Classifying Disadvantaged Districts/Cities in Indonesia: A Support Vector Machine Approach

Wanda Surianto(1), Lia Mauliani(2), Ridha Ferdhiana (3), Nurhasanah(4)

[1,2,3,4] Department of Statistics, Universitas Syiah Kuala, Banda Aceh

Jl. Syech Abdurrauf No. 3 Kopelma Darussalam, Banda Aceh, Aceh

e-mail: wanda.surianto@usk.ac.id(1), lia_liani@mhs.usk.ac.id(2), ridha.ferdhiana@usk.ac.id(3), nurhasanah@usk.ac.id(4)

ABSTRAK

Kesenjangan antara wilayah di Indonesia ditandai dengan adanya istilah daerah tertinggal dan daerah tidak tertinggal. Pemerintah menetapkan status daerah tertinggal setiap lima tahun sekali, surat keputusan peraturan presiden no 63 tahun 2020 menetapkan 62 kabupaten/kota di Indonesia sebagai daerah tertinggal. Penelitian ini bertujuan untuk mengklasifikasikan status daerah tertinggal menggunakan algoritma *Support Vector Machine* (SVM) dengan tiga jenis kernel: linear, polinomial, dan *Radial Basis Function* (RBF). SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi dan tugas klasifikasi non-linear. Dataset yang digunakan berasal dari BPS dan JDIH tahun 2022, mencakup 20 variabel yang merepresentasikan indikator sosial ekonomi, infrastruktur, dan layanan publik. Distribusi data menunjukkan ketidakseimbangan, dengan hanya 62 dari 514 kabupaten/kota yang dikategorikan sebagai daerah tertinggal. Parameter optimal ditentukan secara eksperimental: kernel linear ($C = 0{,}1$), polinomial ($C = 1$, $d = 3$), dan RBF ($C = 1$, $\gamma = 0{,}1$). Berdasarkan hasil evaluasi, kernel linear memberikan performa terbaik pada dataset yang digunakan, dengan akurasi 0,94, presisi 0,91, recall 0,81, dan skor F1 sebesar 0,85. Model mengklasifikasikan 45 kabupaten/kota sebagai tertinggal dan 469 sebagai tidak tertinggal. Sebanyak 29 kabupaten/kota menunjukkan perbedaan klasifikasi dibandingkan dengan klasifikasi resmi pemerintah. Perbedaan ini dapat mencerminkan perubahan kondisi di lapangan atau keterbatasan dalam kriteria kebijakan, yang menunjukkan potensi pendekatan berbasis data untuk mendukung perencanaan pembangunan wilayah yang lebih tepat sasaran dan berkeadilan.

Kata kunci: Daerah tertinggal; Klasifikasi; Kernel; Support Vector Machine (SVM)

ABSTRACT

The terms "underdeveloped" and "non-underdeveloped" regions highlight the gap between regions in Indonesia. The government determines the status of underdeveloped regions every five years. Presidential Decree No. 63 of 2020 determines 62 districts/cities in Indonesia as underdeveloped regions. This study aims to classify disadvantaged district status using the Support Vector Machine (SVM) algorithm with three kernel types: linear, polynomial, and Radial Basis Function (RBF). SVM was selected for its effectiveness in handling high-dimensional data and non-linear classification tasks. The dataset, sourced from BPS and JDIH in 2022, comprises 20 variables covering socioeconomic, infrastructure, and public service indicators. The data distribution is imbalanced, with only 62 out of 514 districts labeled as disadvantaged. Optimal parameters were determined experimentally: linear ($C = 0.1$), polynomial ($C = 1$, $d = 3$), and RBF ($C = 1$, $\gamma = 0.1$). Based on evaluation results, the linear kernel achieved the best performance on the given dataset, with an accuracy of 0.94, precision of 0.91, recall of 0.81, and F1-score of 0.85. The model classified 45 districts as disadvantaged and 469 as non-disadvantaged. A total of 29 districts showed discrepancies compared to the official classification. These differences may indicate either changing ground conditions or limitations in policy criteria, highlighting the potential of data-driven approaches to support more targeted and equitable regional development planning.

