Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Classification of Stunting using a Website-Based Support Vector Machine (SVM) Algorithm (Case Study: Pagar Merbau Community Health Center)

Muhammad Davin Diza Ghifary[1]*, Chairunisah[2], Said Iskandar Al Idrus[3], Faridawaty Marpaung[4], Kana Saputra[5]

[1,2,3,4,5]Computer Science, Universitas Negeri Medan
Dizadavid03@gmail.com[1], denisaziyad0105@gmail.com[2]*

Abstract

Stunting is a chronic nutritional problem that affects the physical growth and cognitive development of toddlers, and remains a serious challenge in Indonesia, including in the working area of the Pagar Merbau Community Health Center. Determining stunting status in the field often faces the risk of error when done manually, so a fast and accurate classification system is needed. This study aims to develop a classification model for the nutritional status of infants (normal, stunted, and severely stunted) using the Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel, and to integrate it into a web-based system to facilitate healthcare workers in conducting automatic data analysis. The data used were collected from infants aged 0–60 months at the Pagar Merbau Community Health Center (2023–2025), with four main attributes: gender, age, weight, and height. Model testing yielded an accuracy of 90%, with precision values of 0.95 (normal), 0.78 (stunting), and 1.00 (severe stunting), and recall values of 0.73, 0.96, and 1.00 for the same class. The implementation of the model on the website allows for the input of infant data, automatic analysis, and real-time visualization of classification results. This system is expected to improve the accuracy, efficiency, and precision of nutritional interventions at the Pagar Merbau Community Health Center.

Keywords: *Stunting, Support Vector Machine, Classfication , Website, Data Mining*

1. Introduction

Stunting is one of the most prevalent chronic nutritional issues worldwide, particularly in developing countries [1]. It is defined as impaired linear growth due to long-term nutritional deficiency, frequent infections, or inadequate care practices [2]. The condition negatively impacts not only children's height but also cognitive and motor development, educational achievement, and future productivity [3]. Indonesia ranks second in Southeast Asia and fifth globally in stunting prevalence [4]. In 2022, the Indonesian Nutritional Status Survey (SSGI) reported a 1.4% increase in stunting cases in Deli Serdang Regency, reaching 13.9% or approximately 541 children. In the Pagar Merbau sub-district, stunting prevalence rose from 0.64% in 2021 to 2.03% in 2022, indicating a growing local health concern that demands immediate intervention. Conventional assessment methods for stunting, which rely on manual anthropometric comparisons to WHO growth standards, are prone to human error [5]. Machine learning provides an opportunity to automate and enhance accuracy in nutritional assessment through pattern recognition and predictive analytics. Among classification algorithms, Support Vector Machine (SVM) is widely recognized for its ability to handle complex, high-dimensional data and to generalize effectively even with limited samples [6][7]. Therefore, this research focuses on designing and developing a web-based stunting classification system using SVM with an RBF kernel, trained on anthropometric data from children at Pagar Merbau Health Center. The objectives are threefold: (1) to construct an accurate SVM-based stunting classification model, (2) to evaluate its performance in classifying normal, stunted, and severely stunted categories, and (3) to implement the model in a user-friendly web system for healthcare personnel.

2. Literature Review

Stunting is defined as a condition of stunted linear growth caused by chronic malnutrition, recurrent infections, or other factors, especially during the first 1,000 days of life [1]. This long-term nutritional deficiency begins during pregnancy and has serious impacts that go beyond physical growth. Specifically, stunting hinders brain development, reduces cognitive abilities, and limits children's future potential [2].

Although physical manifestations in the form of short stature only become apparent after a child reaches the age of two, stunting is essentially a disorder caused by a sustained lack of daily nutritional intake [3]. In addition to affecting developmental aspects, stunting also increases the risk of chronic diseases later in life. Therefore, stunting is not merely a height issue but a public health concern that impacts the quality of human resources holistically.

A website is defined as a collection of digital pages that can be widely accessed through the internet, presenting information in various formats (text, images, animations, and videos) that can be accessed via a URL on a browser [13]. Fundamentally, websites serve as an efficient and widespread medium for disseminating information. In the context of information systems, websites play a crucial role as a means of interactive and real-time communication. Furthermore, websites are the primary platform for data management, online transactions, and the provision of digital services, which collectively support digitization efforts in various sectors.

