RESEARCH ARTICLE • OPEN ACCESS

Support Vector Machine-Radial Basis Function Kernel and K-Nearest Neighbor Differences for Classification Superior Varieties of Rice in Indonesia

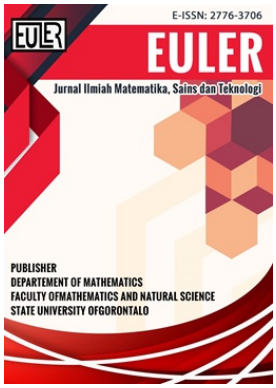
Alissa Chintyana, Kertanah, Siti Hariati Hastuti, and Husnul Khotimah

Volume 12, Issue 2, Pages 157–161, Dec 2024

Received 19 September 2024, Revised 15 November 2024, Accepted 21 November 2024, Published 1 December 2024

To Cite this Article : A. Chintyana, K. Kertanah, S. H. Hastuti, and H. Khotimah,"Support Vector Machine-Radial Basis Function Kernel and K-Nearest Neighbor Differences for Classification Superior Varieties of Rice in Indonesia", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 12, no. 2, pp. 157–161, 2024, *https://doi.org/10.37905/euler.v12i2.27605*

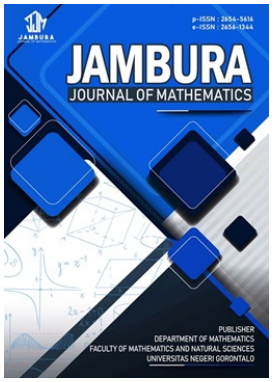
JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
Frequency	:	Biannual (June and December)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/euler
Online ISSN	:	2776-3706
Editor-in-Chief	:	Resmawan
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
Google Scholar ID	:	QF_r-_gAAAAJ
Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics



Jambura Journal of Mathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics

Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi, Volume 12, Issue 2, Pages 157–161, Dec 2024 *https://doi.org/10.37905/euler.v12i2.27605*

Research Article

Check for updates

Support Vector Machine-Radial Basis Function Kernel and K-Nearest Neighbor Differences for Classification Superior Varieties of Rice in Indonesia

Alissa Chintyana[1,*], **Kertanah**[1], **Siti Hariati Hastuti**[1], **and Husnul Khotimah**[1]

[1]*Department of Statistics, Universitas Hamzanwadi, Nusa Tenggara Barat, Indonesia*

ARTICLE HISTORY
Received 19 September 2024
Revised 15 November 2024
Accepted 21 November 2024
Published 1 December 2024

KEYWORDS
Classification
Superior Rice Varieties
Support Vector Machine

ABSTRACT. *Rice is the primary food source for the Indonesian population, making it a priority commodity in Indonesia. Rice production plays a significant role in Indonesia's economic development, with high-yield rice varieties being crucial for enhancing national rice output. Ensuring food security requires the selection of superior rice varieties with optimal quality. This study evaluates various high-yield rice varieties, including INPARA, INPARI, INPAGO, and HIPA, based on characteristic data collected in 2023. Machine learning algorithms, increasingly central to data analysis, were applied, leveraging labeled data suitable for supervised learning methods. During the pre-processing stage, it was determined that the data did not meet the linearity assumption. Thus the Support Vector Machine (SVM) algorithm was modified with the Radial Basis Function (RBF) kernel to better handle non-linear data. Additionally, the K-Nearest Neighbor (KNN) algorithm, a traditional method, was used for comparison. The results indicate that SVM with the RBF kernel achieved faster processing times and the accuracy value reaches 96%, nearly 10% higher than the KNN algorithm.*

1. Introduction

Agriculture is one of the priority sectors for development in Indonesia, with rice as the primary food crop commodity. The increase in population each year requires an increase in rice production to meet the food needs of the community. In addition, rice production results support the development of the Indonesian economy [1]. The factor that plays an important role in increasing national rice production is superior rice varieties. So far, the superior varieties used must be certified by the local agro-ecosystem [2]. Superior rice varieties are divided into four types: Inbred Irrigated Rice (INPARI) which is a variety that is suitable for planting in rice fields, Hybrid Rice (HIPA) rice, which is a composite between two or more divergent varieties and planted in rice fields, Inbred Gogo Rice (INPAGO) this variety is the one that grows in dry land and Inbred Swamp Rice (INPARA) rice that can be grown in swampy areas where this rice is also called tidal rice. The selection of superior rice varieties is carried out carefully to meet the criteria for each type of variety. The classification process can be an alternative for variety selection [3]. Classification is the process of labeling new data or objects based on certain qualities. There are three classification stages, including model construction, model application, and evaluation [4]. Modeling involves creating examples using training data that already have attributes and classes. This information is then used to select new data or object classes. In terms of the precision of developing and applying models to latest data, an evaluation of the data is carried out. The classification process is used to determine records from previously defined new data and the data is entered into several classes [5].

