Vol. 4 No. (1) (June 2026) 43-54

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Classification of Village Development Index Categories in North Maluku Province Using an ADASYN-Based Multiclass Support Vector Machine

Lea Angelina[1], Tiani Wahyu Utami*[2], Saeful Amri[3]
[1,2]Department of Statistics, Universitas Muhammadiyah Semarang, Indonesia
[3]Department of Data Science, Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v4i1.1154

Article Info

Article history:
Received May 18, 2026
Revised May 19, 2026
Accepted June 11, 2026

Keywords:
ADASYN; Classification; Imbalanced Data; Multiclass Support Vector Machine; North Maluku; Village Development Index.

Abstract

Class imbalance is a common challenge in machine learning that may lead to biased predictions and reduced classification performance, particularly for minority classes. This study implements the integration of Adaptive Synthetic Sampling (ADASYN) and Multiclass Support Vector Machine (SVM) to classify the 2024 Village Development Index (IDM) in North Maluku Province. North Maluku Province was selected because it ranks last among the ten developing provinces in Indonesia, with an average IDM score of 0.6159. Furthermore, out of a total of 684 villages, only six have achieved independent status, resulting in extreme disparities and posing a challenge to the accurate classification of village development status. The dataset consists of 684 villages and was obtained from the 2024 IDM database available on the official website of the Ministry of Villages, Disadvantaged Regions, and Transmigration of the Republic of Indonesia. The best-performing model was a linear-kernel SVM with a cost parameter (C) of 100, yielding an accuracy of 98.54%, precision of 98.26%, recall of 99.4%, and an F1-score of 98.83%. Of the total 137 villages evaluated, only two villages were misclassified: Salimuli Village and Dowongimaiti Village. The proposed framework demonstrates the potential of combining ADASYN and SVM for supporting evidence-based village development planning in geographically complex regions.

✉ Correspondence Address:
E-mail: tianiutami@unimus.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Village development is an integral component of Indonesia's national development agenda aimed at achieving equitable distribution of resources and improving the well-being of the broader community. The strategic role of villages is reinforced by Law No. 6 of 2014, which designates villages as the smallest administrative units of government with full authority to manage their own development and serve as a catalyst for local economic growth [1]. The importance of rural development is further highlighted by demographic statistics from Statistics Indonesia (BPS), which report 74093 villages compared to only 8412 urban sub-districts across the country [2]. Despite continuous development efforts, significant regional disparities persist, potentially widening the socioeconomic gap between developed and underdeveloped regions. Therefore, sustainable development based on the optimization of village potential has become the primary strategy for reducing these disparities [3].

To measure progress and determine development priorities in a structured manner, the government has established the Village Development Index (IDM) as a composite evaluation tool that integrates social, economic, and environmental aspects. Based on the aggregation of the Social Resilience Index (IKS), the Economic Resilience Index (IKE), and the Environmental Resilience Index (IKL), villages are classified into five categories: severely underdeveloped, underdeveloped, developing, advanced, and self-reliant [4]. Development programs are often prioritized for developing villages because they are in a transitional stage where targeted interventions can substantially accelerate progress toward higher development status.

In the context of equitable development, North Maluku Province faces highly complex challenges. According to 2024 data from the Indonesian Ministry of Villages, Development of Disadvantaged Regions, and Transmigration, this archipelagic province has the lowest average Human Development Index (HDI) score 0.6159 among provinces classified as developing, making it highly vulnerable to being downgraded to the underdeveloped category [5]. This problem is exacerbated by the uneven improvement in village status; of the 684 existing villages, only six have achieved self-reliant status, while the other eight regencies and cities do not yet have a single self-reliant village. The complexity of setting priorities and these fiscal capacity constraints require an accurate and objective decision-making system to classify IDM status in support of regional development policies.

The need for such a system can be met through classification techniques in data science. One algorithm based on a statistical approach that offers superior performance and good generalization is the Support Vector Machine (SVM) [6]. As a supervised learning method, SVM works by constructing an optimal hyperplane that maximizes the margin between classes. To accommodate IDM classification, which has five categories, this algorithm was developed into Multiclass SVM [7]. The effectiveness of Multiclass SVM in development data has been demonstrated by Latuconsina et al. (2024) in the classification of IDM in Maluku Province, achieving an accuracy of 97.75% using a linear kernel [8].