Classification is a fundamental technique in data mining that serves to group data into predetermined categories or classes. This process involves analyzing data patterns and characteristics to automatically assign each entity to the most appropriate class. Classification has significant advantages in processing large amounts of data, improving the accuracy and efficiency of analysis, and revealing hidden patterns that are difficult to detect manually. The application of classification algorithms—such as Decision Tree, Naive Bayes, Support Vector Machine (SVM), and Neural Network enables the interpretation of complex data, thereby greatly supporting evidence-based decision making. This method is widely used in various sectors, including market trend prediction, fraud detection, medical diagnosis, and sentiment analysis. As machine learning advances, classification accuracy and speed continue to improve, making it an essential component in modern data exploration and analysis. Machine Learning (ML) is a branch of Artificial Intelligence (AI) that focuses on developing intelligent systems capable of learning and adapting independently without the need for explicit reprogramming for each task (Putu et al., 2024). ML utilizes training data to recognize patterns, make predictions, and take decisions, thereby producing dynamic and efficient solutions to complex problems in various sectors, including finance, healthcare, and manufacturing.

Python is an interpretive programming language known for its ease of learning and high code readability, characterized by clear and concise syntax [14]. Its simple yet powerful design enables effective and efficient code writing, thereby increasing software development productivity. Python supports a variety of programming paradigms, including object-oriented, imperative, and functional, providing developers with a high degree of flexibility. Python's main strength lies in its extensive ecosystem and active community support. Various libraries and frameworks are available, making it the top choice in various fields [15].

Support Vector Machine (SVM) is a Machine Learning algorithm based on the principle of Structural Risk Minimization (SRM), which aims to minimize generalization errors in previously unseen data [16] The main objective of SVM is to find the optimal hyperplane that can separate classes in the input space with maximum separation margin, resulting in a highly reliable and accurate classification model. To handle data that cannot be separated linearly (non-linear), SVM utilizes the Kernel Trick [8] This technique implicitly maps data from a low-dimensional space to a higher-dimensional feature space so that the data becomes easier to separate linearly. Commonly used kernel types include linear, polynomial, Radial Basis Function (RBF), and sigmoid

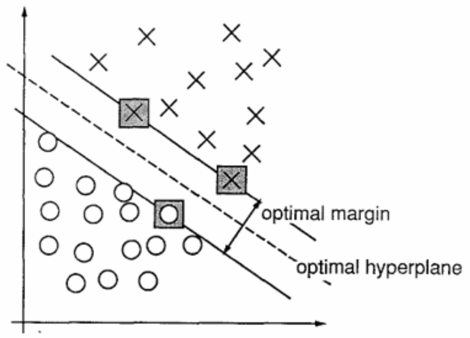

Figure 1: Model Support Vector Machine (Cortes dan Vapnik, 1995)

Figure.1 shows how the Support Vector Machine (SVM) algorithm works, which is influenced by the selection of variables or features used in the model. In its application, SVM focuses on finding a linear function that can separate data into two different classes. This linear function serves as a hyperplane in feature space that aims to maximize the distance (margin) between the two classes. The selection of appropriate variables is very important because it will affect how well this linear function can separate data and produce an accurate model in the form of an equation, which is an important factor in evaluating the effectiveness of the model.

A confusion matrix is a table used to evaluate the performance of a classification model by describing the amount of test data that is classified correctly and the amount of test data that is classified incorrectly. This table provides a detailed overview of how the model's predictions compare to the actual labels of the test data, enabling in-depth analysis of classification errors. Confusion matrices are very useful in various machine learning applications, especially in binary and multi-class classification. For example, in disease detection, false negatives (FN) can be fatal because patients who are actually sick are classified as healthy. Similarly, in fraud detection systems, false positives (FP) can cause legitimate transactions to be rejected. Understanding the confusion matrix and the metrics derived from it is essential for improving the accuracy and reliability of classification models, as well as adjusting models to the specific needs of an application. (Dwi Normawati, 2021).