The classification of superior rice varieties in Indonesia utilizes cross-sectional data, which includes the number of regions cultivating these varieties. The number of regions is classified as a small dataset. Classification algorithms that are sensitive to small datasets include K-nearest neighbor (KNN) and Support Vector Machine (SVM). KNN classification is one of the easy-to-implement algorithms with a high level of effectiveness for handling small datasets [6, 7] and is widely used in cyber-physicsocial systems (CPSS) as a data mining method [8]. The KNN method is a lazy learning technique, which means that this method is used in the classification of closely spaced data [4]. In contrast, the SVM algorithm is employed to find an optimal classifier function that can separate datasets from different classes. SVM is a non-parametric classification method that does not have to meet certain assumptions and distributions [9] and it also offers a fast computational process [10].

Previous research utilizing the Modified K-Nearest Neighbor (MKNN) method for classifying superior rice variety groups achieved an average maximum accuracy of 79.96% and a minimum accuracy of 51.2% [11]. Both the K-Nearest Neighbor (KNN) and SVM algorithms are known for their high accuracy in sentiment analysis [12] and efficient analysis time [10]. Generally, real-world data are seldom linearly separable, most data exhibit non-linear characteristics. Data on superior rice varieties and their characteristics are typically categorical, with relationships

*Corresponding Author.

Email : *alissachintyana@hamzanwadi.ac.id* (A. Chintyana)
Homepage : http://ejurnal.ung.ac.id/index.php/euler/index / E-ISSN : 2776-3706

Table 1. Research variables

Variable	Meaning	Category	Symbol
Plant Shape	Appearance of plant clump stands.	Upright, moderately upright, open, scattered and creeping/irregular	X1
Flag Leaf	The last leaf comes out of the stem	Upright, moderately upright/medium, horizontal and drooping	X2
Grain Shape	The shape of the grain is grouped based on the ratio between the length and width of the grain.	Round, moderately round, medium, and slender/long	X3
Grain Colour	The palea and lemma color at the time of grain ripening	Dark beige, yellow, clean yellow, golden yellow, straw yellow, and dirty yellow	X4
Loss	How easy it is for the grain to fall off when the panicles are held by hand	Resistant, medium, and easy	X5
Sluggishness	Position of plant stand	Resistant, moderately resistant, moderately vulnerable, weak, and very weak	X6
Rice Texture	The texture of the rice	Glutinous, very fluffy, plump, para	X7
Superior Rice Varieties	A secondary taxonomic rank under species	Inpari, Inpara, Hipa, Inpago	Y

between categorical variables being inherently non-linear. This non-linearity arises because changes in one variable do not result in proportional changes in other variables. Consequently, the non-linear SVM algorithm is applied to classify data on superior rice varieties. To address non-linear problems, researchers modify SVM by incorporating a kernel function [13]. For robust optimization, researchers use the Radial Basis Function (RBF) kernel within SVM. Both KNN and SVM algorithms are applied for classifying superior rice varieties, as they do not rely on specific assumptions. Based on the description above, this study aims to evaluate the effectiveness of SVM with the RBF kernel and KNN in classifying superior rice varieties.

2. Methods

The data on superior rice varieties is categorical and consists of observations collected in 2023. In practice, categorical data types cannot be analyzed with classical parametric methods because there will be violated assumptions. To overcome this, researchers use machine learning methods with supervised learning.

2.1. Data Description

This study uses a quantitative research method. The data source in this study was taken from the official website of the Rice Crop Research Center (BBPADI). The object of the study is the area with superior rice varieties in Indonesia as many as 95 observations. The research variables consisted of four types of superior rice varieties as dependent variables and plant shape, flag leaves, grain shape, grain color, shedding, recumbentness, and rice texture as independent variables. Table 1 presents more complete information related to the research data.

Classification is applied to determine the quality of the best superior rice varieties. In the classification, data is first split into training data and testing data. This study uses a 70:30 ratio for training data:testing data. SVM and KNN analysis in classifying superior rice varieties using statistics software. In this study, the best method can be seen from the value based on the results of the Accuracy and Sensitivity values.