Although SVM demonstrates superior performance, the primary challenge in classifying IDM in North Maluku is data imbalance, where independent villages account for only 0.9% of the total (an "extreme imbalance" category) [9]. This condition can lead to modeling bias toward the majority class and reduce the prediction accuracy for the minority class. To address this, the Adaptive Synthetic Sampling (ADASYN) oversampling method was applied because it can adaptively generate synthetic data based on the classification difficulty level of the minority class [10]. Several previous studies have confirmed that the integration of ADASYN and Multiclass SVM successfully improves accuracy and recall significantly in various cases of imbalanced data [11].

The effectiveness of integrating ADASYN and SVM in handling imbalanced data has been demonstrated in various studies. Ramadhan (2021) demonstrated that the ADASYN–SVM method achieved an accuracy of 87.3%, higher than the SMOTE–SVM method, which achieved an accuracy of 85.4% in the detection of type 2 diabetes [12]. Additionally, Ramadhani et al. (2025) reported that the application of Multiclass ADASYN–SVM was able to improve the accuracy of traffic accident classification from 70.75% to 96.38%, with precision, recall, and F1-score values each exceeding 96% [13].

Although previous studies have applied Multiclass SVM to Village Development Index classification, limited attention has been given to regions characterized by archipelagic geography, fragmented accessibility, and extreme class imbalance. Furthermore, the effectiveness of ADASYN-based oversampling in addressing severe imbalance in IDM classification remains underexplored. Therefore, this study investigates the integration of ADASYN and Multiclass SVM for IDM classification in North Maluku Province, one of Indonesia's most geographically dispersed island regions.

METHOD

This research method was systematically designed to address the challenges of classifying imbalanced data through a statistics-based data science approach. Broadly speaking, the experimental stages in this study are divided into five main phases: secondary data collection, data preprocessing, class distribution balancing using the Adaptive Synthetic Sampling (ADASYN) method, predictive modeling with the Multiclass Support Vector Machine (SVM) algorithm, and a comprehensive evaluation of model performance. All preprocessing, oversampling, model training, hyperparameter tuning, and evaluation procedures were implemented using Python.

2.1 Data Sources and Research Variables

The data used in this study consists of secondary data in the form of the 2024 Village Development Index (IDM) for North Maluku Province, covering 684 villages. The data was obtained from the official website of the Ministry of Villages, Disadvantaged Regions, and Transmigration of the Republic of Indonesia, accessible via the official webpage https://idm.kemendesa.go.id/.

Table 1. Research Variables

Variable	Description	Consideration for Variable Selection
IDM Status (Y)	1: Independent 2: Advanced 3: Developing 4: Underdeveloped 5: Severely Underdeveloped	These five classification classes have been established by the Ministry of Villages, Disadvantaged Regions, and Transmigration.
IKS (X_1)	Social Resilience Index	Sub-index component of the Village Development Index (IDM)
IKE (X_2)	Economic Resilience Index	Sub-index component of the Village Development Index (IDM)
IKL (X_3)	Environmental Resilience Index	Sub-index component of the Village Development Index (IDM)

2.2 Data Preprocessing

Before moving on to the machine learning modeling phase, the dataset first undergoes preprocessing to ensure data integrity [14]. The steps involved include:

1. Checking for Missing Values: Ensuring there are no empty data points in the 684 sample rows.
2. Label Encoding: Converting the format of the dependent variable (Y) which was originally categorical text data into numbers (1 through 5) in accordance with the computational standards of the SVM algorithm.
3. Data Splitting: The dataset was divided into an 80% training set comprising 547 villages for model development and a 20% testing set comprising 137 villages for evaluating the model's performance. A fixed random seed was used to ensure reproducibility of the experimental results.

2.3 Handling Imbalanced Data ADASYN

The IDM classification case in North Maluku is characterized by extreme class imbalance, with the "Independent" category acting as an extreme minority class (<1%) [15]. To address modeling bias toward the majority class, the Adaptive Synthetic Sampling (ADASYN) method was applied to the training data. ADASYN generates synthetic samples adaptively by focusing on minority instances that are more difficult to classify [16]. The procedure can be summarized through the following mathematical formulation [17].