3. Research Methods

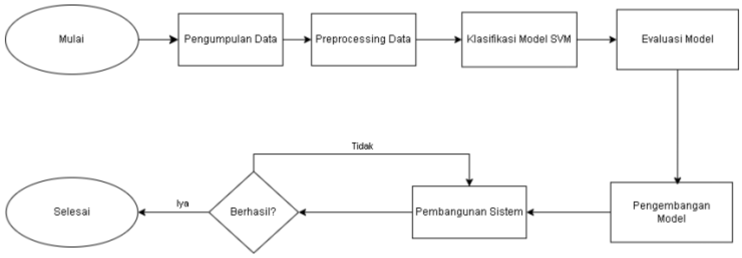

Figure 2: Research Stages

3.1. Data Collecting

This study uses data obtained from the Pagar Merbau Community Health Center as its main source of data. The data covers four main attributes that serve as independent variables, namely age, weight, and height. These attributes were selected to provide a comprehensive picture of the factors that can affect the health conditions of toddlers in the region.

3.2. Preprocessing Data

Preprocessing is the initial stage that aims to prepare raw data so that it can be analyzed or used in machine learning models. This step is very important because the quality and characteristics of the data used have a significant impact on the performance of the resulting model. The preprocessing process involves processing raw data into a ready-to-use dataset, including filtering and selecting relevant data for further processing in the document. This stage also involves removing unnecessary columns (drop column) to make the dataset more focused and efficient, checking for missing values to ensure data completeness, and removing duplicate data to avoid redundancy that could affect the analysis results. Thus, preprocessing ensures that the data used is appropriate for the analysis requirements and supports more accurate and efficient model results.

3.3. Classification Support Vector Machine

The process of classifying the nutritional status of toddlers was carried out using a Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel. The preprocessed dataset was divided into training data and test data. The RBF kernel was selected based on its ability to handle non-linear data separation. For model optimization, parameter tuning was performed by testing combinations of C and γ (gamma) values. The classification stages included model training, testing, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The test results showed that the optimal parameter combination was $C=100$ and $\gamma=0.25$, which produced the best classification performance on the research dataset.

3.4. Figures and tables

After the model has been successfully trained, performance evaluation is a crucial step, in which the confusion matrix is the most common evaluation tool. A confusion matrix is a table that presents a comparison between the model's classification results and the actual labels of the test data. This tool provides a detailed overview of the classification model's performance by visualizing the number of correct predictions and incorrect predictions. Thus, the confusion matrix is used to evaluate the model's accuracy in more detail (Normawati, 2021).

3.5. Model Development

The main objective of developing this model is to create an effective classification system to predict the nutritional status of toddlers (normal, stunted, severely stunted) with high accuracy. During the training process, model performance analysis was conducted using accuracy, precision, recall, and F1-score metrics to ensure that the model was not only accurate but also showed balance in classifying each nutritional category. The final results of this model development were then integrated into a website-based system for direct implementation in the classification process at the Pagar Merbau Community Health Center.

3.6. System Development

A website-based system was developed as a follow-up to the classification results evaluation. This system functions as a visualization dashboard that aims to facilitate medical personnel at the Pagar Merbau Community Health Center in analyzing and identifying the risk of stunting in toddlers. In its application, this system allows for the input of infant data (gender, age, weight, and height), which is then processed by a Support Vector Machine (SVM) model that has been trained to directly determine nutritional status and potential stunting.

4. Result and Discussion

4.1. Description Data

Data collection in this study was conducted using the documentation method through the utilization of medical records of toddlers at the Pagar Merbau Community Health Center. Based on the results of the survey, 1,548 toddler data were obtained. The data covered five main variables, namely gender, age, height, weight, and toddler status.

Table 1: Collecting Data

Category	Amount of Data
Normal	493
Stunting Parah	260
Stunting	795
Total	1548

4.2. Preprocessing Data

After the data has been collected, the next step is to perform data preprocessing. This stage aims to improve data quality and ensure that the data is in a format that is suitable for the classification algorithm used. In this study, the preprocessing process consists of six main stages, namely: data transformation using the label encoding method, column deletion (drop column), handling missing values, removing duplicate data, balancing data using the undersampling technique, and normalizing data through standardization (Z-Score Normalization) or scale transformation.

4.2.1 Transformation Data

Data preprocessing is an essential step before model training. One important step in this stage is data transformation to ensure that the data format is suitable for modeling purposes. In this study, Label Encoding is used to convert categorical variables in the dataset into numerical representations. This conversion is based on the index or unique sequence of each variable category, which will be explained further.

