11	K	k_19	K\k_19.png	1490654	1024	1003
12	K	k_2	K\k_2.png	1696101	1023	1005
13	K	k_20	K\k_20.png	1674077	1024	1003
14	K	k_21	K\k_21.png	1657654	1024	1007
15	K	k_22	K\k_22.png	1583796	1020	1005
16	K	k_23	K\k_23.png	1623002	1023	1006
17	K	k_24	K\k_24.png	1564847	1023	1006
18	K	k_25	K\k_25.png	1671042	1023	1006
19	K	k_26	K\k_26.png	1587893	1023	1005
20	K	k_27	K\k_27.png	1669752	1024	1007
21	K	k_28	K\k_28.png	1586256	1023	1006
22	K	k_29	K\k_29.png	1688225	1023	1006
23	K	k_3	K\k_3.png	1675136	1023	1005
24	K	k_30	K\k_30.png	1674459	1023	1005
25	K	k_31	K\k_31.png	1700141	1024	1005
26	K	k_32	K\k_32.png	1908140	1024	1007

FIGURE 8 . Screenshot of Image Data Dimensions 1 to 26

PREPROCESSING DATASET

The embedding applied in this study is SqueezeNet, which is an efficient neural network model designed to reduce resource usage while maintaining good performance in image classification tasks. The embedding produces 1000 features for each data, labeled from n0 to n999. A screenshot of the image embedding results obtained is presented in **FIGURE 9** .

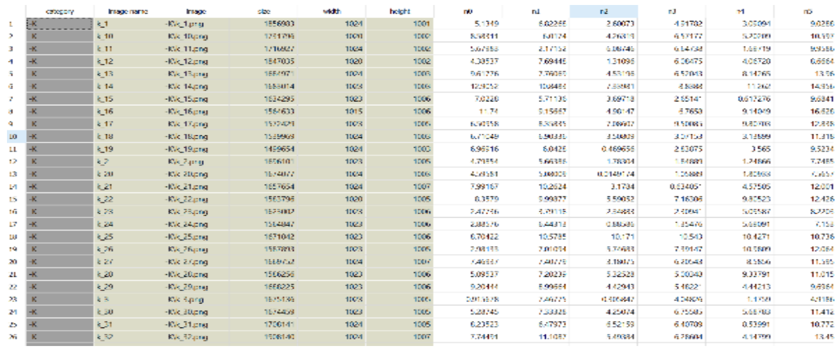

FIGURE 9 . Image Embedding Result Screenshot

IMPLEMENTATION OF SVM ALGORITHM

The next step is to applied the SVM algorithm by conducting experiments using SVM kernels, namely: Linear, Polynomial, RBF, and Sigmoid. The resulting confusion matrix for each SVM kernel is presented from **FIGURE 10** to **FIGURE 13.** FIGURE **10** presents the Linear SVM confusion matrix.

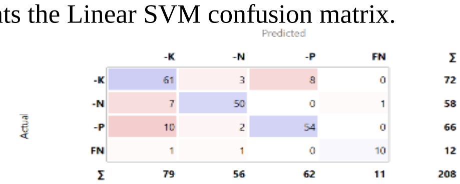

FIGURE 10. Linear SVM Confusion Matrix

FIGURE **11** presents the confusion matrix of Polynomial SVM.

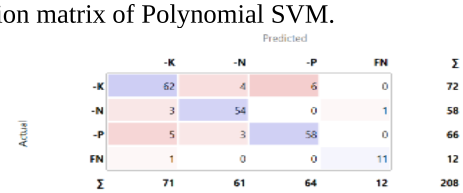

FIGURE 11. Polynomial SVM Confusion Matrix

FIGURE **12** presents the SVM RBF confusion matrix.

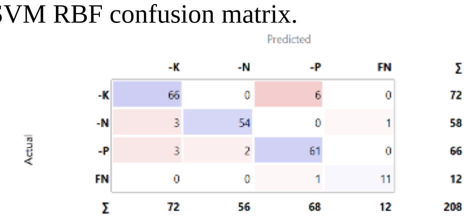

FIGURE 12. SVM RBF Confusion Matrix

FIGURE **13** presents the confusion matrix of the Sigmoid SVM.

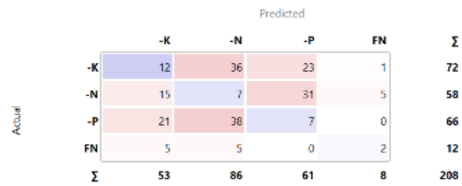

FIGURE 13. Sigmoid SVM Confusion Matrix

MODEL PERFORMANCE EVALUATION

Model evaluation is done by calculating the accuracy value. Accuracy is calculated based on data generated from the confusion matrix using equation (6). The finding of the accuracy calculation are presented in **TABLE 2.**

TABLE 2 . Summary of Model Accuracy

SVM Kernel	Accuracy Value
Linear	84%
Polynomial	89%
RBF	92%
Sigmoid	14%

Based on **TABLE 2** , the RBF kernel deliver upmost accuracy of 92%, followed by the Polynomial kernel with an accuracy of 89%, indicating that both kernels are able to handle the data well and provide accurate classification. The Linear kernel has an accuracy of 84%, which is still quite good but not as good as the previous two kernels. On the contrary, the Sigmoid kernel shows the lowest performance with an accuracy of only 14%, indicating the ineffectiveness of this kernel in handling the dataset. Therefore, the RBF and Polynomial kernels are the best choices for this data.