1. Calculating the imbalance ratio of the minority class relative to the majority class.

$$d = \frac{m_s}{m_l} \tag{1}$$

2. Calculate the total amount of synthetic data (G) that must be generated with an oversampling parameter of $\beta = 0{,}99$.

$$G = (m_l - m_s) \times \beta \tag{2}$$

3. Calculate the difficulty score (r_i) for each minority sample based on the K-nearest neighbors using the Euclidean distance.

$$r_i = \frac{\Delta_i}{K},\ i = 1, \dots, m_s \tag{3}$$

4. Normalize the values of (r_i) to obtain the density distribution weights $(\widehat{r_i})$:

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \tag{4}$$

5. Calculate (g_i), which is the number of new data observations generated for each nearest-neighbor cluster of the minority data.

$$g_i = \hat{r}_i \times G \tag{5}$$

6. Generate new synthetic data (s_i), where λ is a random number between 0 and 1. using the following formula:

$$s_i = x_i + (x_{zi} - x_i) \times \lambda \tag{6}$$

2.4 Multiclass Support Vector Machine (SVM)

Support Vector Machines (SVM) work by finding the optimal hyperplane that maximizes the separation margin between classes [18]. Since the target variable has 5 categories, this algorithm was adapted into a Multiclass SVM using the One-Against-All (OAA) approach. The OAA approach constructs n separate binary SVM models, where each model separates a specific class from the combined set of other classes [19]. Mathematically, this hyperplane can be defined by the following equation:

$$w_i x_i + b = 0 \tag{7}$$

To account for the varying characteristics of data distribution, this study tests and compares three types of basic kernel functions, namely:

1. Linear Kernel: Used when the decision boundary between classes tends to form a straight line. In general, the function of a linear kernel can be expressed as follows [20].

$$K(x_i, x_j) = x_i . x_j \tag{8}$$

2. Polynomial Kernel: Maps data to a high-dimensional space with a complexity parameter (degree). In general, the polynomial kernel function can be expressed as follows [21].

$$K(x_i, x_j) = (x_i . x_j + 1)^h \tag{9}$$

3. Radial Basis Function (RBF) Kernel: Used to handle complex nonlinear data by relying on the gamma parameter (γ). In general, the RBF kernel function can be expressed as follows [22].

$$K(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}), \sigma \in \mathbb{R}^+ \tag{10}$$

2.5 Hyperparameter Optimization

Hyperparameter optimization was conducted using the Grid Search method combined with 10-fold cross-validation on the balanced training data [8]. Determine the range of parameter weights for each kernel function:

1. Linier C = [0.001, 0.01, 0.1, 1, 10, 100],
2. Polynomial C = [0.001, 0.01, 0.1, 1, 10, 100], degree = [2, 3, 4, 5, 6], and
   RBF C = [0.001, 0.01, 0.1, 1, 10, 100], γ = [0.001, 0.01, 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10].

The best parameter combination for each kernel was selected based on the highest average cross-validation score.

2.6 Model Evaluation

Once the best model from each kernel was selected, final testing was conducted using the original test data (20% of the data). Model performance was quantified using a multi-class confusion matrix to calculate four key evaluation metrics, namely [23]:

a. Accuracy: Describes the overall percentage of correct classifications made by the model.

$$Accuracy = \frac{\Sigma(TP)}{Total\ Data} \times 100\% \tag{11}$$

b. Precision: Measures the ratio of true positives to the total number of positive predictions made by the model.

$$Precision = \frac{\sum_{i=1}^{n} precision_i}{n} \times 100\% \tag{12}$$

c. Recall: Describes the proportion of actual positive data that the model correctly identifies.

$$Recall = \frac{\sum_{i=1}^{n} recall_i}{n} \times 100\% \tag{13}$$

d. F1-Score: A harmonic mean that combines the precision and recall metrics as an indicator of model performance.

$$\mathrm{F1} - score = \frac{2 \times recall \times precision}{recall + precision} \times 100\% \quad (14)$$

2.7 Types of Villages

The Village Development Index (IDM) is a tool designed by the Ministry of Villages, Disadvantaged Regions, and Transmigration to classify the development status of villages as a basis for determining priorities in the utilization of village resources. The IDM is based on the accumulation of social, economic, and environmental factors that interact to support sustainable development and enhance the self-reliance of village communities [24]. The determination of village status categories is based on threshold values established by the Indonesian Ministry of Villages, Disadvantaged Regions, and Transmigration, as presented in Table 2 [25].