Table 2: Label encoding on Toddler Status Variables

Infant Status	Label
Normal	0
Stunting Parah	1
Stunting	2

Table 3: Gender label encoding

Infant Status	Label
Laki - Laki	0
Perempuan	1

4.2.2 Drop Column

After data transformation through Label Encoding, the next step is data cleaning by removing irrelevant or duplicate columns (drop column). The Name column was removed because it is a unique identifier that does not contribute to modeling. The Unnamed: 5 column was eliminated because it did not have a clear name and was suspected to be an artifact of data format errors. Furthermore, the Sex and Month columns were deleted because they were duplicates and only differed in naming (English) from the existing Gender and Age columns. This deletion aimed to eliminate redundancy and inconsistency, resulting in a clean and relevant dataset for stunting status analysis and modeling.

4.2.3 Missing Value

After removing irrelevant columns, the next preprocessing step is to check for missing values. Missing values occur when some data entries are empty, usually due to recording errors or unavailability of information. Based on the checks performed, no empty data was found in this research dataset, ensuring data integrity for the next stages of analysis and modeling.

4.2.4 Check Duplicate

After handling missing values, the next preprocessing step is to check for duplicate data. Duplication, which may arise due to input errors or data merging, can cause bias in modeling. Based on the examination, 12 data points were indicated as duplicates in this research dataset, which required further handling before analysis.

4.2.5 Data Balancing

After removing duplicate data, the next step is to perform data balancing because class imbalance was identified in the Stunting Status column. The amount of data for the "Severe Stunting" category far exceeds the 'Stunting' and "Normal" categories. This imbalance can cause model bias towards the majority class. Therefore, the undersampling technique was applied by randomly removing some data from the majority categories ("Severe Stunting" and 'Normal') until the number was equal to that of the minority class ("Stunting"). This process produced a balanced final dataset, in which each class had 260 entries, with a total of 780 data ready for model training.

Table 4 Class Distribution After Data Balancing

Category	Amount of Data
Normal	260
Stunting Parah	260
Stunting	260
Total	780

4.2.6 Normalization Test

After data balancing using undersampling, the next preprocessing step is data normalization. In this study, the Standardization or Z-Score Normalization method was applied to numerical features (age, weight, and height). The application of standardization is crucial because differences in scale or range of values between features have the potential to interfere with the analysis process and machine learning model training. Normalization ensures that each feature has an equal contribution to the model.

4.3. Aplication of SVM Classification

After going through all stages of preprocessing including data cleaning, removal of irrelevant attributes, and normalization the dataset was divided into training data and test data with a ratio of 90:10. The training data was used for model construction, while the test data served to evaluate the model's performance on unseen data. In the modeling stage, the Support Vector Machine (SVM) algorithm was selected, using the Radial Basis Function (RBF) kernel. The RBF kernel was chosen based on its proven reliability in handling data that cannot be separated linearly. The model was trained using four independent variables with the γ (gamma) parameter set at a value of 0.25.

4.4. Model Evaluation

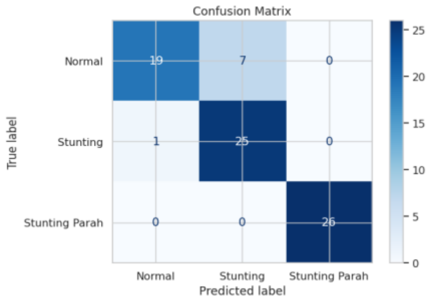

Figure 3: Confusion Matrix

After successfully implementing the SVM classification process, the next step is to evaluate the model. Model evaluation is carried out using a confusion matrix. A confusion matrix is used to evaluate the accuracy of a model by comparing the classification prediction results with the actual classes of the data. The following shows the confusion matrix generated from the application of the classification model using Python on the Google Colab platform.

