IAIC International Conference Series
Vol. 4, No. 1, 2023, pp. 115 ~ 121
E-ISSN: 2774-5899 | P-ISSN: 2774-5880
DOI: https://doi.org/10.34306/conferenceseries.v4i1.639

Predictions using Support Vector Machine with Particle Swarm Optimization in Candidates Recipient of Program Keluarga Harapan

[1*]Arie Satia Dharma, [2]Evi Rosalina Silaban, [3]Hana Maria Siahaan
[1,2,3]Informatics, Institut Teknologi Del, Indonesia
e-mail: [1*]ariesatia@del.ac.id, [2]evi@gmail.com,
***Corresponding author**

(Received November 8, 2023 Revised November 17, 2023 Accepted November 25, 2023, Available online December 19, 2023)

Abstract

Program Keluarga Harapan (PKH) is a conditional social assistance program as an effort to alleviate poverty which is allocated to poor vulnerable households. The determination of candidates for the Program Keluarga Harapan assistance recipients is still carried out in village meetings, so it takes quite a long time and there is potential for subjectivity in the assessment carried out by Village Government officials which can lead to differences of opinion between deliberation participants in assessing the eligibility of residents as PKH recipients. For this reason, this research will use an optimization method, namely Particle Swarm Optimization (PSO) to select the most optimal attribute out of 39 attributes. After that, a classification algorithm, namely the Support Vector Machine (SVM), was chosen to form a classification model for Candidates for Social Assistance for the Program Keluarga Harapan (PKH). The classification of Candidates for Social Assistance Recipients of the Program Keluarga Harapan (PKH) was carried out in 2 experiments, namely before and after optimization. Experiments before optimization give an accuracy value of 92.44%. While the Support Vector Machine accuracy value after optimization gives an accuracy value of 92.51%. Based on the experimental results, it can be concluded that the Particle Swarm Optimization method can increase the accuracy of the Support Vector Machine algorithm by 0.07%. And the best model is the Support Vector Machine after optimizing Particle Swarm Optimization by using the 17 most optimized attributes in determining class targets.

Keywords: *Program Keluarga Harapan, Particle Swarm Optimization, Support Vector Machine, PSO, SVM*

1. Introduction

Poverty is a major social problem experienced by developing countries, especially Indonesia [1]. According to the Central Bureau of Statistics, poverty is a condition of being unable to meet basic needs, both in the form of food and clothing, economically [2]. The government continues to make efforts to create various social protection programs that are planned, directed and sustainable with the aim of improving the living standards of underprivileged families [3]. This government program consists of assistance to individuals, families, and poor vulnerable groups in the categories of education, food, energy, health, socio-economic, agriculture, marine, fisheries and housing [4]. Program Keluarga Harapan (PKH) is a conditional social assistance program as an effort to alleviate poverty which is allocated to vulnerable poor households which consists of three components, namely the health component, the education component and the social welfare component [5].

Based on information obtained from Village Government Officials, the process of determining potential recipients of the Program Keluarga Harapan (PKH) currently still relies on village deliberations. There are several challenges that need to be overcome in this process [6]. Firstly, the village deliberation process takes quite a long time and secondly there is the potential for subjectivity in the assessment carried out by village government officials which can lead to differences of opinion between deliberation participants in assessing the suitability of residents as PKH recipients. Third, the government, in this case the social service, already has criteria for determining the eligibility of PKH recipients based on the form that prospective PKH recipients must fill out. Therefore, this research aims to help determine potential PKH recipients by predicting the eligibility of PKH beneficiaries [7].

In overcoming this problem, one approach that can be used to overcome this problem is to utilize the science of artificial intelligence such as machine learning [8]. Machine learning is the science of artificial intelligence to study data so that it can work according to instructions [9]. Supervised learning is a method in machine learning that is intended for mapping an input into an output whose value has been provided [10]. In Supervised Learning, training data containing input and output information will be given to the system, so that the system will process the data that has been entered. By looking for existing data patterns, based on these patterns it will become a reference for the next data set [11].

