Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Classification of Program Keluarga Harapan Assistance Recipients Using a Website-Based Support Vector Machine Algorithm (Case Study: Panyabungan Kota Subdistrict)

Khoirul Ahyar[1]*, Yulita Molliq Rangkuti [2], Eng. Mansur. AS[3], Zulfahmi Indra[4], Kana Saputra S[5]

[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Indonesia
parinduriahyar@gmail.com [1]*

Abstract

Program Keluarga Harapan (PKH) is a social assistance program aimed at reducing poverty by providing financial aid to eligible families. This research focuses on the development and implementation of the Support Vector Machine (SVM) algorithm to classify PKH recipients in Panyabungan Kota Subdistrict, Mandailing Natal Destrict. The classification process utilizes factors such as family income, number of family members, and the presence of elderly members. These three factors are chosen due to their availability from public records, ensuring the privacy of participants. The classification model developed in this study is implemented in a web-based system built with PHP and JavaScript, designed to facilitate the automatic classification of PKH recipients. This system helps streamline the registration to be more precise and effective, providing an efficient solution for local government officials to identify eligible families for the PKH program. The evaluation results show that this system can classify PKH recipients well with an accuracy of 93%, offering an automated approach that supports decision-making in the distribution of social assistance.

Keywords*:* PKH, Support Vector Machine (SVM), classification, family income, eligibility, Panyabungan Kota Subsdistrict, Mandailing Natal District, web-based system, PHP, JavaScript.

1. Introduction

Poverty is a condition in which an individual is unable to obtain adequate resources to meet their basic needs, such as education, health, and social welfare [1]. The Central Bureau of Statistics (BPS) defines poverty as a state in which a person cannot fulfill their basic food and non-food needs, measured by the individual's monthly expenditure. Poverty occurs due to several factors, including a high number of unemployed individuals within the productive age group and the low level of education among the working-age population in a region [2]. Poverty is often described as a continuous poverty cycle because individuals find it difficult to escape from its grip due to limited access to opportunities that would improve their quality of life [2]. Poverty is a common and problematic issue found in many developing countries, including Indonesia.

Based on BPS data from March 2023, there are 25.90 million poor people in Indonesia [3], or 9.36% of the total population, with most residing in rural areas. In North Sumatra Province, the number of poor residents reached 1.34 million people (8.23%), particularly in Medan City, Langkat Regency, Deli Serdang, Simalungun, and Mandailing Natal [3] This study focuses on Panyabungan Kota District, Mandailing Natal Regency, because the area serves as the economic and administrative center of the regency and has a relatively high number of PKH recipients. Additionally, the researcher resides in the area, which provides direct understanding of the local socio-economic conditions and ease of obtaining data from the Social Service Office and village authorities. For these reasons, Panyabungan Kota District is considered a representative and relevant location for this research.

To reduce poverty levels, the Indonesian government has implemented numerous social protection policies, one of which is the *Program Keluarga Harapan* (PKH). PKH is a Conditional Cash Transfer (CCT) program that provides cash assistance aimed at accelerating poverty reduction in Indonesia. PKH includes several forms of assistance for poor households, such as access to health services, access to educational services, food support, and nutritional improvement for poor [4]. PKH assistance is provided in cash and is divided into two types: fixed assistance and component assistance. Fixed assistance is given consistently to poor households, while component assistance is given when families meet certain criteria, with a maximum claim for four individuals in one household [5]. In addition to poor households, PKH is also provided to people with disabilities and the elderly.

Individuals who wish to receive PKH assistance must be registered in the Integrated Social Welfare Data (DTKS). Registration is carried out by bringing an Identity Card (KTP) and Family Card (KK) to the village office or head of the local village. Once the data is submitted, it is validated and verified before being forwarded to the Social Service Office for further processing by the mayor or regent. If the applicant meets the criteria, the Ministry of Social Affairs will distribute the PKH funds to the eligible individual. Although this procedure is considered effective, it still has several weaknesses, such as the potential for subjectivity in selecting candidates, long verification times due to multiple administrative levels, and the risk of human error. As a solution, the Ministry of Social Affairs launched the "Cek Bansos" application to speed up self-registration. However, the application has not functioned optimally, so offline registration remains the primary method.