Keywords: Disadvantaged regions; Classification; Kernel; Support Vector Machine (SVM)

INTRODUCTION

Disadvantaged regions refer to regencies whose areas and communities are underdeveloped compared to other regions on a national scale [1]. The designation of disadvantaged status is determined by the government through a Presidential Regulation (Perpres) issued every five years. The most recent regulation, Perpres No. 63 of 2020, identifies 62 districts/cities as disadvantaged, with the next update scheduled for 2025 [2]. These disparities may be driven by various factors, including limited economic activity, low human resource quality, inadequate infrastructure, high incidence of disasters and conflicts, and restricted access to transportation, telecommunications, and information services [3]. These multidimensional issues necessitate urgent and coordinated responses from both local and central governments to ensure social and economic equity and to prevent regions from falling behind.

One of the government's efforts to address this issue is the issuance of Presidential Regulation No. 63 of 2020 concerning Disadvantaged, Frontier, and Outermost (3T) regions. According to this regulation, 62 districts/cities are classified as disadvantaged regions [2]. Through the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration, several intervention programs have been implemented, improving the status of some districts. However, with ongoing changes in regional administrative structures such as expansion, creation, and mergers, continuous reassessment of regional development status is necessary.

A region is categorized as disadvantaged or not based on a composite index that includes three key components: the Social Resilience Index, the Economic Resilience Index, and the Ecological/Environmental Resilience Index [4]. Additionally, a region's ability to manage resources particularly in terms of potential, information/value, innovation, and entrepreneurship supports its progress toward becoming non-disadvantaged [5].

Various statistical and machine learning methods can be applied to classification analysis, including Support Vector Machine (SVM) [6]. SVM is a technique that separates two classes of data by finding the optimal hyperplane that maximizes the margin between them [7][8]. With kernel functions such as linear, polynomial, and radial basis function (RBF), SVM can handle both linear and non-linear classification problems, particularly in binary classification tasks [9].

A study in Maluku Province applied SVM with a linear kernel (C = 1) and eight variables, achieving 76.31% accuracy in classifying districts/cities [10]. Another study using the Naïve Bayes algorithm on 208 districts/cities yielded 90.5% accuracy with an 80:20 train-test split [11]. However, these methods have limitations Naïve Bayes assumes feature independence, and decision trees may overfit with high-dimensional data. Therefore, this study investigates whether SVM can effectively classify the disadvantaged status of all districts/cities in Indonesia using publicly available indicators. This research is titled *"Classification of Disadvantaged District/City Status in Indonesia Using Support Vector Machine (SVM)"*.

METHOD

Data Source

This study uses secondary data sourced from the 2022 publication of the Central Statistics Agency (BPS) and Presidential Regulation No. 63 of 2020 issued by the Government of the Republic of Indonesia. The regulation is available through the Regulatory Database of the Legal Documentation and Information Network (JDIH BPK). The dataset comprises 514 observations, covering all districts and cities across 34 provinces in Indonesia.

Research Variable

This study utilizes a total of 20 variables, which include one dependent variable (denoted as Y) and nineteen independent variables (denoted as X_1 to X_{19}). The independent variables are categorized into three composite indices: the Social Resilience Index, comprising variables X_1 to X_7; the Economic Resilience Index, comprising variables X_8 to X_{15}; and the Ecological/Environmental Resilience Index, comprising variables X_{16} to X_{19}.