Table 2. Village Status Categories

Category	**Village Status**	**IDM Score Range**
1	Independent	IDM > 0.8155
2	Advanced	0.7072 < IDM ≤ 0.8155
3	Developing	0.5989 < IDM ≤ 0.7072
4	Underdeveloped	0.4907 < IDM ≤ 0.5989
5	Severely Underdeveloped	IDM ≤ 0.4907

According to data from Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Regulation No. 2 of 2016 on the IDM, the IDM calculation classifies villages in Indonesia into five status categories.

1) Independent Village
   Independent village is one that has developed rapidly and possesses a strong capacity to drive development, expand, and maintain village infrastructure in order to improve community well-being. Such villages are also characterized by sustainable social, economic, and ecological resilience.
2) Advanced Village
   Advanced village is a village that has the capacity to utilize social, economic, and environmental resources to improve community well-being, the quality of life for residents, and reduce poverty levels.
3) Developing Village
   Developing village is a village that has the potential to become a progressive village, with available social, economic, and environmental resources, but its management is not yet optimal in efforts to improve community well-being, the quality of life for residents, and poverty alleviation.
4) Underdeveloped Village
   A Underdeveloped Village is a village that faces vulnerability due to natural disasters, economic shocks, and social conflicts, rendering it unable to manage its potential social, economic, and ecological resources and resulting in various forms of poverty.
5) Severely Underdeveloped Villages
   Severely Underdeveloped villages are those in a vulnerable condition due to natural disasters, economic pressures, or social conflicts, rendering them unable to manage their social, economic, and ecological resources and still facing various forms of poverty.

2.8 Village Development Index (IDM)

The IDM is a composite indicator used by the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration to assess a village's level of progress based on three main dimensions, namely [26]:

1. Social Resilience Index (IKS)
   The IKS measures the social resilience of rural communities, covering aspects of social capital, health, education, and housing.
2. Economic Resilience Index (IKE)
   The IKE indicates a rural community's ability to withstand economic pressures through economic strength, regional access, financial institutions, logistics distribution, and trade hubs.
3. Environmental Resilience Index (IKL)
   The IKL reflects a village's capacity to manage the environment sustainably, including environmental health and preparedness for potential disasters.

RESULTS AND DISCUSSION

The results and discussion section presents the findings obtained from the implementation of the ADASYN-based Multiclass Support Vector Machine (SVM) model for classifying the Village Development Index (IDM) in North Maluku Province. This section covers descriptive analysis of the dataset, preprocessing results, handling of imbalanced data using ADASYN, hyperparameter optimization, and evaluation of classification performance using several metrics, including accuracy, precision, recall, and F1-score. The discussion further explains the effectiveness of each SVM kernel in classifying IDM categories and interprets the resulting confusion matrix.

3.1 Descriptive Statistics

Descriptive analysis was conducted to provide an overview of the Village Development Index (IDM) classification in North Maluku Province in 2024. The distribution of village categories is presented in Figure 1.

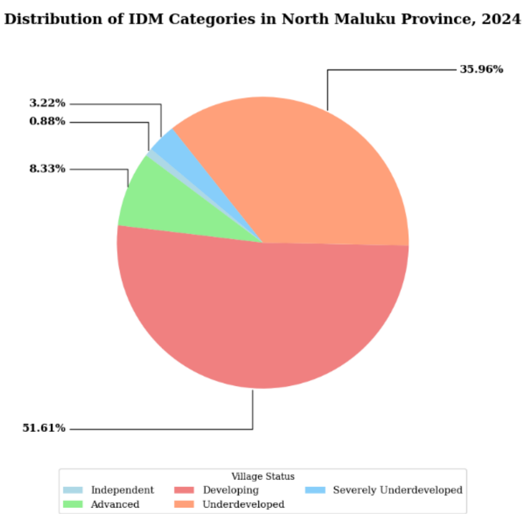

Figure 1. Distribution of IDM Categories in North Maluku Province

Based on Figure 1, most villages were classified as developing villages (51.61%), followed by underdeveloped villages (35.96%). Meanwhile, advanced villages accounted for 8.33%, severely underdeveloped villages 3.22%, and independent villages only 0.88%. This distribution clearly indicates a severe class imbalance problem, particularly for the independent village category, which represents less than 1% of the total observations.

These findings suggest that most villages in North Maluku are still in the transition stage toward higher development status. The very small proportion of self-reliant villages indicates that development acceleration and policy intervention are still required, particularly in strengthening social, economic, and environmental resilience in rural areas. Furthermore, descriptive statistics of the predictor variables used in this study are presented in Table 3.