Over time, technological advancements have led to the development of new fields that can assist human tasks, one of which is Machine Learning. Machine Learning is a subfield of Artificial Intelligence that focuses on computational methods that can improve themselves through learning from a dataset to produce accurate predictions [6]. Machine Learning consists of several subsets, including Supervised Learning, Unsupervised Learning, Reinforcement Learning, and Semi-Supervised Learning [7]. Supervised Learning is a branch of machine learning used to learn patterns in labeled data [8]. It is used to build classification and regression models. A classification model works by learning the patterns in training data that already contain labels so that it can classify new data accordingly [9]. Classification is divided into two types: one-class classification and multi-class classification. Popular machine learning algorithms for classification include K-Nearest Neighbor (KNN), Naïve Bayes, and Support Vector Machine (SVM) [6].

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks. SVM works by finding the maximum margin between data points using a hyperplane in an N-dimensional feature space [4]. A hyperplane is a decision boundary used to classify data points. Data points that fall on one side of the hyperplane represent one class. When the data consists of only two classes, the hyperplane forms a linear boundary. The hyperplane is determined by the data points that are closest to it, known as support vectors. These support vectors are used to define the shape and width (margin) of the hyperplane to determine the best separation between classes [4].

Support Vector Machine (SVM) has strong accuracy and low computational cost compared to other machine learning algorithms. This is supported by several studies on PKH classification. For example, [2] showed that SVM achieved an accuracy of 93.5% in classifying PKH recipients based on numerical data. [5] found that a web-based SVM system for classifying PKH recipients in remote areas helped local Social Service Offices manage data more easily, achieving an accuracy of 90.8%. [6]6 also reported that SVM with an RBF kernel produced 92.7% accuracy for PKH classification using numerical data. Meanwhile, [8] showed that Naïve Bayes achieved only 80% accuracy in classifying PKH recipients in Timbang Deli. Based on these findings, SVM consistently outperforms other algorithms such as Naïve Bayes, making it a suitable choice for PKH classification tasks.
With the background and previous research related to SVM, this study aims to develop a model for classifying poor households to accelerate the verification process for PKH recipients and reduce subjectivity, making the registration process more independent. The model will then be developed and implemented into a web-based platform to make it more accessible to the public. The internet is not a significant obstacle for Indonesian society, as BPS reports that approximately 89.45% of Indonesian households have internet access [3]. Therefore, internet accessibility should not hinder the adoption of such a system.

2. Literature Review

The Family Hope Program (Program Keluarga Harapan/PKH) is one of Indonesia's major conditional cash transfer initiatives aimed at reducing poverty through financial assistance tied to education and health requirements. The program targets low-income households and seeks to break the intergenerational cycle of poverty by ensuring children stay in school, pregnant mothers receive proper health care, and households meet basic welfare standards [3]. Despite its positive impact since its launch in 2007, issues such as inaccurate beneficiary data and uneven distribution remain challenges in its implementation.

Poverty itself is a multidimensional issue characterized not only by insufficient income but also by limited access to education, health services, employment opportunities, and basic infrastructure. In Indonesia, poverty measurement relies on the national poverty line defined by minimum expenditure requirements for food and non-food needs. Factors such as household income, family size, number of school-age children, elderly dependents, access to employment, and level of education significantly influence poverty levels within households [2]. These variables are also central to determining PKH eligibility, making them highly relevant to classification models.

To address data inaccuracies and reduce subjectivity in PKH beneficiary verification, machine learning has emerged as a promising solution. Machine learning enables automated pattern recognition from historical data, improving decision-making consistency and efficiency. Support Vector Machine (SVM), a supervised learning algorithm, is widely used for classification tasks due to its robustness in handling small-to-medium datasets, ability to maximize class separation margins, and capacity to model non-linear relationships through kernel functions [9]. SVM's strengths make it suitable for socioeconomic classification problems where features are strongly interrelated.

In the context of PKH classification, SVM can learn patterns from poverty-related variables—income, family size, number of school-age children, and presence of elderly dependents—to predict whether a household qualifies for assistance. Although SVM performs well in many applications, optimal parameter selection (kernel, C, gamma) remains essential to ensure high accuracy and prevent overfitting. Furthermore, validation techniques such as K-Fold Cross Validation help assess the stability of the model across different subsets of data [10.
The development of the PKH classification system in this study also involves the use of Laravel, a PHP-based web framework known for its efficient routing, database management via Eloquent ORM, security features, and extensive ecosystem. Laravel enables the integration of the SVM model into an accessible web application, allowing real-time classification through a user-friendly interface [11]. The application was developed and tested in a local environment using Laragon, a lightweight development platform that simplifies database management, server configuration, and API testing [12].