Table 1. Research Variables

Variable	Variable Name	Unit
Y	Disadvantaged Region Status	Classification label (0 = Disadvantaged; 1 = Non-disadvantaged)
X_1	Expected Years of Schooling	Years
X_2	Life Expectancy	Years
X_3	Mean Years of Schooling	Years
X_4	Number of Midwives	Persons
X_5	Number of Community Health Centers	Units
X_6	Number of Family Planning Clinics	Units
X_7	Percentage of Households with Electricity	Percent
X_8	Per Capita Expenditure	Rupiah
X_9	Open Unemployment Rate	-
X_{10}	Percentage of Population Living in Poverty	Percent
X_{11}	Gross Regional Domestic Product	Rupiah
X_{12}	Per Capita Food Expenditure	Rupiah
X_{13}	Poverty Depth Index	-
X_{14}	Poverty Severity Index	-
X_{15}	Number of Markets	Units
X_{16}	Number of Earthquake Events	Events
X_{17}	Number of Landslide Events	Events
X_{18}	Number of Flood Events	Events
X_{19}	Paved Roads	Km

Analysis Methods

Classification is applied when the target variable in a study is categorical. Its primary objective is to predict labels or assign data to specific categories [12]. When the aim is to uncover the predictive structure of a problem, classification focuses on identifying the variables or interactions among variables that contribute to a particular outcome [13].

In machine learning, algorithms often adopt a binary classification framework, which involves only two target classes. For multi-class problems, one class is treated as positive and the others as negative, with the process repeated for each class [14]. Support Vector Machine (SVM) aims to construct a hyperplane that separates data points with positive and negative labels. Rather than using any separating line, SVM seeks the hyperplane that maximizes the margin the shortest distance to the nearest data points, known as support vectors which defines the optimal boundary between classes

[15]. By using kernel functions, SVM is capable of projecting data into a higher-dimensional space to facilitate the separation of classes that cannot be linearly separated in the original space [16].

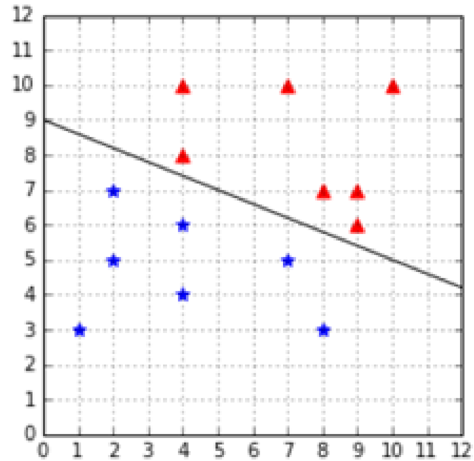

Figure 1. Data partitioning with a hyperplane

Figure 1 illustrates a hyperplane that separates two classes, represented as the positive class (+1) and the negative class (−1). The hyperplane in the above illustration can be expressed by Equation (1), while Equations (2) and (3) indicate that the illustrated method divides the dataset into positive and negative classes..

$$\boldsymbol{wx} + b = 0 \tag{1}$$

$$\boldsymbol{wx} + b > 0 \tag{2}$$

$$\boldsymbol{wx} + b < 0 \tag{3}$$

The weight vector (W) represents the line perpendicular to the origin and the hyperplane. The bias (b) refers to the position of the hyperplane relative to the origin. Equations (4) and (5) below are used to calculate the values of **b** and **W**.

$$b = -1/2\ (\boldsymbol{w}.\boldsymbol{x}^{+} + \boldsymbol{w}.\boldsymbol{x}^{-}) \tag{4}$$

$$\boldsymbol{w} = \sum_{i=1}^{n} a_i y_i \boldsymbol{x_i} \tag{5}$$

H_1 is the supporting hyperplane for the +1 class, represented by the function $wx_1 + b = +1$, while H_2 is the supporting hyperplane for the −1 class. To determine the optimal hyperplane between the two classes, Equations (7) and (8) is used.

$$Margin = |dH_1 - dH_2| = 2\frac{2}{\|W\|} \tag{6}$$

$$Minimize\ j_1[\boldsymbol{w}] = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N} \xi_i \tag{7}$$

$$yi(xi.\boldsymbol{w} + b) - 1\ \geq 0, i = 1, \dots, n \tag{8}$$

The value of the support vectors influences the level of accuracy in determining the most suitable method the higher the support vector value, the greater the resulting accuracy. Therefore, parameter selection becomes a critical aspect of the problem. A trial and error technique is employed to obtain the optimal accuracy value [17].