2.2. Support Vector Machine

The Support Vector Machine (SVM) is a dividing machine with the main idea of building a hyperplane that can effectively separate data points into different classes. In SVM, there are several terms: hyperplane is the decision boundary as the separator among classes, pattern is the member of the class, and margin is the distance between the hyperplane and the nearest pattern of each class. This closest pattern is referred to the support vector. The effort to find the location of the hyperplane is the core of the learning process in SVM [14]. Hyperplanes are said to be optimal if they are right in the middle of the classes so that they have the farthest distance to the outermost data in both classes. The best hyperplane can be found by measuring the margin and maximum support vector. The number of input variables determine the dimensions of the hyperplane [15]. The dimensions of the hyperplane defined by $p-1$, p is the input features.

In general, the data problems obtained are rarely linear, many are non-linear. The classification rule for the non-linear problem is: for all observation $i = 1,2,\ldots,n$ [10],

$$y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) - 1 + \xi_i \geq 0, \tag{1}$$

y_i is the labeled data, $\boldsymbol{x}_i$ is input variables, $\boldsymbol{w} = (w_1, w_2, \ldots, w_p)^{\mathrm{T}}$ is the vector of coefficients and b be a constant. The primal problem in non-linear SVM that can be solved is [10]:

$$h\left(\boldsymbol{x}\right) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b \text{ and } \xi, \tag{2}$$

that minimizes $E_p\left(\boldsymbol{w}, \xi\right) = \frac{1}{2}\|\boldsymbol{w}\|^2 + c\sum_{i=1}^{n}\xi_i$ and subject to Eq. (1). Eq. (2) is the hyperplane function that minimizes $E_p\left(\boldsymbol{w}, \xi\right)$ with additional c. Parameter c is the regularization parameter that controls the trade-off between margin maximization and missclassification. For linearization aim, the kernel can be used to transform data into a higher dimensional space [10]. Data is stored in the form of a kernel that measures the similarity or inequality of data objects. The kernel can be built for a variety of data objects ranging from continuous data and discrete data through data sequences and graphs. With the additional kernel function, the dual problem of the optimization for SVM is maximize [10]:

$$E_D\left(\boldsymbol{\alpha}\right) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right), \tag{3}$$

α_i is the lagrange multiplier associated with training data $(\boldsymbol{x}_i, y_i)$ and $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the kernel function.

The radial base function (RBF) kernel leverages the Gaussian function to evaluate the degree of similarity between two input vectors in feature space. The RBF kernel is one of the most robust kernels for all types of data. In an RBF kernel, the smaller the kernel value, the greater the distance between the input vectors. The result of this kernel will affect the contribution of the vector to the decision boundary generated by the SVM model. The smaller the kernel value, the less influence it has on the separation process [16]. The kernel function that will be used in this study is the Radial Basis Function (RBF) Kernel. Eq. (4) is a function of the RBF kernel [10].

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \parallel \boldsymbol{x}_i - \boldsymbol{x}_j \parallel^2\right), \tag{4}$$

where gamma (γ) is parameter of RBF kernel.

2.3. K-Nearest Neighbor

The KNN method is one of the methods used to classify data based on training data. This method is part of supervised learning, the latest data is categorized based on the proximity of the number of existing categories. This means that the KNN algorithm classifies new data based on the distance of the datum and its proximity to other data [17]. In this case, the distance used is the Euclidean distance. The Euclidean distance is shown in Eq. (5) [4].

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}, \tag{5}$$

where $\boldsymbol{x}_1 = (x_{11}, x_{12}, \ldots, x_{1n})$ and $\boldsymbol{x}_2 = (x_{21}, x_{22}, \ldots, x_{2n})$ are input variables. Distance calculations are performed for each input variable in pairs. The basic concept of the KNN method is to find the shortest distance between the reclassified data in the environment *k* (minimum distance) that is closest to the training data. The largest number of classes/categories with the smallest distance becomes new data, namely classes/categories of testing data. The advantage of the KNN method is that it is simple and effective when the data is large enough and works well enough, and it is resistant to noisy data storage [18]. However, this method also has a disadvantage, namely that the more training data is used, the longer the calculation time will be used.