To evaluate the system's functionality, Black Box testing was employed, focusing on input–output behavior without examining internal code structure. This ensures that the classification system performs according to requirements and functions effectively under various user scenarios [13].

Overall, previous literature highlights the importance of accurate targeting in social assistance programs, the multidimensional nature of poverty indicators, and the strong suitability of SVM for classification tasks. By integrating SVM into a Laravel-based web application and validating it through Black Box testing, this study contributes a practical and efficient solution for improving the accuracy and objectivity of PKH beneficiary verification at the local government level [14].

3. Methodology

3.1. Research Stages

This study follows the Research and Development (R&D) approach with the objective of developing a Support Vector Machine (SVM)-based classification system for determining PKH beneficiary eligibility. The research stages begin with literature review and proceed through data collection, preprocessing, SVM model implementation, model evaluation, and finally integration of the model into a web-based application. The overall workflow of the research is shown in Fig. 1.

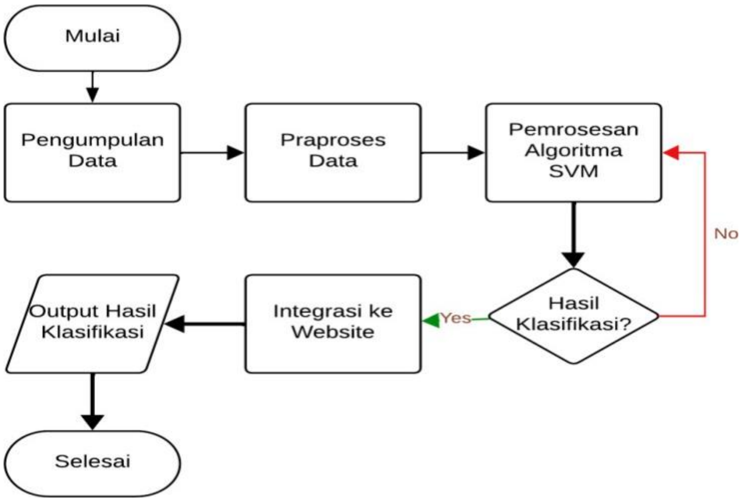

Fig. 1: Flowchart of the research stages

The process starts with designing the classification system and preparing the structure of the web platform. Afterward, data related to poverty indicators were collected, processed, and transformed into a suitable format for model training. The SVM algorithm was then trained, evaluated, and integrated into the final website system.

3.2. Data Collection

Data were collected from several villages in Panyabungan Kota District, Mandailing Natal Regency. The dataset was obtained through the PKH Coordinator with official authorization. The collected data include socioeconomic indicators relevant to PKH eligibility, such as household income, number of family members, number of school-age children, and presence of elderly dependents.

Data collection took place between April 2024 and July 2024. To protect privacy, personal identifiers such as names were anonymized. The collected dataset serves as the primary source for training and testing the SVM classification model.

3.3. Data Pre-processing

Data preprocessing was conducted to ensure that the dataset was clean and suitable for modeling. The steps included:

a. Data Cleaning: Removing duplicate entries, fixing inconsistent values, and eliminating incomplete records.
b. Normalization: Applying Min–Max Scaling to transform numerical features into the 0–1 range.
c. Class Balancing: Addressing imbalance between PKH and non-PKH classes by applying class weight in the SVM model.
d. Dataset Splitting: Dividing the dataset into training data (80%) and testing data (20%).

These preprocessing steps ensured that the SVM model received well-structured, balanced, and normalized input data.

3.4. Support Vector Machine Model

The SVM algorithm was applied to classify households into PKH-eligible and non-eligible categories. The model was implemented in Python and later integrated into a Laravel-based environment. The SVM classifier was trained using the preprocessed training data, and its performance was then evaluated using the testing dataset.

The features used in the SVM model include household income, number of family members, number of school-age children, and presence of elderly dependents. These features represent the poverty indicators considered in PKH eligibility assessment.

3.5. Model Evaluation

The performance of the classification model was assessed using common machine learning evaluation metrics, namely:
a. Accuracy – The proportion of correct predictions.
b. Precision – The correctness of positive predictions (PKH recipients).
c. Recall – The model's ability to correctly identify actual PKH recipients.
d. F1-Score – The harmonic mean of precision and recall.