Analysis Steps

The data analysis in this study was performed using Microsoft Excel and Python, supported by the Google Colab programming platform. The analysis was conducted through the following stages:

1. Data cleaning;
   The dataset used in this study contains fewer than 100 missing values per variable. To address this issue, the missing values were imputed using the mean value of each respective variable.
2. Descriptive statistical analysis;
3. Data normalization using Z-Score Standardization to standardize the variables;
   Data normalization transforms variable values to a standard scale to prevent dominance caused by scale differences. One commonly used method is Z-score Standardization, which transforms the data so that it has a mean of 0 and a standard deviation of 1 [23]. Z-normalization for a value x in a dataset can be calculated using the equation (9).
   $$x_i' = \frac{x_i - \mu}{\sigma} \tag{9}$$
4. Dataset split;
   To obtain a more accurate model with better generalization capability, the training and evaluation processes were carried out using K-Fold Cross Validation. In this method, the dataset is divided into K equal parts (folds), where each fold is used as the testing set in turn, while the remaining folds are used for training. This approach ensures that all data are used for both training and testing, leading to more stable evaluation results that are not dependent on a single train-test split. In this study, data splitting was performed using K = 5 and K = 10 as variations.
5. Inferential analysis using the Support Vector Machine (SVM) method, which involved the application of three kernel functions:
   Linear function: The linear kernel can be expressed in the following equation:
   $$\phi\left(x_i, x_j\right) = x_i^T x_j \tag{10}$$
   $K\ (x, x')$ is a kernel function where xxx and $x.x'$ are feature vectors of two data points to be processed. This linear kernel function is the simplest type of kernel, representing the dot product of the two vectors [18][20]. The parameter **C** values used are 0.1, 1, and 10. The selection of these parameters is based on a previous study on the classification of disadvantaged regions [22].
   Polynomial Kernel Function: The polynomial kernel has two parameters: c, which represents the constant term, and d, which denotes the degree of the kernel. The polynomial kernel equation can be written as follows:
   $$\phi\left(x_i, x_j\right) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \tag{11}$$
   $\boldsymbol{x}$ and $\boldsymbol{x'}$ are the feature vectors of two data points to be processed, while d represents the predefined degree of the polynomial. As the degree increases, the kernel function maps the input data into higher-dimensional space, enabling the SVM to perform non-linear classification on more complex datasets. The values of parameter **C** used are 0.1, 1, and 10, while the values of parameter **d** used are 2, 3, and 4. The selection of these parameters follows a previous study on the classification of disadvantaged regions [22].
   Radial basis function (RBF): RBF kernel projects feature vectors into an infinite-dimensional space using the parameter γ (gamma). In the SVM test, treatment using RBF produces a higher

level of accuracy. This is because usually linear and polynomial kernels take less time and provide lower accuracy than rbf or Gaussian kernels [19]. However, unlike the polynomial kernel, the RBF kernel is less dependent on the degree of the function, which makes it more effective in addressing overfitting issues commonly encountered in polynomial kernels [21]. The following equation represents the Radial Basis Function (RBF) kernel.

$$\phi(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \quad (12)$$

This equation indicates that the closer $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the larger the resulting value. The RBF kernel has the property that higher weights are assigned to data points that are closer together, and these weights decrease exponentially as the distance between the points increases. The values of parameters **C** and **γ** used in the RBF kernel are 0.1, 1, and 10. The selection of these parameters considers the combinations of **C** and **γ** values used in previous studies [22].

6. Result interpretation and conclusion

   In classification tasks, model performance is commonly evaluated using several metrics, including the confusion matrix, accuracy, precision, and recall.