Table 3. Descriptive Statistics of Research Variables

Variable	Mean	Median	Min	Max	Std Dev	Variance
IKS	0.6808	0.68	0.4514	0.9771	0.0818	0.0067
IKE	0.4803	0.4667	0.15	0.9333	0.1416	0.02
IKL	0.6850	0.6667	0.2	1	0.1201	0.0144

Table 3 shows that the Social Resilience Index (IKS) and Environmental Resilience Index (IKL) had relatively higher average values compared to the Economic Resilience Index (IKE). This finding indicates that economic resilience remains the weakest dimension of village development in North Maluku Province. Furthermore, the relatively small standard deviation values suggest that the data distribution among villages was fairly homogeneous.

3.2 Data Balancing Using ADASYN

Before oversampling, the training data showed a severe class imbalance, where the "independent villages" category had far fewer observations than the "developing" and "underdeveloped" categories. This imbalance could potentially bias the classification model toward majority classes. The class distribution before balancing is presented in Figure 2.

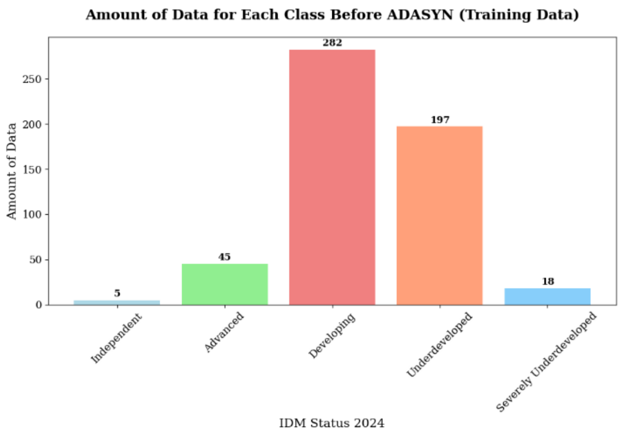

Figure 2. Distribution of Training Data Classes Before ADASYN Balancing

Figure 2 shows that the training data were highly imbalanced, with the "developing" category dominating the dataset with 282 observations, followed by the "underdeveloped" category with 197 observations. In contrast, the minority classes contained far fewer data points, namely "advanced" villages (45), "very underdeveloped" villages (18), and "independent villages" with only 5 observations. This condition indicates an extreme imbalance problem that required balancing before model training.

After applying the ADASYN oversampling method to the training data, the class distribution became more balanced through the generation of synthetic minority samples. The balanced class distribution after ADASYN is presented in Figure 3.

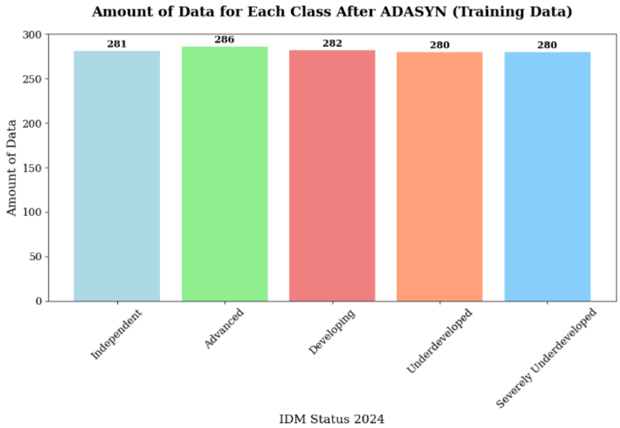

Figure 3. Distribution of Training Data Classes After ADASYN Balancing

Figure 3 shows that the minority classes increased significantly after applying ADASYN through the generation of 862 synthetic observations. As a result, the dataset became more balanced and proportional, reducing the risk of model bias toward majority classes. The adaptive mechanism of ADASYN generated synthetic samples mainly around difficult minority observations near the decision boundaries, thereby improving the classifier's ability to recognize minority classes more effectively.

3.3 Hyperparameter Optimization Results

Hyperparameter tuning was performed using Cross Validation (CV) for each kernel type. The optimal parameters obtained from the tuning process are presented in Table 4.