These metrics provided insights into how well the SVM classifier performed on the testing dataset.

3.6. System Integration

After model training and evaluation, the SVM classifier was integrated into a web application developed using PHP Laravel and JavaScript. The website enables users to input household socioeconomic data and automatically obtain classification results from the trained SVM model.

4. Results & Discussion

4.1. Data Collection and Labeling

The data collection process was carried out from April to July 2024 in several villages located in Panyabungan Kota District. The dataset was obtained from the related institution in Excel format and was fully anonymized to ensure privacy and confidentiality. A total of 500 records were collected, consisting of:
a. 400 households eligible for PKH
b. 100 households not eligible for PKH

The dataset contained three main variables: family income, number of family members, and the presence of elderly members. These variables were selected because they are highly relevant to the criteria used by the Indonesian government in determining PKH eligibility.

Table 1: Presents A Portion Of The Data Used In This Study

No	Salary	Age	Number of Children	Results
1.	9000000	28	1	Not Eligible
2.	7000000	50	0	Not Eligible
3.	500000	87	3	Worthy
4.	1000000	65	5	Worthy
5.	3500000	53	0	Not Eligible
6.	9500000	30	0	Not Eligible
7.	2000000	30	4	Worthy
8.	2000000	56	5	Worthy
9.	500000	49	3	Worthy
10.	1500000	31	5	Worthy
⋮				
497.	1000000	51	2	Layak
498.	9000000	57	0	Not Eligible
499.	1000000	49	3	Layak
500.	1000000	56	4	Layak

The descriptive analysis shows that most eligible households have low income, a large number of dependents, and often include elderly members. These observations indicate that the selected variables are strong determinants of PKH eligibility.

4.2. Data Pre-processing

The data preprocessing stage was conducted to ensure that the dataset was ready for model training using the Support Vector Machine (SVM) algorithm. The preprocessing consisted of several steps:
1. Data Cleaning
a. Removing duplicate entries

b. Handling incomplete or inconsistent records
c. Replacing missing values with mean or mode

These steps ensured that the dataset was free from irregularities that could affect classification performance.
2. Data Transformation

a. Converting categorical attributes into numeric format
b. Encoding data when necessary
c. Standardizing the structure of data columns

3. Data Normalization (Min-Max Scaling)

Normalization was applied to transform numerical values into a range between 0 and 1 using the formula:

$$X_{\text{new}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Normalization was conducted on variables such as income, age, and number of children to prevent any feature from dominating the SVM decision boundary.

4. Class Balancing (Class Weight)

Since the dataset was imbalanced (400 eligible vs. 100 not eligible), class weights were applied during model training. This technique ensured that the model learned both classes fairly and avoided bias toward the majority class.
After preprocessing, the dataset was split into 80% training data and 20% testing data.

4.3. SVM Model Training Results

The SVM model was trained using the preprocessed training dataset. During training, SVM generated weights for each feature, representing their relative importance in determining eligibility.

Example of training weights:

a. Income: 1.5
b. Age: –2.0
c. Number of children: –1.0
d. Bias: 0.3

Prediction was performed using the SVM linear decision function:

a. If **score ≥ 0 → Eligible**
b. If **score < 0 → Not Eligible**

The model was then evaluated using the testing dataset.

4.4. Model Evaluation (Accuracy, Precision, Recall, F1-Score)

The performance of the SVM model was evaluated using accuracy, precision, recall, and F1-score. The confusion matrix results are as follows:

Table 2: F1-Score Calculation Summary

Remarks	Quantity
True Positive (TP)	74
False Positive (FP)	5
False Negative (FN)	6
True Negative (TN)	15

Precision

$$\frac{TP}{TP + FP} = \frac{74}{79} = 0.94$$

Recall

$$\frac{TP}{TP + FN} = \frac{74}{80} = 0.93$$

F1-Score

$$2 \times \frac{(0.94 \times 0.93)}{(0.94+0.93)} = 0.93$$

Interpretation

The resulting **F1-score of 93%** indicates that the SVM model:
a. performs with high accuracy,
b. shows a good balance between precision and recall,
c. effectively classifies both eligible and non-eligible households,
d. is not biased despite the imbalanced dataset.

Thus, the model can reliably support PKH eligibility classification.

4.5. Web-Based System Implementation

The trained SVM model was implemented into a web-based application. The system was developed using:
a. Laravel (PHP) as the backend framework
b. HTML, CSS, and JavaScript for the frontend

The application includes four main interfaces:
1. Home Page

Fig. 2: View Home Page

Displays general information about the system and its purpose.