CONCLUSIONS

Based on the accuracy results of the four SVM kernels, the subsequent findings could be inffered:

1. The RBF kernel provides the highest accuracy of 92%, so it can be concluded that the SVM model with this kernel is better at handling the tested dataset, capable of producing more accurate classification than other kernels.
2. The Polynomial Kernel has an accuracy of 89%, slightly lower than the polynomial kernel. This shows that RBF is also a strong choice for handling non-linear data.
3. Linear Kernel gives 84% accuracy, which is still quite good but not as good as the previous two kernels. Linear kernel is generally suitable for use on datasets that have higher linearity.
4. The Sigmoid kernel has the lowest accuracy, only 14%, indicating that the model with the sigmoid kernel is not able to produce optimal results on this dataset. This kernel may be less suitable for the data being tested.

ACKNOWLEDGEMENTS

The researcher would like to express gratitude to the Lampung State Polytechnic, especially the P3M, for providing funding through the 2024 DIPA Research and Community Service Grant.

REFERENCES

[1] N. Inaya, D. Armita, and H. Hafsan, “Identification of nutritional problems of various types of plants in Palajau Village, Jeneponto Regency,” *J. Mhs. Phylogeny. Biol.* , vol. 1, no. 3, pp. 94–102, 2021.

[2] R. Wati, “Effect of Kinds of Fertilizer on Growth and Yield of Several Sweet Corn Varieties,” *J. Floratek* , vol. 7, pp. 107–114, 2018.

[3] A. Fakhrezi, RE Saputra, and FC Hasibuan, "Design and Construction of an IoT-Based Monitoring System for Nutrients, Humidity, Soil PH and Air Temperature Using an ESP32 Microcontroller," *e-Proceeding Eng.* , vol. 10, no. 1, pp. 778–786, 2023.
[4] B. Fauziyah, "Analysis of red-yellow podzolic soil nutrients as a recommendation for fertilizing cocoa plants aged 0–1 year," *el-Qudwah* , no. 1906, 2010.
[5] R. Ariana, *Plant Diseases Due to Nutrient Deficiency* , no. January. 2016.
[6] D. Armita, W. Wahdaniyah, H. Hafsan, and H. Al Amanah, "Visual Diagnosis of Essential Nutrient Problems in Various Types of Plants," *Teknosains Media Inf. Science and Technology.* , vol. 16, no. 1, pp. 139–150, 2022.
[7] M. Dahria, R. Kustini, R. Gunawan, M. Hutasuhut, and P. Purwadi, "Expert System for Diagnosing Nutrition Definition in Hydroponic Plants Using Certainty Factor Method," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. and Sist. Komput. TGD)* , vol. 6, no. 1, p. 216, 2023.
[8] IW Wiraatmaja, "Mineral Nutrient Deficiency and Toxicity and Their Response to Results," *Teaching Materials* , p. 5, 2017.
[9] S. Agarwal, *Data mining: Data mining concepts and techniques* . 2014.
[10] K. Ahmed, TR Shahidi, SM Irfanul Alam, and S. Momen, "Rice leaf disease detection using machine learning techniques," *2019 Int. Conf. Sustain. Technol. Ind. 4.0, STI 2019* , no. May 2020, p. 1–5, 2019.
[11] A. Handayanto, K. Latifa, ND Saputro, and RR Waliansyah, "Analysis and Application of Support Vector Machine (SVM) Algorithm in Data Mining to Support Promotion Strategy," *JUITA J. Inform.* , vol. 7, no. 2, p. 71, 2019.
[12] E. Haryatmi and S. Pramita Hervianti, "Application of Support Vector Machine Algorithm for On-Time Student Graduation Prediction Model," *J. RESTI (Information Systems and Technology Engineering)* , vol. 5, no. 2, pp. 386–392, 2021.
[13] K. Saputra, "Comparison of Kernel Function Performance of Support Vector Machine Algorithm on Rice Disease Classification," *Ijccs* , vol. 17, no. 1, pp. 119–131, 2023.
[14] Vapnik and VN, "The Nature of Statistical Learning," *Theory* . matter. 334, 1995.
[15] HCS Ningrum, "Comparison of Linear Support Vector Machine (SVM), Radial Basis Function (RBF), and Kernel Polynomial Methods in Classifying Advanced Study Fields Chosen by UII Alumni," *Final Assignment Stat. Univ. Islam Indones.* , pp. 1–90, 2018.
[16] H. Al Azies, D. Trishnanti, and E. Mustikawati PH, "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," *IPTEK J. Proc. Ser.* , vol. 0, no. 6, p. 53, 2019.
[17] AZ Praghakusma and N. Charibaldi, "Comparison of Kernel Functions of Support Vector Machine Methods for Instagram and Twitter Sentiment Analysis (Case Study: Corruption Eradication Commission)," *JSTIE (Jurnal Sarj. Tek. Inform.* , vol. 9, no. 2, p. 88, 2021.