2.4. Model Goodness

The Confusion Matrix can be interpreted as a tool that has a function to analyze whether the classifier is good at recognizing tuples from different classes. To calculate the accuracy of a performance system, a confusion matrix is generally used. Confusion Matrix is a method that is often applied to determine the performance of an algorithm. This method is in the form of a diagram that contains actual information about the approximate classification of a system using data in the form of a matrix as a result. The values of True-Positive (TP) and True-Negative (TN) provide information when the classifier in classifying the data is correct, while False-Positive (FP) and False-Negative (FN) provide information when the classifier is wrong in classifying the data [17]. Table 2 is a form of the confusion matrix.

Table 2. Confusion Matrix

Actual	Predict	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

where

TP : the sum of the correct predicted outcomes with a positive original class.

FP : the sum of the wrong predicted outcomes with a positive original class.

TN : the sum of the correct predicted results with a negative original class.

FN : the sum of the wrong prediction results with the original class being negative.

$$A = \frac{TP+TN}{TP+TN+FP+FN}, \tag{6}$$

$$S = \frac{TP}{TP+FN}. \tag{7}$$

Table 2 shows the confusion matrix that will be used to calculate the accuracy of the classification method. Accuracy (A) and sensitivity (S) are formulated in Eq. (6) and Eq. (7) [4], respectively.

2.5. Analysis Step

The analysis techniques used in this study are the Support Vector Machine (SVM) and the K-Nearest Neighbors (KNN) algorithm. Data processing using statistics software. The measure of the model's goodness is the value of accuracy and consistency such as equations Eq. (3) and Eq. (4). The following is a description of the SVM and KNN algorithms.

1. Split data into training and testing data (Section 2.1).
2. Support Vector Machine analysis steps:
   (a) Labeled data as reference that grouped into certain classes.
   (b) Select RBF kernel function (Eq. (4)).
   (c) Determine parameter C and Gamma value.
   (d) Maximize the dual problem (Eq. (3)).
   (e) Forming the hyperplane function (Eq. (2)).
   (f) Classification predictions.
   (g) Confusion Matrix Creation (Table 2).
3. K-Nearest Neighbour analysis steps:
   (a) Specifies the *k* parameter to determine the number of nearest neighbors.
   (b) Calculate the Euclidean distance using training data (Eq. (5)).
   (c) Sort the distance of the calculation result at 2) from the smallest distance.
   (d) Collecting classifications based on *k*-values.
   (e) Find the number of classes from neighbour which have the closest distance.
   (f) Classification predictions.
   (g) Confusion Matrix Creation (Table 2).

3. Results and Discussion

The machine learning algorithm starts with a preprocessing stage. During this process, it was found that the data

Table 3. Data Characteristics

Varieties	Mode						
	Plant shape	Flag Leaf	Grain Shape	Grain Colour	Loss	Sluggishness	Rice Texture
Inpari	Upright	Upright	Slim	Clean Yellow	Moderate	Moderate	Fluffier
Inpara	Upright	Upright	Medium	Yellow	Moderate	Moderate	Pera
Hipa	Upright	Upright	Slim	Straw Yellow	Moderate	Resistant	Fluffier
Inpago	Upright	Perpendicular	Medium	Straw Yellow	Moderate	Resistant	Medium

contains no outliers and is non-linear. The analysis continues with data exploration in Section 3.1 and the selection of the best model in Section 3.2. In the SVM algorithm, besides the RBF kernel, we also tested the sigmoid, polynomial, and additive kernels. The RBF kernel is presented in this article as the one with the highest accuracy.

3.1. Exploratory Data Analysis

The dataset to be analyzed is first studied how the pattern and characteristics. This is related to the suitability of the method to be used. The patterns and characteristics in question are such as data distribution and data descriptive [19]. Data description was carried out to determine the characteristics of observational data. Descriptive analysis can be used to see an overview of the characteristics of each variable used. This study used 95 observations of superior rice varieties with the Inpari type dominating the data. The mode value is used to determine the characteristics of superior rice variates.

Table 3 is a characteristic based on the value of the mode of each superior rice variety based on its independent variables. Inpari differs from Inpara based on the shape of the grain, the color of the grain, and the texture of the rice. Inpago can be distinguished from the other three types of varieties because it has flag leaves that tend to be upright and medium rice texture. Other characteristics can also be known by referring to the results of the mode in Table 3. From Table 3, it can also be known the characteristics of superior rice varieties in Indonesia. Most of the superior varieties of rice are characterized by upright plant shapes, upright flag leaves, slender grain shapes, straw yellow grain color, resistant reclining, and fluffy rice texture.