2. SVM Classification Results Page

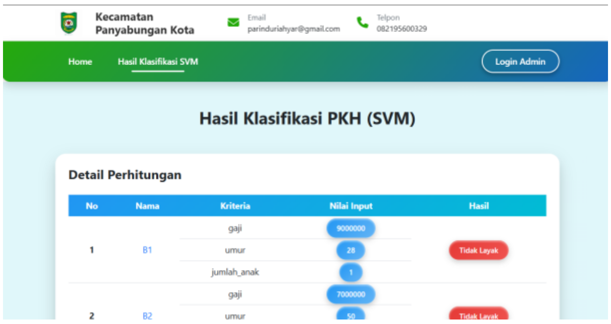

Fig. 3: View SVM Classification Results Page

Shows classification results in tabular or graphical form based on family income, number of members, number of children, and the presence of elderly members.

3. Admin Login Page

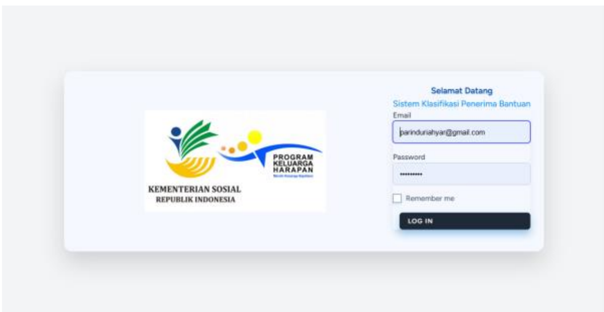

Fig. 4: View Admin Login Page

Allows authorized administrators to access the system securely.
4. Admin Dashboard

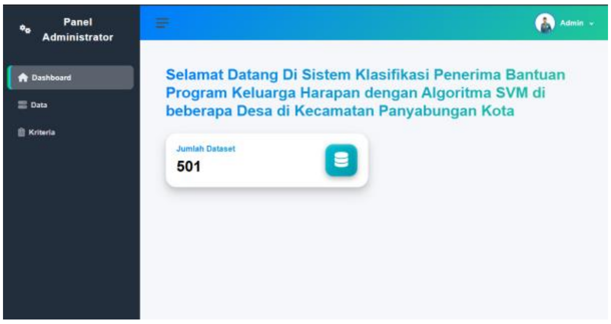

Fig. 5: View Admin Dashboard

Enables administrators to:
a. manage household data
b. edit classification criteria
c. adjust feature weights
d. view summary statistics

The database was created using Laravel's migration feature, ensuring consistency and maintainability.

4.6. System Testing Using Black Box Method

Black Box testing was performed to evaluate the functional aspects of the system without inspecting its internal code. The testing scenarios included:
a. login functionality
b. data input and storage
c. feature suitability
d. system flexibility
e. interface clarity
f. navigation ease
g. visual layout and readability
h. system response time

Each test item was evaluated using the **Likert Scale (1–4)**.
The final score was calculated using:

$$\text{Final Score} = \frac{\text{Total Points Obtained}}{\text{Maximum Possible Points}} \times 100\% \quad (2)$$

Results show that the system falls under the **"Satisfactory to Very Satisfactory"** category, indicating that it meets user needs effectively.

4.7. System Strengths and Weaknesses

The strengths of the developed system include:
1. High model performance with a 93% F1-score.
2. Classification process is objective and faster than manual verification.
3. Web-based system is accessible from any device.
4. User interface is simple and easy to understand.
5. Database is well-integrated through Laravel.

The identified weaknesses of the system include:
1. Dataset is limited to a single region (Panyabungan Kota).
2. The model uses only three primary features.
3. Internet connection is required to access the system.

5. Conclusion and Recommendations

5.1. Conclusion

This study successfully developed a classification system for determining the eligibility of households to receive the *Program Keluarga Harapan* (PKH) assistance using the Support Vector Machine (SVM) algorithm. Based on the research conducted, several conclusions can be drawn:

1. Development of the Classification System:
   The SVM-based classification model was able to categorize PKH eligibility using three main variables: family income, number of family members, and the presence of elderly household members. The model achieved a 93% accuracy rate, indicating that SVM performs effectively and can be relied upon in classifying household eligibility.
2. Effective Use of Preprocessed Data:
   The dataset used in this study—after undergoing preprocessing steps including data cleaning, handling missing values, and normalization—provided valid and reliable input for the classification process. These preprocessing steps ensured that the data were consistent and suitable for automatic PKH eligibility classification.
3. Web-Based System Integration:
   The trained SVM model was successfully integrated into a web-based system developed using the Laravel framework. This integration enables users to automatically classify PKH recipients through an interactive and responsive web interface, making the system accessible and user-friendly.