There are many machine learning algorithms that can be used to study previous data models, one of which is the Support Vector Machine (SVM). The Support Vector Machine (SVM) algorithm was chosen because based on previous research entitled "Naive Bayes Algorithm, Decision Tree, and SVM for Classification of Sharia Cooperative Customer Financing Approval" a system was obtained that could predict the credibility of future prospective customers with accuracy the Naive Bayes algorithm is 77.29%, the Decision Tree is 89.02% and the highest is the Support Vector Machine (SVM) 89.86% [12]. Support Vector Machine (SVM) also has high accuracy results with a fairly low error rate and can be used for classification, but besides these advantages SVM also has weaknesses in selecting optimal parameters [13]. For the optimization method, researchers will use the Particle Swarm Optimization (PSO) method. This method was chosen because Particle Swarm Optimization (PSO) can optimize problems in selecting optimal attributes in a classification model resulting in increased accuracy [14]. Based on research conducted by Musyaffa & Rifai (2018), with the research title "Support Vector Machine Model Based on Particle Swarm Optimization for Liver Disease Prediction" the SVM classification model resulted in an accuracy of 71.36%, while the SVM classification model with PSO with an accuracy of 77, 36%, which shows that the PSO-based SVM model has a better accuracy of 6% [15].

Based on this, researchers will create a classification system that will assist village government officials in determining PKH beneficiary candidates using a machine learning approach that will use a Support Vector Machine (SVM) algorithm based on Particle Swarm Optimization (PSO). Using the Support Vector Machine (SVM) algorithm based on Particle Swarm Optimization (PSO) is expected to increase accuracy compared to only using the Support Vector Machine (SVM) algorithm [16].

2. Research Method

Support Vector Machine (SVM) is a Supervised Machine Learning algorithm that is intended for handling Classification and Regression problems [17]. This algorithm works to characterize each data item as a notation in n-dimensional space (where n is the number of attributes in the data set) with the value of each attribute being a certain coordinate value [18]. SVM in classification aims to find hyperplane values to divide into 2 classes, then the data will be classified with the large values of both classes properly [19]. Hyperplane is a dividing line that serves as a separator between classes. In general, the goal of a Support Vector Machine is to maximize the hyperplane boundaries. The formula number 1, 2, and 3 is used to calculate the hyperplane [20] as follows.

$$w * x + b = 0 \quad (1)$$

$$\frac{1}{2}||w||^2 \quad (2)$$

With the provision of,

$$y_i \, (w * x_i + b) \geq 1 \quad (3)$$

Particle Swarm Optimization (PSO) was introduced by Dr. Eberhart and Dr. Kennedy in 1995 [21]. Particle Swarm Optimization (PSO) is an optimization algorithm that mimics the processes that occur in the survival of populations of birds (flock of birds) and fish (schools of fish). In Particle Swarm

Optimization (PSO), swarms are assumed to have a certain size with each particle initially located at a random location in a multidimensional space. Each particle is assumed to have two characteristics, namely position and velocity [22]. Each particle moves in a certain space or space and remembers the best position that has been passed or found with respect to food sources or objective function values [23]. Each particle conveys information or the best position to the other particles and adjusts the position and speed of each based on the information received about the good position [24]. The stages in Particle Swarm Optimization for optimizing through feature selection are as follows;

1. Initialization
   Assume that the size of the herd or herd (number of particles) is N. The initial velocity and position of each particle in N dimensions is determined randomly in the range of values [0,1]. This stage is the initial stage which aims to create a population of particles that are candidate solutions to find the best solution among these candidates.
2. Update Velocity
   After finding pbest and gbest on initialization, then each particle will move so that a new speed is needed so that the particles can move. The equation number 4 used for updating the particle velocity is as follows.
   $$v_i(t+1)=w*v_i(t)+c1*rand_1*(pbest_i(t)-x_i(t)+c2*rand_2(gbest_i(t)-x_i(t)) \quad (4)$$
3. Update Position
   Particles that already have a new speed will move to a position that is different from the previous position so that the particle's position needs to be updated.
   The equation number 5 used to update the particle position is as follows.
   $$x_i(t+1)=x_i(t) + v_i(t+1) \quad (5)$$
   The xmin and xmax threshold values are [0.1]. If the current position value x[i] is less or exceeds the xmin and xmax limits, then the current position value x[i] will be repaired or reflected so that it is within the specified limits, with the formula number 6.
   $$x[i+1] = max(min(xi,\ 0), 1) \quad (6)$$
4. Update Pbest and Gbest
   After the particle has a new position, then the pbest and gbest values must also be updated. The new pbest is obtained based on the smallest fitness value, while the gbest value is obtained from the updated smallest pbest value.
   In determining whether a feature will be selected, a threshold is needed as a limit that if the position is greater than the threshold the feature will be selected or represented to a value of 1, if it is smaller than the threshold then the feature is not selected or represented to a value of 0. This determination uses the number 7 equation:
   $$x_i = \begin{cases} 1, & \text{if } x_i > 0.5 \\ 0, & others \end{cases} \quad (7)$$
   Then optimization is carried out with the objective function to find the fitness value of the number of features that have been selected using the number 8 equation:
   $$f(x) = \alpha * (1 - P) + (1 - \alpha) * \frac{NSelected}{NFeatures} \quad (8)$$
   α as a parameter used to control the contribution between classification performance and the number of features selected.