Table 4. Optimal Parameters and Cross Validation Results

Kernel	Optimal Parameters	Average CV Score
Linear	$C = 100$	0.8686
Polynomial	$C = 10$, $degree = 3$	0.7219
RBF	$C = 100$, $\gamma = 0.1$	0.6634

Based on Table 4, the results showed that the Support Vector Machine (SVM) model with a linear kernel produced the highest average Cross Validation score compared to the polynomial and RBF kernels, suggesting superior generalization capability and a stronger ability to capture the underlying structure of the IDM dataset. Furthermore, the linear kernel generated 59 support vectors, fewer than the polynomial kernel 165 support vectors and RBF kernel 127 support vectors, indicating that the linear kernel formed a simpler and more efficient classification boundary.

3.4 Classification Performance Evaluation

The evaluation results showed that the ADASYN-based Multiclass SVM model achieved excellent classification performance. The comparison of evaluation metrics for each kernel is presented in Table 5.

Table 5. Performance Comparison of SVM Kernels

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	98.54%	98.26%	99.40%	98.83%
Polynomial	97.08%	97.31%	97.08%	97.16%

Kernel	Accuracy	Precision	Recall	F1-Score
RBF	96.35%	96.53%	96.35%	96.39%

Based on Table 5, the linear kernel achieved the best overall performance, with an accuracy of 98.54%, precision of 98.26%, recall of 99.40%, and an F1-score of 98.83%. These results indicate that the integration of ADASYN and SVM effectively addressed the extreme imbalanced class problem in IDM classification. The high recall value indicates that the model is highly effective in accurately identifying village categories, including minority classes, while the high F1-score indicates a good balance between precision and recall. These findings are consistent with previous research reporting the effectiveness of ADASYN–SVM in handling imbalanced multi-class data. Ramadhan (2021) reported that ADASYN–SVM achieved an accuracy of 87.3% in detecting type 2 diabetes, while Ramadhani et al. (2025) achieved an accuracy of 96.38% in classifying traffic accidents. However, this study achieved a higher accuracy rate of 98.54%, indicating that the combination of ADASYN and SVM is highly suitable for classifying IDM status based on the IKS, IKE, and IKL indicators.

3.5 Confusion Matrix Analysis

Based on the evaluation results, the Support Vector Machine (SVM) model with a linear kernel achieved the best classification performance compared to the polynomial and RBF kernels. Therefore, the confusion matrix analysis was focused on the linear kernel model to evaluate its classification performance in more detail. The confusion matrix of the best model is presented in Figure 4.

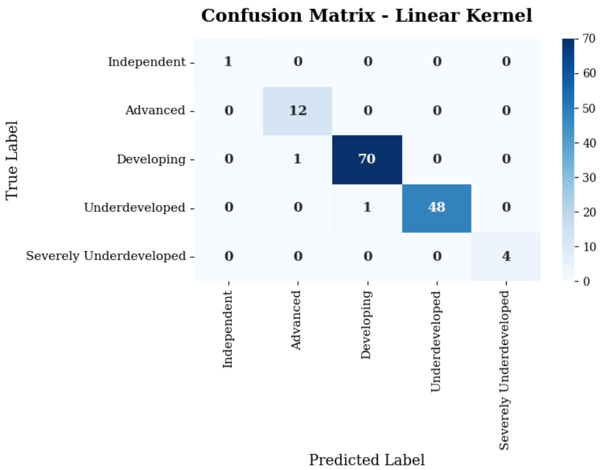

Figure 4. Confusion Matrix of the Linear Kernel SVM Model

Figure 4 shows that the linear kernel model classified almost all testing data correctly. Out of 137 testing observations, only two villages were misclassified, namely Salimuli Village and Dowongimaiti Village. Most village categories were successfully predicted into their correct classes, indicating that the predictor variables, namely the Social Resilience Index (IKS), Economic Resilience Index (IKE), and Environmental Resilience Index (IKL), provided strong discriminatory information for IDM classification.

The superior performance of the linear kernel suggests that the predictor variables (IKS, IKE, and IKL) provide a feature space that is largely linearly separable after the ADASYN balancing process. This finding is consistent with previous studies reporting that linear SVM models often outperform more complex kernels when the underlying data structure is relatively simple and the dimensionality of the feature space is limited. The high recall value obtained by the proposed model is particularly important from a policy perspective because misclassification of disadvantaged villages may lead to inappropriate allocation of development resources. Therefore, the ability of the model to accurately identify minority categories contributes directly to more effective policy targeting.