4.5. Model Test

```
Classification Report for RBF Kernel:
              precision    recall  f1-score   support

           0       0.95      0.73      0.83        26
           1       0.78      0.96      0.86        26
           2       1.00      1.00      1.00        26

    accuracy                           0.90        78
   macro avg       0.91      0.90      0.90        78
weighted avg       0.91      0.90      0.90        78
```

Figure 4: Classification Report

Based on the classification report, the Support Vector Machine (SVM) model with RBF kernel proved to be very effective in classifying the nutritional status of toddlers with 90% accuracy, supported by high macro precision and recall (0.91 and 0.90, respectively), and achieved perfect precision and recall (1.00) in Class 2 (Severe Stunting).

5. Conclusion

Based on the results obtained from this study, several conclusions can be drawn as follows:

1. The data preprocessing process includes label transformation, duplicate data cleaning, class balancing using undersampling, and Z-Score normalization. The SVM model is optimized with the best parameters (C=100 and γ=0.25) to separate the data into three categories: normal, stunting, and severe stunting.
2. Evaluation of the model on the test data (10% of the total dataset) showed excellent performance with an accuracy value of 90%. In detail, the model achieved an average macro precision value of 0.91 and a recall of 0.90. The model showed significant advantages in detecting the severe stunting category with perfect precision and recall values (100%). This performance indicates that the developed SVM model has clear and consistent classification capabilities.
3. The SVM algorithm is integrated into a website developed using Python with the Flask framework and Scikit-learn (sklearn) library. This system is implemented at the Pagar Merbau Community Health Center and provides a number of key features, including user authentication (login and registration), information dashboard, child data management (input and details),

measurement data addition, and stunting classification page. These features are designed to support effective toddler data management and practical application of stunting classification.

Acknowledgement

The author would like to express sincere gratitude to Chairunisah, S.Si., M.Si. for her crucial guidance and support, as well as to Medan State University for providing facilities and an academic environment that supported the implementation of this research.

References

[1] Apriastini, N., et al., “Determinants of Stunting in Early Childhood,” *Public Health Journal*, 2024.
[2] Haskay, R., “Chronic Malnutrition and Cognitive Delay in Indonesian Children,” *Journal of Nutrition Studies*, 2024.
[3] Kurniati, D., Djuwita, A., & Istiqfani, N., “Nutritional Deficiency and Growth,” *Child Health Research*, 2023.
[4] Gurning, R., Mulyani, A., & Setiawan, T., “Stunting Statistics in Indonesia,” *Indonesian Health Review*, 2022.
[5] Prasetiya, T., et al., “Manual Anthropometric Assessment Errors,” *BMC Pediatrics*, 2020.
[6] Wahyuni, S.D., & Kusumodestoni, R.H., “Implementation of Support Vector Machine in Medical Classification,” *IJCA*, 2024.
[7] Cortes, C., & Vapnik, V., “Support Vector Networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995
[8] Prastyo, D., et al., “Kernel Functions in SVM,” *Data Science Review*, 2020.
[9] Fajri, R., & Primajaya, D., “Parameter Optimization in SVM,” *Computational Intelligence Journal*, 2023.
[10] Dwi Normawati, “Evaluation Metrics in Classification,” *Information Systems Research*, 2021.
[11] Widhari, W., Triyadi, A., & Sari, R.T.K., “Naïve Bayes and KNN in Diagnosing Stunting,” *Int. J. Health Informatics*, 2022.
[12] Prasetya, T., Ali, I., & Rohmat, C.L., “Classification of Stunting in Slangit Village Using KNN,” *Applied Computing Journal*, 2020.
[13] Kurniawan, B., Gunawan, R., & Elanda, A. (2022, May). Redesain Sistem Aplikasi Web Sumber Reload. In Prosiding Seminar Nasional Inovasi dan Adopsi Teknologi (INOTEK) (Vol. 2, No. 1, pp. 240-249).
[14] Gustiandi, B. (2023). Langkah Awal Menguasai Bahasa Pemrograman Python.
[15] Santoso, H.R., Yulina, S., & Muslim, I. (2024). Pelatihan Pemrograman Dasar Python pada SMKN 7 Pekanbaru. JITER-PM (Jurnal Inovasi Terapan - Pengabdian Masyarakat) [14] Gustiandi, B. (2023). Langkah Awal Menguasai Bahasa Pemrograman Python.
[16] Tommy Rustandi, S.R., Suhaedi, D., & Pemanasari, Y. (2023). Pemetaan Hyperlane Pada Support Vector Machine. Bandung Conference Series: Mathematics.