RESULT AND DISCUSSION

Descriptive Analysis

Descriptive analysis provides a general overview of the data. Indonesia consists of 514 districts/cities across 34 provinces. According to Presidential Regulation No. 63 of 2020 on 3T regions, 62 districts/cities (12%) are classified as disadvantaged, while the remaining 452 (88%) are non-disadvantaged. This shows that the majority of regions fall into the non-disadvantaged category. The distribution highlights regional disparities and can guide policymakers in formulating development strategies. The descriptive statistics are presented in Table 2.

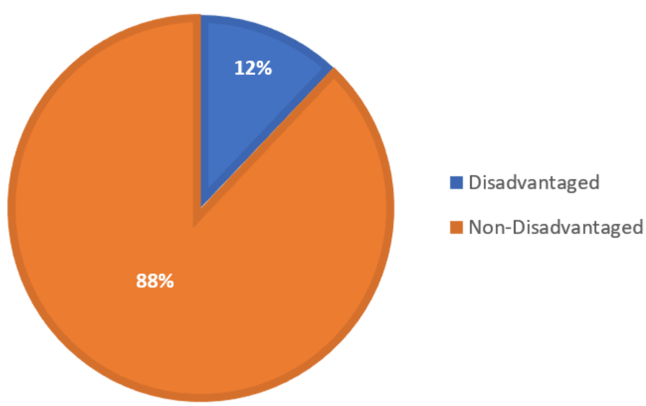

Figure 2. The disadvantaged region status of districts/cities in Indonesia.

The province with the highest number of disadvantaged districts/cities is Papua, with a total of 22, followed by East Nusa Tenggara, which has 13 disadvantaged districts/cities. Furthermore, there are three provinces Papua, Maluku, and East Nusa Tenggara where the number of disadvantaged regions exceeds the number of non-disadvantaged ones. Java and Kalimantan Island have no disadvantaged districts/cities, while Sumatra Island has only a small number. Overall, it can be observed that disadvantaged districts and cities are predominantly concentrated in the eastern region of Indonesia.

Table2. Descriptive Statistics of Study

Variable	Minimum	Maximum	Mean
X_1	4.07	17.81	13.09
X_2	55.70	77.82	69.93
X_3	1.58	13.03	8.55
X_4	0.00	10281.00	641.33
X_5	3.00	101.00	22.26
X_6	7.00	1060.00	76.00
X_7	27.70	100.00	97.54
X_8	4190.00	24221.00	10643.32
X_9	0.12	75.64	5.96
X_{10}	2.28	42.03	11.68
X_{11}	255.92	72961651.00	591218.80
X_{12}	43.84	5740045.00	601974.40
X_{13}	0.17	13.90	1.97
X_{14}	0.02	7.59	0.53
X_{15}	0.00	137.00	19.35
X_{16}	0.00	11.00	0.12
X_{17}	0.00	103.00	1.68
X_{18}	0.00	30.00	3.16
X_{19}	0.00	105589.00	477.30

According to BPS, the national average for expected years of schooling in Indonesia is 12.85 years, which closely aligns with the dataset average of 13.09 years. However, there is a wide disparity across regions, with values ranging from as low as 4.07 years to as high as 17.81 years. This indicates that while some districts still experience very limited access to education, others have expectations reaching higher education levels. Such gaps reflect serious inequalities in educational opportunities.

Similar disparities are observed in other indicators within the Social Resilience Index. The number of midwives ranges from 0 to 10,281, and community health centers from 1 to 1,010 units highlighting unequal access to basic health services. In contrast, life expectancy shows relatively low variability, ranging from 55.70 to 77.82 years, with an average of 69.93 years. This suggests that despite differences in health infrastructure, longevity remains fairly consistent. Overall, these descriptive statistics emphasize the need for targeted classification and intervention to strengthen social resilience in underdeveloped regions.

Selection of the Best Ratio and Parameters

Following the evaluation of various parameters across two different ratios for each kernel function, the subsequent step involves selecting the parameters that demonstrate the best performance based on the classification test results. The optimal parameters are determined using classification performance indicators, namely accuracy, precision, recall, and f1 score.