Hyperparameters are parameters that are used to control or regulate the behavior of machine learning algorithms during the training process [25]. Each algorithm has a hyperparameter value that can be adjusted according to needs. The selection of the right hyperparameters can greatly affect the performance and results of the created machine learning algorithms. Following are some of the hyperparameters contained in the SVM algorithm:

1. Kernel
   The kernel is a transformation function that is useful for changing data from one representation to another by transforming data into a higher feature space, when linear separation is not possible. SVM has four types of kernels namely Linear, Polynomial, Radial basis function (RBF) as well as Sigmoid and linear kernel the best solution among existing kernel types [26].
2. Parameter C

The regularization parameter controls the tradeoff between achieving low training error and low test error. Generally, when C is small, the margin is maximized (wider) and the number of misclassified samples increases (small penalty) whereas when C is large the margin width is minimized (smaller) and the number of misclassified samples is smaller (large penalty) [27].

Confidence score is a measure or score that indicates the level of confidence or trust in the model for the predictions made [28][29]. The confidence score can also be used as a standard for setting priorities [30]. The confidence score is generated by calculating the distance of the new data to the hyperplane formed by the SVM model. The formula number 9 used to calculate the distance is.

$$distance = (w^T * x) + b \quad (9)$$

After getting the distance, then the distance will be converted into a confidence score using the function or transformation using the sigmoid function using the formula number 10.

$$confifence\_score = 1/(1 + np.exp(-distance)) \quad (10)$$

3. Results and Analysis

Before selecting the optimal attributes, there are 40 attributes that are carried out in building the model using the Support Vector Machine. The variables used in this study are divided into two categories, namely the independent and dependent categories. The following 40 variables used can be seen in Table 1 Data Variable

Table 1. Variable Data

No	Variable	Type	No	Variable	Type
1.	jumlah_anggota_rumah_tangga	Numeric	21.	ada_sepeda	Category
2.	kepemilikan_rumah	Category	22.	ada_motor	Category
3.	kepemilikan_lahan	Category	23.	ada_mobil	Category
4.	jenis_lantai	Category	24.	ada_perahu	Category
5.	jenis_dinding	Category	25.	ada_kapal	Numerik
6.	jenis_atap	Category	26.	jumlah_sapi	Numerik
7.	jumlah_kamar	Numerik	27.	jumlah_kerbau	Numerik
8.	sumber_air_minum	Category	28.	jumlah_kuda	Numerik
9.	cara_peroleh_air_minum	Category	29.	jumlah_kambing	Numerik
10.	sumber_penerangan_utama	Category	30.	jumlah_babi	Category
11.	fasilitas_buang_air_besar	Category	31.	status_usaha_art	Category
12.	buang_tinja	Category	32.	status_kks	Category
13.	ada_tabung_gas	Category	33.	status_kip	Category
14.	ada_lemari_es	Category	34.	status_kis	Category

15.	ada_ac	Category	35.	status_bpjs_mandiri	Category
16.	ada_pemanas	Category	36.	status_jamsostek	Category
17.	ada_telepon	Category	37.	status_asuransi	Category
18.	ada_tv	Category	38.	status_rastra	Category
19.	ada_emas	Category	39.	status_kredit_usaha_ rakyat	Category
20.	ada_laptop	Category	40.	status_pkh	Category

After implementation using 40 attributes, the accuracy value is obtained which can be seen in Table 3. Results of the Classification Model. From the table it can be obtained information that the accuracy value obtained before selecting the optimal attribute has an average accuracy value of 92.44% which is a very good classification. After obtaining the results of the support vector machine model without optimization, SVM modeling will then be carried out using PSO optimization to obtain the most optimal sub-attribute results consisting of 17 attributes as shown in Table 2 Best Attributes.