3.2. Model Selection

This research involves categorical data, for which poisson or logistic models can be considered appropriate for modeling [20]. Categorical data is generally non-linear and does not follow a normal distribution. Data mining techniques are commonly used as alternatives for handling non-linear data. The data mining process is preceded by a pre-processing. Pre-processing includes data cleansing from missing values and outliers. The next step is to split the data into training and test dataset. The ratio used is $70:30$ for the training: test data so that 66 observations of the training data and the rest of the test data. The Machine Learning algorithm adjusts its parameters based on the data provided during training. The model is trained using the training dataset, and its performance during this process is evaluated using a validation (test) dataset [4]. This evaluation assesses the model's ability to recognize general patterns within the dataset. The selection of the best model is based on accuracy and sensitivity metrics (Table 4 and Table 6).

Table 4. Goodness of Model SVM

Varieties	Accuracy	Sensitivity
Inpari	95.65%	100.00%
Inpara		100.00%
Hipa		100.00%
Inpago		93.33%

The SVM algorithm is implemented using the RBF kernel. This RBF method requires the tuning parameter cost (c) and gamma values. In this study, a range of 1 – 100 was used for the c parameter and a range of 1 – 5 for the gamma parameter. The final results showed that a c value of 50 and a gamma of 5 resulted in the highest accuracy of 95.65%. Thus, the analysis of research data using the SVM method uses the values c = 50 and gamma = 5. Based on Table 4, it can be seen that the goodness of the model obtained is close to perfect according to the value produced close to 100%. Sensitivity indicates a measure of how well a classification method correctly classifies per category. The greater the sensitivity value and the closer to the value of 100, the better the related classification method. In general, the goodness value of a model is indicated by its accuracy value.

Table 5. Goodness of K Value KNN

K-value	Accuracy	Time
2	75%	5.28 s
4	72%	5.47 s
6	68%	5.48 s
8	72%	5.75 s

Table 6. Goodness of Model KNN

Varieties	Accuracy	Sensitivity
Inpari	75.86%	90.00%
Inpara		50.00%
Hipa		40.00%
Inpago		83.33%

Before carrying out the KNN classification process, first determine the value of K. The value of K is tried from 2 to 10. The selection of the K-value range is based on the accuracy value which continues to decrease as the K-value increases. The process of finding the best K-value can be seen in Table 5. From Table 5, it can be seen that as the number of K-increases, the accuracy decreases, and the analysis process time gets longer. This can happen in the search for the central value of each cluster, so that the larger the number of K, the more often the search for the central value will be carried out, which results in the process will take longer. The classification analysis with KNN then uses

the value of $K = 2$ which produces the highest accuracy value. The best model is selected based on the highest accuracy value. This selected model can then be applied to identify superior rice varieties in Indonesia. These superior rice varieties are chosen based on their above-average quality, with environmental factors such as moisture content and temperature playing a critical role in the selection process [21]. By incorporating these environmental factors, the benefits to the agricultural sector can be further enhanced.

4. Conclusion

The data on superior rice varieties is categorized as categorical data, which makes classification methods the primary approach for managing labeled categorical information. In practice, this data often demonstrates non-linear characteristics. The Support Vector Machine (SVM) algorithm, particularly when using the Radial Basis Function (RBF) kernel, provides a clear and systematic solution for handling non-linear data, with the added benefit of not requiring specific assumptions to be met. The RBF kernel has consistently been shown to deliver higher accuracy compared to other kernels. In fact, SVM with the RBF kernel achieves an accuracy rate 8% higher than that of the K-Nearest Neighbors (KNN) algorithm. This superior performance is further supported by the sensitivity values, where the sensitivity of the KNN model is significantly lower than that of the SVM model. For instance, in over half of the categories, the sensitivity exceeds 50%, indicating that the RBF SVM is a more suitable method for classifying superior rice varieties in Indonesia in 2023. Moreover, in cases involving new superior rice varieties, the RBF SVM can be effectively applied to assist in their classification.

Author Contributions. **Alissa Chintyana:** Conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. **Kertanah:** Validation, writing—review and editing, supervision, funding acquisition. **Siti Hariati Hastuti:** Validation, writing—review and editing, funding acquisition. **Husnul Khotimah:** Software, formal analysis, investigation, resources. All authors discussed the results and contributed to the final manuscript.

Acknowledgement. The authors are thankful Universitas Hamzanwadi for funding this research. The authors are also grateful the editors and reviewers who have supported us in improving this manuscript.

Funding. This research was funded by Universitas Hamzanwadi.

Conflict of interest. The authors declare that there are no conflicts of interest related to this article.