Table 3. Best Parameters for Linear Kernel Function

K-folds	Parameters	Classification Performance			
		accuracy	precision	recall	f1 score
10-folds	$c = 0.1$	0,94231	0,97	0,85	0,89
	$c = 1$	0,92308	0,88	0,88	0,88
	$c = 10$	0,94231	0,90	0,93	0,91
5-folds	$c = 0.1$	0,96117	0,98	0,88	0,92
	$c = 1$	0,95146	0,92	0,89	0,91
	$c = 10$	0,93204	0,88	0,86	0,87

Based on Table 3, it can be seen that the best parameter for classifying districts/cities in Indonesia using the linear kernel is C = 10 with 10-folds, while for the 5-folds, the best C parameter for classifying the status of disadvantaged districts/cities is 0.1.

Table 4. Best Parameters for Polynomial Kernel Function

K-folds	Parameters	Classification Performance			
		accuracy	precision	recall	f1 score
10-folds	$d = 2; c = 0.1$	0,86538	0,84	0,69	0,73
	$d = 2; c = 1$	0,88462	0,86	0,74	0,78
	$d = 2; c = 10$	0,88462	0,83	0,78	0,80
	$d = 3; c = 0.1$	0,88462	0,94	0,70	0,75
	$d = 3; c = 1$	0,92308	0,96	0,80	0,85
	$d = 3; c = 10$	0,88462	0,86	0,74	0,78
	$d = 4; c = 0.1$	0,86538	0,84	0,69	0,73
	$d = 4; c = 1$	0,88462	0,86	0,74	0,78
	$d = 4; c = 10$	0,86538	0,80	0,73	0,75
5-folds	$d = 2; c = 0.1$	0,89320	0,88	0,68	0,73
	$d = 2; c = 1$	0,91262	0,95	0,72	0,78
	$d = 2; c = 10$	0,92233	0,88	0,80	0,83
	$d = 3; c = 0.1$	0,90291	0,95	0,69	0,75
	$d = 3; c = 1$	0,92233	0,96	0,75	0,81
	$d = 3; c = 10$	0,92233	0,91	0,78	0,82
	$d = 4; c = 0.1$	0,89320	0,88	0,68	0,73
	$d = 4; c = 1$	0,90291	0,89	0,71	0,76
	$d = 4; c = 10$	0,92233	0,88	0,80	0,83

Based on Table 4, the best parameter combination for the polynomial kernel function in the simulation test is d = 3 and C = 1. By comparing the classification model performance indicators, the same optimal parameter combination is obtained for both 10-folds and 5-folds training and testing data ratios.

Table 5. Best Parameters for RBF Kernel Function

K-folds	Parameters	Classification Performance			
		accuracy	precision	recall	f1 score
10-folds	$\gamma = 0.1; c = 0.1$	0,80769	0,40	0,50	0,45
	$\gamma = 0.1; c = 1$	0,84615	0,80	0,64	0,67
	$\gamma = 0.1; c = 10$	0,86538	0,84	0,69	0,73
	$\gamma = 1; c = 0.1$	0,80769	0,40	0,50	0,45
	$\gamma = 1; c = 1$	0,80769	0,48	0,50	0,45
	$\gamma = 1; c = 10$	0,80769	0,40	0,50	0,45
	$\gamma = 10; c = 0.1$	0,80769	0,40	0,50	0,45
	$\gamma = 10; c = 1$	0,80769	0,40	0,50	0,45
	$\gamma = 10; c = 10$	0,80769	0,40	0,50	0,45
5-folds	$\gamma = 0.1; c = 0.1$	0,84466	0,42	0,50	0,46
	$\gamma = 0.1; c = 1$	0,90291	0,95	0,69	0,75
	$\gamma = 0.1; c = 10$	0,88350	0,82	0,68	0,72
	$\gamma = 1; c = 0.1$	0,84466	0,42	0,50	0,46
	$\gamma = 1; c = 1$	0,84466	0,42	0,50	0,46
	$\gamma = 1; c = 10$	0,84466	0,42	0,50	0,46
	$\gamma = 10; c = 0.1$	0,84466	0,42	0,50	0,46
	$\gamma = 10; c = 1$	0,84466	0,42	0,50	0,46
	$\gamma = 10; c = 10$	0,84466	0,42	0,50	0,46