Table 2. Best Attribute

No	Variable	Type	No	Variable	Type
1.	jumlah_anggota_rumah_ tangga	Numeric	10.	sumber_penerangan_ utama	Category
2.	kepemilikan_rumah	Category	11.	fasilitas_buang_air_ besar	Category
3.	kepemilikan_lahan	Category	12.	buang_tinja	Category
4.	jenis_lantai	Category	13.	ada_tabung_gas	Category
5.	jenis_dinding	Category	14.	ada_lemari_es	Category
6.	jenis_atap	Category	15.	ada_ac	Category
7.	jumlah_kamar	Category	16.	ada_pemanas	Category
8.	sumber_air_minum	Category	17.	ada_telepon	Category
9.	cara_peroleh_air_minum	Category			

After implementation using 17 optimal attributes, the accuracy value is obtained which can be seen in Table 3. Results of the Classification Model. From the table it can be obtained that the accuracy value obtained after selecting optimal attributes has an average accuracy value of 92.51% which is a very good classification. Based on Table 3. The results of the Classification Model can be seen that after using the optimal attribute selection, the model can increase the accuracy value significantly so as to produce a better value. The accuracy value obtained from the support vector machine classification algorithm and support vector machine with optimization after using the optimal attribute shows an increase in accuracy value. An increase in the accuracy value can occur because the attribute values used are the optimal attributes in influencing the determination of PKH assistance beneficiary candidates. From these results it can be seen that the accuracy value of the support vector machine algorithm with optimization is better than the support vector machine algorithm even though there is no significant difference. The following is Table 3. Classification Model Results.

Table 3. Classification Model Result

Accuracy	SVM without optimization	SVM with optimization
Model Accuracy	92.44%	92.51%
Validasi Accuracy	86.70%	91.10%

In the context of a Support Vector Machine model without optimization, overfitting tends to occur due to the lack of an effective mechanism for managing complexity and feature selection. However, using PSO optimization, the problem of overfitting can be overcome by finding relevant subsets of attributes, eliminating unimportant attributes, and reducing model complexity.

In conclusion, the implementation of Particle Swarm Optimization optimization on the Support Vector Machine model has brought significant benefits. The resulting model is easier to interpret, has better generalization abilities, and is able to overcome overfitting. In the context of classifying data, the use of Particle Swarm Optimization optimization is an effective approach to improve model performance and obtain more accurate predictions on new data that have never been seen before.

4. Conclusion

Based on the results obtained, it can be concluded that this study implemented the Support Vector Machine (SVM) algorithm and the Particle Swarm Optimization optimization method for the classification of PKH social assistance recipient candidates.

Particle Swarm Optimization is able to improve the accuracy of the Support Vector Machine algorithm by finding the 17 most optimal attributes that can be used to classify PKH beneficiary candidates.

The Support Vector Machine (SVM) model is prone to overfitting with a training accuracy of 92.44% and test data performance of 86.70%. However, with Particle Swarm Optimization optimization, the SVM model has a training accuracy of 92.51% and new data accuracy of 91.10%, showing good generalization ability in real-world situations.

References

[1] Lestari U, Targiono M. Sistem Pendukung Keputusan Klasifikasi Keluarga Miskin Menggunakan Metode Simple Additive Weighting (Saw) Sebagai Acuan Penerima Bantuan Dana Pemerintah (Studi Kasus: Pemerintah Desa Tamanmartani, Sleman). J TAM (Technology Accept Model. 2017;8(1).

[2] M. A. A. Budi, M. G. L. Putra, and L. H. Atrinawati, "Improving Helpdesk Capability in Perum Peruri Through Service Catalog Management Based on ITIL V3," *Int. J. Cyber IT Serv. Manag.*, vol. 2, no. 2, pp. 117–126, 2022.

[3] Ghahremani nahr J, Nozari H, Sadeghi ME. Artificial intelligence and Machine Learning for Real-world problems (A survey). Int J Innov Eng. 2021;1(3).

[4] M. H. R. Chakim, M. Hatta, A. Himki, A. R. A. Zahra, and N. N. Azizah, "The Relationship Between Smart Cities and Smart Tourism: Using a Systematic Review," *ADI J. Recent Innov.*, vol. 5, no. 1Sp, pp. 33–44, 2023.

[5] Abijono H, Santoso P, Anggreini NL. ALGORITMA SUPERVISED LEARNING DAN UNSUPERVISED LEARNING DALAM PENGOLAHAN DATA. J Teknol Terap G-Tech. 2021;4(2).

[6] R. Rarmizi, I. Y. Nasaruddin, and N. Hidayah, "Analysis of the influence of corporate governance on the financial performance of Islamic banks in Indonesia 2016-2021," *APTISI Trans. Manag.*, vol. 7, no. 2, pp. 179–190, 2023.

[7] Ngurah G, Susena E, Furqon MT, Wihandika RC. Optimasi Parameter Support Vector Machine (SVM) dengan Particle Swarm Optimization (PSO) Untuk Klasifikasi Pendonor Darah Dengan Dataset RFMTC. J Pengemb Teknol Inf dan Ilmu Komputer,. 2018;2(12).