References

[1] A. Faqih, "Analisis Komoditas Unggulan Sektor Pertanian", *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 7, no. 4, pp. 550-559, 2021, doi: 10.29210/020211242.

[2] W. Waluyo, S. Suparwoto, J. Johanes, and N. Wahyu, "Development of Production of Sources of New Upper Variety of Seeds of Rice Results in South Sumatera Province", *Jurnal KaliAgri*, vol. 3, no. 2, pp. 51-60, 2022, doi: 10.56869/kaliagri.v3i2.413.

[3] A. Rachman, M. T. Furqon, and F. Ramdani, "Klasifikasi Varietas Unggul Padi menggunakan Algoritma C4.5", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 9, pp. 9244-9249, 2019.

[4] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3^{rd} ed. Boston: Morgan Kaufmann Publisher, 2012.

[5] M. Arhami and M. Nasir, *Data Mining Algoritma dan Implementasi*, Yogyakarta: Penerbit ANDI, 2020.

[6] Y. P. Baita, "Perbandingan Algoritme Klasifikasi untuk Prediksi Kinerja Siswa di Kelas", *INFOS Journal-Information System Journal*, vol. 1, no. 4, pp. 1–4, 2019.

[7] I. K. Hasan, R. Resmawan, and J. Ibrahim, "Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa," *Indonesian Journal of Applied Statistics*, vol. 5, no. 1, pp. 58–66, May 2022, doi: 10.13057/ijas.v5i1.58056.

[8] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi menggunakan Metode KNN pada Data Iris", *Jurnal Sains dan Edukasi Sains*, vol. 5, no. 4, pp. 28-37, 2022, doi: 10.24246/juses.v5i1p28-37.

[9] I. F. Yulianti, "Penerapan Metode SVM dan BPNN dalam Pengklasifikasian PUS di Jawa Barat", *Jurnal Statistika dan Aplikasinya*, vol. 4, no. 1, pp. 23-34, 2020.

[10] B. Clarke, E. Fokoue, and H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*. New York: Springer, 2009.

[11] A. C. Imanda, N. Hidayat, and M. T. Furqon, "Klasifikasi Kelompok Varietas Unggul Padi menggunakan Modified K-Nearest Neighbor", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 8, pp. 2392-2399, 2018.

[12] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter", *Jurnal Informatika*, vol. 6, no. 2, pp. 212-218, 2019, doi: 10.31294/ji.v6i2.5129.

[13] P. R. Sihombing, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia", *Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 417-426, 2021, doi: 10.30812/matrik.v20i2.1174.

[14] I. Asana and N. Yanti, "Sistem Klasifikasi Pengajuan Kredit dengan Metode Support Vector Machine", *Jurnal Sistem Cerdas*, vol. 6, no. 2, pp. 123-133, 2023.

[15] T. Wurijanto, H. B. Setiawan, and A. B. Tjandrarini, "Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko menggunakan Algoritma Support Vector Machine", *Jurnal Ilmiah Scroll: Jendela Teknologi Informasi*, vol. 10, no. 1, pp. 1-6, 2022, doi: 10.30640/ejournalscroll.v10i1.291.

[16] V. D. Yunanda and N. Hendrastuty, "Perbandingan Kernel Polynomial dan RBF pada Algoritma SVM untuk Analisis Sentimen Skincare di Indonesia", *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 726-735, 2024.

[17] K. Karthikeya and K. H. Sudarshan, "Prediction of Agriculture Crops using KNN Algorithm", *International Journal of Multidisplinary Educational Research*, vol. 10, no. 5, pp. 36-39, 2021.

[18] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max Normalization and Z-score Normalization in the K-Nearest Neighbor (KNN) Algorithm to Test the Accuracy of Types of Breast Cancer", *International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13-20, 2021.

[19] R. A. Johnson and G. K. Bhattacharyya, *Statistics Principles and Methods*, 6^{th} ed. Hoboken: John Wiley & Sons, Inc.

[20] A. Chintyana, A. Choiruddin, and Sutikno, "Cox Point Process with Ridge Regularization: A Better Approach for Statistical Modeling of Earthquake Occurrences," in *Soft Computing in Data Science*, 2023, pp. 163–177, doi: 10.1007/978-981-99-0405-1_12.

[21] P. P. V. T. Pertanian, *Deskripsi Varietas Tanaman yang telah Dilepas Triwulan IV Tahun 2023*. Jakarta: Kementerian Pertanian, 2023.