Based on Table 5, the optimal parameters for the RBF kernel were $\gamma = 0.1$ and C = 10 for both 10-fold and 5-fold dataset splits. However, the best overall performance was achieved using the linear kernel with C = 0.1 on the 5-fold dataset, resulting in an accuracy of 0.96117, precision of 0.98, recall of 0.88, and F1 score of 0.92.

Table 6. Best Model Confusion Matrix (Linear Kernel, c = 0.1)

Actual	Prediction		Total
	disadvantaged	non-disadvantaged	
Disadvantaged	39	23	62
Non-disadvantaged	6	446	452
Total	45	469	514

After evaluating all dataset split configurations, the linear kernel with C = 0.1 was identified as the most effective for classifying disadvantaged districts/cities in Indonesia using the SVM algorithm. The best-performing model was the linear kernel, as determined by the SVM-based analysis conducted in this study [24]. Based on Table 6, the model correctly classified 39 districts/cities as disadvantaged and 446 as non-disadvantaged, while 23 disadvantaged areas were misclassified as non-disadvantaged and 6 non-disadvantaged areas were misclassified as disadvantaged. Using this optimal setting, the model achieved an accuracy of 0.94, precision of 0.91, recall of 0.81, and an F1 score of 0.85.

CONCLUSION

This study identified the optimal kernel functions and parameter configurations for classifying the disadvantaged status of districts and cities in Indonesia using the Support Vector Machine (SVM) algorithm. The best-performing settings were c = 0.1 for the linear kernel, demonstrated the highest performance, achieving an accuracy of 94%, precision of 91%, recall of 81%, and an F1 score of 85%. When compared to the official 2020 classification outlined in Presidential Regulation No. 63, the SVM model predicted a reduction in the number of disadvantaged regions from 62 to 45 in 2022, with 29 districts/cities showing a change in status 23 improved to non-disadvantaged, while 6 previously non-disadvantaged areas were reclassified as disadvantaged. These findings underscore the potential of SVM as a valuable tool to support policy decisions and monitor regional development progress. However, the study is not without limitations, as it relies solely on cross-sectional data from a single year (2022), without incorporating temporal dynamics. To address this, future research should consider using longitudinal data, techniques to improve model interpretability, and applying ensemble methods to enhance robustness. Furthermore, involving policy stakeholders through feedback mechanisms could increase the model's practical relevance and effectiveness in guiding targeted development interventions and early identification of at-risk regions.

REFERENCE

[1] Pemerintah Indonesia. Peraturan Presiden Republik Indonesia No 131 Tahun 2015 Tentang Penetapan Daerah Tertinggal Tahun 2015-2019. Sekretariat Negara. Jakarta. 2015.

[2] Pemerintah Indonesia. Peraturan Presiden Republik Indonesia Nomor 63 Tahun 2020 tentang Penetapan Daerah Tertinggal Tahun 2020–2024. Jakarta: Sekretariat Negara Republik Indonesia. 2020.

[3] E. A. Sari, Meilani T, I. A Shariati, S. Sofyan, R. A. Baihaqi, R. Nooraeni. Klasifikasi Kabupaten Tertinggal Di Kawasan Timur Indonesia Dengan Support Vector Machine. JIKO (Jurnal Informatika dan Komputer). Vol. 3, No. 3, pp 188-195. 2020.

[4] Direktorat Jenderal Pembangunan Desa dan Perdesaan. Indeks Desa Membangun. Jakarta. 2020.