CONCLUSION

This study successfully applied the ADASYN-based Multiclass Support Vector Machine (SVM) method for classifying the Village Development Index (IDM) in North Maluku Province. The results showed that the data suffered from extreme class imbalance, which was successfully addressed through the application of ADASYN by generating 862 synthetic data points, thereby making the training data distribution more balanced. The best model was obtained using a linear kernel with a parameter of C = 100, yielding an accuracy of 98.54%, precision of 98.26%, recall of 99.40%, and an F1-score of 98.83%. During the testing phase, there were only two classification errors out of 137 observations, demonstrating excellent generalization ability and predictive performance. Overall, the findings demonstrate that integrating ADASYN with Multiclass SVM provides an effective framework for handling severe class imbalance while maintaining high predictive accuracy in village development classification tasks. The proposed approach can support data-driven policy formulation and development prioritization, particularly in geographically complex island regions such as North Maluku Province. For future studies, additional predictor variables and alternative machine learning algorithms may be explored to further improve classification performance and model interpretability.

ACKNOWLEDGMENTS

The authors would like to express gratitude to the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration of Indonesia for providing access to the dataset used in this study.

REFERENCES

[1] N. Sari and T. Oktavianor, "Indeks Desa Membangun (IDM) di Kabupaten Barito Kuala," *J. Adm. Publik dan Pembang.*, vol. 2, no. 1, p. 36, 2021, doi: 10.20527/jpp.v2i1.2768.

[2] E. Setyowati, "Tata Kelola Pemerintahan Desa Pada Perbedaan Indeks Desa Membangun (IDM): Studi Tiga Desa di Kabupaten Malang," *JISPO J. Ilmu Sos. dan Ilmu Polit.*, vol. 9, no. 2, pp. 170–188, 2019, [Online]. Available: https://doi.org/10.15575/jispo.v9i2.5324

[3] H. Salsabillah and B. Fitanto, "Determinan Yang Mempengaruhi Disparitas Antar Wilayah Di Kawasan Timur Indonesia Tahun 2014-2019," *J. Dev. Econ. Soc. Stud.*, vol. 3, no. 2, pp. 468–482, 2024, doi: 10.21776/jdess.2024.03.2.11.

[4] A. N. Astika and N. S. Subawa, "Evaluasi Pembangunan Desa Berdasarkan Indeks Desa Membangun," *J. Ilm. Muqoddimah J. Ilmu Sos. Polit. dan Hum.*, vol. 5, no. 2, pp. 223–232, 2021, doi: https://doi.org/10.31604/jim.v5i2.2021.223-232.

[5] Kementerian Desa PDTT RI, *Peringkat Indeks Desa Membangun (IDM Tahun 2024)*. 2024. [Online]. Available: https://drive.google.com/drive/folders/1LD_LR4O7ojKyf8nNXYA3bUXzsOSsq1MB

[6] N. C. Woods, V. E. Agada, and A. K. Ojo, "Predicting Phishing Websites Using Support Vector Machine and Multi-Class Classification Based on Association Rule Techniques," *Univ. Ibadan J. Sci. Logics ICT Res.*, vol. 2, no. 1, pp. 28–39, 2025, doi: 10.48550/arXiv.2505.21141.

[7] K. G. S, K. C, and R. Sethuraman, "Using Multi Class SVM for Creating an Algorithm for Human Activity Recognition," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India: IEEE, 2024, pp. 1–5. doi: 10.1109/ACCAI61061.2024.10602276.

[8] F. Latuconsina, M. S. N. Van Delsen, and Yudistira, "Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass pada Data Indeks Desa Membangun (IDM) di Provinsi Maluku," *J. Math. Comput. Stat.*, vol. 7, no. 2, pp. 380–395, 2024, doi: 10.35580/jmathcos.v7i2.3624.

[9] R. Obiedat *et al.*, "Sentiment Analysis of Customers ' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[10] A. Alhudhaif, "A novel multi-class imbalanced EEG signals classi fi cation based on the adaptive synthetic sampling (ADASYN) approach," *PeerJ Comput. Sci.*, vol. 7, no. e523, 2021, doi: 10.7717/peerj-cs.523.

[11] C. Magnolia, A. Nurhopipah, and B. A. Kusuma, "Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka pada Aplikasi Twitter," *Edu Komputika J.*, vol. 9, no. 2, pp. 105–113, 2022, doi: 10.15294/edukomputika.v9i2.61854.