Overall, the findings demonstrate that the SVM algorithm, combined with proper preprocessing and a web-based interface, can support a more objective, efficient, and transparent PKH classification process.

5.2. Recommendations

Based on the research findings and limitations identified in this study, several recommendations can be proposed for future development:

1. Addition of New Variables:
   Future research may consider incorporating additional variables—such as the education level of the household head, housing conditions, and access to healthcare facilities—to enhance the accuracy and relevance of PKH eligibility classification.
2. Periodic Model Evaluation:
   The SVM model should be evaluated periodically to maintain optimal performance, especially if there are changes in the dataset or PKH eligibility criteria in the future.
3. System Feature Enhancements:
   The system can be further improved by adding features such as statistical analysis tools, visual reporting dashboards, and recommendations for policy interventions based on classification outcomes.

Through continuous development and evaluation, this PKH classification system has the potential to support more effective and transparent decision-making processes, contributing positively to the management of social assistance programs in Indonesia.

References

[1] Todaro, M. P., & Smith, S. C. (2020). Economic Development. Thirteenth Edition.In *Pearson* (Issue 13th Edition). https://www.mkm.ee/en/objectives-activities/economic-development

[2] Made Ariasih, N. L., & Yuliarmi, N. N. (2021). Pengaruh Tingkat Pendidikan, Tingkat Kesehatan dan Pengangguran Terbuka Terhadap Tingkat Kemiskinan di Provinsi Bali. *Cerdika: Jurnal Ilmiah Indonesia*, *1*(7), 821–839.

[3] Badan Pusat Statistik. (2023). Profil Kemiskinan di Indonesia Maret 2023. *Badan Pusat Statistik*, *47*, 1–16.

[4] Mirtaheri, S. L., & Shahbazian, R. (2022). Machine Learning Theory to Applications. In *Machine Learning Theory to Applications*.

[5] Nazarudin, P. (2021). Pedoman Pelaksanaan Program Keluarga Harapan 2021. In *Direktur Jaminan Sosial Keluarga Direktorat Jendral Perlindungan Dan Jaminan Sosial Kementrian Sosial RI* (Vol. 5, Issue 2).

[6] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2019). Foundations of Machine Learning. In *SSRN Electronic Journal*.

[7] Jo, T. (2021). Machine learning foundations: Supervised, unsupervised, and advanced learning. In *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*.

[8] Muller, A., & Guido, S. (2018). *Introduction to Machine Learning with Python: A Guide for Beginners in Data Science*.

[9] Doshi, D. R., Hiran, D. K. K., Jain, R. K., & Lakhwani, D. K. (2021). Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition). In *BPB Publications* (Vol. 2, Issue 5).

[10] Rismawandi, R., Pasek, I. G., Wijaya, S., & Nugraha, G. S. (2022). *Implementasi Metode Convolutional Neural Network Untuk Penegenalan Huruf Aksara Sasak Pada Android (Implementation Convolutional Neural Network Method for Recognition of Sasak Characters in Android)*. *4*(1), 11–20. http://jtika.if.unram.ac.id/index.php/JTIKA/

[11] Halvani, O., Winter, C., & Graner, L. (2018). Unary and Binary Classification Approaches and their Implications for Authorship Verification. *Litbang : Media Penelitian Dan Pengembangan*, 1–15.

[12] Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3–29.

[13] Tian, Y., Shi, Y., & Liu, X. (2012). Recent advances on support vector machines research. *Technological and Economic Development of Economy*, *18*(1), 5–33.

[14] Susanti, D. H., Maritim, U., Ali, R., Pratiwi, D., Maritim, U., Ali, R., Hani, F., Wahyuni, S., Maritim, U., & Ali, R. (2022). Implementasi kebijakan pkh dalam rangka mengatasi kemiskinan di kecamatan rowokangkung dimasa pandemi. *Jurnal Hukum, Politik Dan Ilmu Sosial*, *1*(2), 38–51.