[8] R. D. Affandi, H. Pratiwi, and M. I. Sa'ad, "Application of the SMARTER Method in Determining the Whitening of Study Permits and Teacher Study Tasks," *Aptisi Trans. Technopreneursh.*, vol. 5, no. 2, pp. 205–215, 2023.

[9] Musyaffa N, Rifai B. Model Support Vector Machine Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Liver. JITK (Jurnal Ilmu Pengetah Dan Teknol Komputer). 2018;3(2).

[10] T. A. D. Lael and D. A. Pramudito, "Use of Data Mining for The Analysis of Consumer Purchase

Patterns with The Fpgrowth Algorithm on Motor Spare Part Sales Transactions Data," *IAIC Trans. Sustain. Digit. Innov.*, vol. 4, no. 2, pp. 128–136, 2023.
[11] Freitas D, Lopes LG, Morgado-Dias F. Particle Swarm Optimisation: A historical review up to the current developments. Vol. 22, Entropy. 2020.
[12] U. Rahardja, "Using Highchart to implement business intelligence on Attendance Assessment system based on YII Framework," *Int. Trans. Educ. Technol.*, vol. 1, no. 1, pp. 19–28, 2022.
[13] Gad AG. Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. Arch Comput Methods Eng. 2022;29(5).
[14] U. Rahardja, Q. Aini, D. Manongga, I. Sembiring, and I. D. Girinzio, "Implementation of Tensor Flow in Air Quality Monitoring Based on Artificial Intelligence," *Int. J. Artif. Intell. Res.*, vol. 6, no. 1, 2023.
[15] Fajri M, Primajaya A. Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search. J Appl Informatics Comput. 2023;7(1).
[16] Goel A, Srivastava SK. Role of kernel parameters in performance evaluation of SVM. In: Proceedings - 2016 2nd International Conference on Computational Intelligence and Communication Technology, CICT 2016. 2016.
[17] S. Kosasi, U. Rahardja, I. D. A. E. Yuliani, R. Laipaka, B. Susilo, and H. Kikin, "IT Governance: Performance Assessment of Maturity Levels of Rural Banking Industry," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, IEEE, 2022, pp. 1–6.
[18] Gerds TA, van de Wiel MA. Confidence scores for prediction models. Biometrical J. 2011;53(2).
[19] A. Lubis, I. Iskandar, and M. M. L. W. Panjaitan, "Implementation of KNN Methods And GLCM Extraction For Classification Of Road Damage Level," *IAIC Trans. Sustain. Digit. Innov.*, vol. 4, no. 1, pp. 1–7, 2022.
[20] Berthon A, Han B, Niu G, Liu T, Sugiyama M. Confidence Scores Make Instance-dependent Label-noise Learning Possible. In: Proceedings of Machine Learning Research. 2021.
[21] Putra MI. Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest pada BRI Kantor Cabang Pelaihari. *J Tek Inform dan Sist Inf. 2019*;13(1).
[22] Riaddy, Aditya I, Sibaroni Y, Aditsania A. Ekstraksi Informasi pada Makalah Ilmiah dengan Pendekatan Supervised Learning. e-Proceeding Eng. 2019;3(1).
[23] K. Kano and E. Dolan, "Data Compression Analysis of Multimedia Video on Demand and DEMAND TV Broadcast Systems on the Network," *Int. J. Cyber IT Serv. Manag.*, vol. 3, no. 1, pp. 48–53, 2023.
[24] Achyani YE. Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung. J Inform. 2018;5(1).
[25] Fauziah, Tiro MA, Ruliana. Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua. ARRUS J Math Appl Sci. 2022;2(2).
[26] Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. Soft Comput. 2018;22(2).
[27] Qiuyun T, Hongyan S, Hengwei G, Ping W. Improved Particle Swarm Optimization Algorithm for AGV Path Planning. IEEE Access. 2021;9.
[28] Xie T, Yao J, Zhou Z. DA-based parameter optimization of combined kernel support vector machine for cancer diagnosis. Processes. 2019;7(5).\
[29] Chardin D, Gille C, Pourcher T, Humbert O, Barlaud M. Learning a confidence score and the latent space of a new supervised autoencoder for diagnosis and prognosis in clinical metabolomic studies. BMC Bioinformatics. 2022;23(1).
[30] Nurajijah N, Riana D. Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah. J Teknol dan Sist Komput. 2019;7(2).