[12] N. G. Ramadhan, "Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus," *Sci. J. Informatics*, vol. 8, no. 2, pp. 276–282, 2021, doi: 10.15294/sji.v8i2.32484.

[13] M. F. Ramadhani, T. A. Y. S. Siswa, and W. J. Pranoto, "Penerapan Metode Forward Selection dan ADASYN pada Algoritma SVM untuk Klasifikasi Data Kecelakaan Lalu Lintas," *J. Buffer Inform.*, vol. 11, no. 2, pp. 25–34, 2025, [Online]. Available: https://journal.fkom.uniku.ac.id/buffer/article/view/472/132

[14] D. Ardiyansyah and N. Oktafiani, "Perbandingan Metode Pengukuran Jarak Pada K-Nearest Neighbour Dalam Klasifikasi Data Teks Cardiovaskular," *J. Inf. Syst. Manag. Digit. Bus.*, vol. 1, no. 2, pp. 116–122, 2024, doi: 10.59407/jismdb.v1i2.260.

[15] I. Reviyana, "Perbandingan Algoritma K-Nearest Neighbor (KNN) Menggunakan Metode SMOTE dan ADASYN Dalam Mengatasi Data Tidak Seimbang," 2023.

[16] E. Ileberi and Y. Sun, "Advancing Model Performance With ADASYN and Recurrent Feature Elimination and Cross-Validation in Machine Learning-Assisted Credit Card Fraud Detection : A Comparative Analysis," *IEEE Access*, vol. 12, pp. 133315–133327, 2024, doi: 10.1109/ACCESS.2024.3457922.

[17] Z. Wan, "Prediction and Visualization Analysis of Lung Cancer Risk by Machine Learning," in *Highlights in Science, Engineering and Technology*, 2023, pp. 221–229. doi: 10.54097/hset.v39i.6531.

[18] B. Chen, Y. Hu, G. Zhang, D. Cheng, and S. Zhang, "Certainty-Aware Support Vector Machines for Fair Classification," *Neurocomputing*, vol. 655, p. 131330, 2025, doi: https://doi.org/10.1016/j.neucom.2025.131330.

[19] T. Thayatharsan and A. Ramanan, "Improving Multiclass SVM Classification Using a Hybrid Alternating Decision Strategy," Kelaniya, Gampaha, Sri Lanka: IEEE International Research Conference on Smart Computing and Systems Engineering (SCSE), 2026, pp. 1–6. doi: 10.1109/SCSE70081.2026.11499835.

[20] H.-Y. Huang and C.-J. Lin, "Linear and Kernel Classification: When to Use Which?," Society for Industrial and Applied Mathematics, 2016, pp. 216–224. doi: 10.1137/1.9781611974348.25.

[21] Z. Song, D. P. Woodruff, Z. Yu, and L. Zhang, "Fast Sketching of Polynomial Kernels of Polynomial Degree," in *Proceedings of Machine Learning Research*, 2021, pp. 9812–9823.

[22] Y. Julianto, D. H. Setiabudi, and S. Rostianingsih, "Analisis Sentimen Ulasan Restoran Menggunakan Metode SVM," *J. Infra*, vol. 10, no. 1, 2022.

[23] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies*, vol. 9, no. 4, p. 81, 2021, doi: https://doi.org/10.3390/technologies9040081.

[24] D. W. Saputra, "Analisis Pembangunan Desa Berdasarkan Capaian Indeks Desa Membangun Di Kecamatan Tempuling Kabupaten Indragiri Hilir," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2023. [Online]. Available: http://repository.uin-suska.ac.id/id/eprint/75623

[25] Kementerian Desa PDTT RI, *Standar Operasional Prosedur (SOP) Update Data Indeks Desa Membangun Tahun 2020.* Direktorat Jenderal Pembangunan dan Pemberdayaan Masyarakat Desa: Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi, Republik Indonesia., 2020.

[26] A. Yakin, E. Rustiadi, and D. O. Pribadi, "Analisis Spasial Desa Membangun Berbasis Jaringan Jalan dan Penggunaan Lahan di Kabupaten Brebes," *J. Perenc. Wilayah, Kota, dan Permukim.*, vol. 6, no. 1, pp. 42–59, 2024, doi: 10.20961/desa-kota.v6i1.77483.42-59.