[5] Hamidi, H., Harioso, & Huda. Indeks Desa Membangun. Kementrian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi, Jakarta. 2015.

[6] R. Primartha, Belajar Machine Learning Teori dan Praktik. Bandung : Informatika Bandung, 2018

[7] Mariyam, P. Ana, Ratnawati, D. Eka and Wahyu, Widodo Agis. Klasifikasi Penyakit Gigi dan Mulut Menggunakan Metode Support Vector Machine. Pengembangan Teknologi Informsi dan Ilmu Komputer. Vol. 2, pp. 802-810. 2018.

[8] W. C. Hsu, C. C. Chang and C. J. Lin. A Practical Guide to Support Vector Machine., Taipei: Departement of Computer Science National Taiwan University. 2014.

[9] Domingos, P. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Book, New York. 2015.

[10] Palisoa, N. F., Sinay, L. J., & Matdoan, M. Y. Penerapan Support Vector Machine (SVM) untuk Klasifikasi Kabupaten Tertinggal di Provinsi Maluku. Jurnal Matematika, Statistika dan Terapannya, vol. 02, pp. 79–86. 2023.

[11] Lidya, W., Yozza, H., & Yanuar, F. Klasifikasi Daerah Tertinggal Di Indonesia Menggunakan Metode Naive Bayes Classifier Yanuar. Jurnal Matematika UNAND, IX (1), pp 23–29. 2020.

[12] Wizner, W. Python programming for beginners: 3 books in 1. Springer, London. 2020.

[13] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. Classification and Regression Trees. Chapman & Hall, London. 2020.

[14] T. Jo, Machine learning foundations: Supervised, unsupervised, and advanced learning. Korea: Springer International Publishing. 2021.

[15] C. C. Aggarwal, Data Classification Algorithms and Applications. New York: CRC Press, 2015.

[16] Yalsavar, M., Karimaghaee, P., Sheikh-Akbari, A., Khooban, M.-H., Dehmeshki, J., Al- Majeed, S. Kernel parameter optimization for support vector machine based on sliding mode control. IEEE Access 10, 17003–17017. 2022.

[17] J. A. K. Suykens, M. Signoretto, and A. Argyriou. Regularization, Optimization, Kernels, and Support Vector Machines. 2014.

[18] I Nyoman Setiawan, Robert Kurniawan, Budi Yuniarto, Rezzy Eko Caraka, Bens Pardamean, Parameter Optimization of Support Vector Regression Using Harris Hawks Optimization, Procedia Computer Science, vol. 179, pp. 17-24. 2021.

[19] Wiryawanto T. M. P., Hawani Z., and Ramadhani M. A. Comparison of Support Vector Machine (SVM) and Autoregressive Integrated Moving Average (ARIMA) Methods for Predicting Air Quality Using Python and KNIME", *J Statistika*, vol. 16, no. 1, pp. 384–394, Jul. 2023

[20] M. Athoillah, E. Purnaningrum, and R. K. Putri, "Modified Multi-Kernel Support Vector Machine for Mask Detection," CommIT (Communication and Information Technology) Journal, vol. 16, no. 2, pp. 159–166, 2022.

[21] Cambell, C., & Ying, Y. Learning with Support Vector Machines : Synthesis Lecturers on Artificial Intelligence and Machine Learning. Morgan & Claypool. 2011.

[22] Al-Azies, H., & Anuraga, G. Klasifikasi Daerah Tertinggal di Indonesia Menggunakan Algoritma SVM dan k-NN. Jurnal Ilmu Dasar, 22(1), 31–38. 2021.

[23] James, G., Witten, D., Hastie, T., & Tibshirani, R. An Introduction to Statistical Learning. Springer, London. 2017.

[24] Roshafara F. Forecasting Average Rice Prices at Milling Level According to Quality Using Support Vector Regression ", J Statistika, vol. 17, no. 1, pp. 664–671, Jul. 2024.