Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13208
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison PSO And IWPSO Performance In Optimizing Decision Tree Algorithm On Heart Disease Dataset

Inggit Dwi Oktaviani[1)*], Ferian Fauzi Abdulloh[2)]
[1,2)]Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
[1)]inggitdwioktaviani@students.amikom.ac.id, [2)]ferian@amikom.ac.id

Submitted : Nov 29, 2023 | **Accepted** : Dec 14, 2023 | **Published** : Jan 1, 2024

Abstract: Heart disease, one of the most common and potentially fatal chronic diseases, has become a major focus in global health efforts. In this study, researchers used the decision tree algorithm on the heart disease dataset with the stages of the decision algorithm including the EDA, Split Data, and Decision tree modeling stages. Furthermore, hyperparameters use PSO and IWPSO to optimize the algorithm. The purpose of this research is to analyze the performance of Particle Swarm Optimization (PSO) and Inertia Weight Particle Swarm Optimization (IWPSO) in heart disease prediction based on relevant datasets. PSO and IWPSO were applied to the heart disease dataset, with the results showing an accuracy rate of 78% for PSO and 84% for IWPSO. These results indicate that IWPSO provides significant performance improvement compared to PSO in the context of heart disease prediction. The implications of these findings can support the development of more efficient prediction systems for early detection of heart disease, making a positive contribution to prevention efforts and further treatment of this critical health condition. In addition, the purpose of this research is to continue research in the form of C4.5 on heart disease with a result of 80.43%. In this study, IWPSO got the best accuracy of 84.23% greater than previous research. The results of this study are to provide insight that PSO and IWPSO hyperparameters can optimize decision trees in handling heart disease datasets and continue research.

Keywords: Comparison, Decision Tree, Heart Disease, IWPSO, PSO

INTRODUCTION

One of the important organs in the human body is the heart. The human heart is in the chest cavity and has 4 chambers. The rooms are the right and left ventricles and left and right and left atria, where the ventricles are below the atria. At heart there are valves that are useful so that there is no mixing of blood. blood. The heart is included in the cardiovascular system, which is where this system has circulations, for example, there is pulmonary circulation (pulmonary). The heart as an important organ is not inevitable from diseases that can attack it (Aniamarta et al., 2022).

Heart attack is a condition where the blood flow of the coronary arteries blood flow is stopped so that the heart muscle is deprived of oxygen, causing an infarction. So, a heart attack is also referred to as an acute myocardial infarction. This heart attack is an urgent emergency so that This heart attack is an urgent emergency so that it requires proper and fast handling which is useful so that the heart damage is not too severe heart damage is not too severe. For example, in treatment there are various articles that discuss such as treatment patterns, cardiac rehabilitation, compliance with blood pressure control, reperfusion therapy. In addition, there are also various studies that link technology with heart attacks so

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13208 p-ISSN : 2541-044X

that heart attacks so that patients can be treated more quickly. Whereas it has been said earlier that this disease requires proper diagnosis and rapid treatment (Aniamarta et al., 2022).

In this research, the solution used in the classification of heart disease is the application of data analysis and machine learning techniques. and machine learning techniques. Classification methods, such as Decision tree, which can prove effective in identifying patterns that may signal the presence of phishing emails. In identifying a person with heart disease. Decision trees are used to handle categorical data in the context of heart disease analysis. In previous research classifying the same thing, namely the classification of heart disease data using the decision tree algorithm with rapid miner. The heart disease data that has been obtained is carried out the EDA process and then decision tree modeling is carried out. The performance of the decision tree was obtained at 80.43% (Agus Oka Gunawan et al., 2023). So, I want to continue this research with the same algorithm and dataset but with the context of hyperparameter comparison that I will do in this research.

The effectiveness of each hyperparameter in handling the heart disease dataset is aimed at optimization of the decision tree model. PSO and IWPSO were chosen because they help optimize the distribution and processing speed of the decision tree model. IWPSO itself is an improvement of PSO where IWPSO optimizes the Inertia Weight of the PSO. The relationship between PSO and IWPSO in decision tree is to find the optimal combination of parameters in the parameter space of the decision tree algorithm, such as the maximum depth of the tree and the minimum number of samples for splitting. This It is hoped that the results of this research will provide deeper insight into the advantages and limitations of each variant in classifying emails. PSO and IWPSO in classifying heart disease. The stages of this method itself start from data collection, EDA, visualization, data splitting, modeling, and hyperparameter. The targeted output of this research is a better understanding of how PSO and IWPSO work on heart disease datasets.

With a better understanding of the performance improvement of decision trees as well as their optimization with PSO and IWPSO in the context of heart disease detection, it is expected that more effective and adaptive heart disease prediction measures can be implemented. it is expected that more effective and adaptive heart disease prediction measures can be implemented. This research can contribute to the development of early detection techniques for new heart disease data, which in turn will aid in the prediction and adaptive management of heart disease. which in turn will aid prediction and treatment. The results of this study can provide practical guidance for medical professionals and researchers in selecting the most suitable hyperparameters for decision tree model tasks.

LITERATURE REVIEW

This study was conducted to compare between PSO and IWPSO as a hyperparameter in decision tree optimization of heart disease data for. In this modeling, classification is used which is the process of finding a class model that will be categorized. PSO is based on the idea of a colony of particles moving in the solution search space. IWPSO is a variation or improvement of PSO that includes a weighting element on the particles in the colony (Setiawan et al., 2019).

Research by (Agus Oka Gunawan et al., 2023), Researchers use the algorithm decision tree. In this study, researchers predicted heart disease. With the decision tree algorithm, researchers got 80.43% accuracy. Research by (Riansyah et al., 2023), In this study, researchers conducted research on improving the accuracy of the C5.0 decision tree using Adaboost. From the results obtained before optimizing the decision tree, the accuracy was 80.58%, while when optimization was carried out there was an increase of 82.98%. Research by (Juliane, 2023), This study conducted research on the comparison between decision tree and naive bayes on diabetes dataset. The final result obtained is that the decision tree gets the best accuracy with croton optimization of 91.30%. Research by (Kristiyanti & Normah, 2019), This research applies PSO to optimize 2 algorithms, namely SVM and Naive Bayes. The results obtained from this study show an increase in each algorithm with the best performance being the SVM algorithm of 82.85% when PSO is performed. Research by (Purwaningsih, 2019), This research conducted PSO optimization to optimize 2 algorithms, namely SVM and Neural Network. The results obtained from this study show an increase in each algorithm with the best performance being the Neural Network algorithm of 84.55% when PSO is performed. Research by (Sekyere et al., 2023), In this study, researchers conducted research to improve the performance of PSO by optimizing the inertia

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13208 p-ISSN : 2541-044X

weights in the PSO or IWPSO. The results obtained show the effectiveness of the inertial weight function in PSO in optimizing the problem and can produce better performance than linear or exponential sigmoid-based functions.

METHOD

The research aims to compare the performance of PSO and IWPSO in decision tree algorithm optimization. The data used in this study are data heart disease data obtained from open source Kaggle with the name Heart disease. This dataset contains personal data of patients affected by heart disease and not affected by heart disease. The label of this data itself contains two features, namely 0 identifies normal patients and 1 identifies patients with heart disease. The results of this study are used as a comparison of PSO and IWPSO and optimization of the decision tree algorithm. The program that I did, I program on google colab with the research flow with the research flow in the figure below.

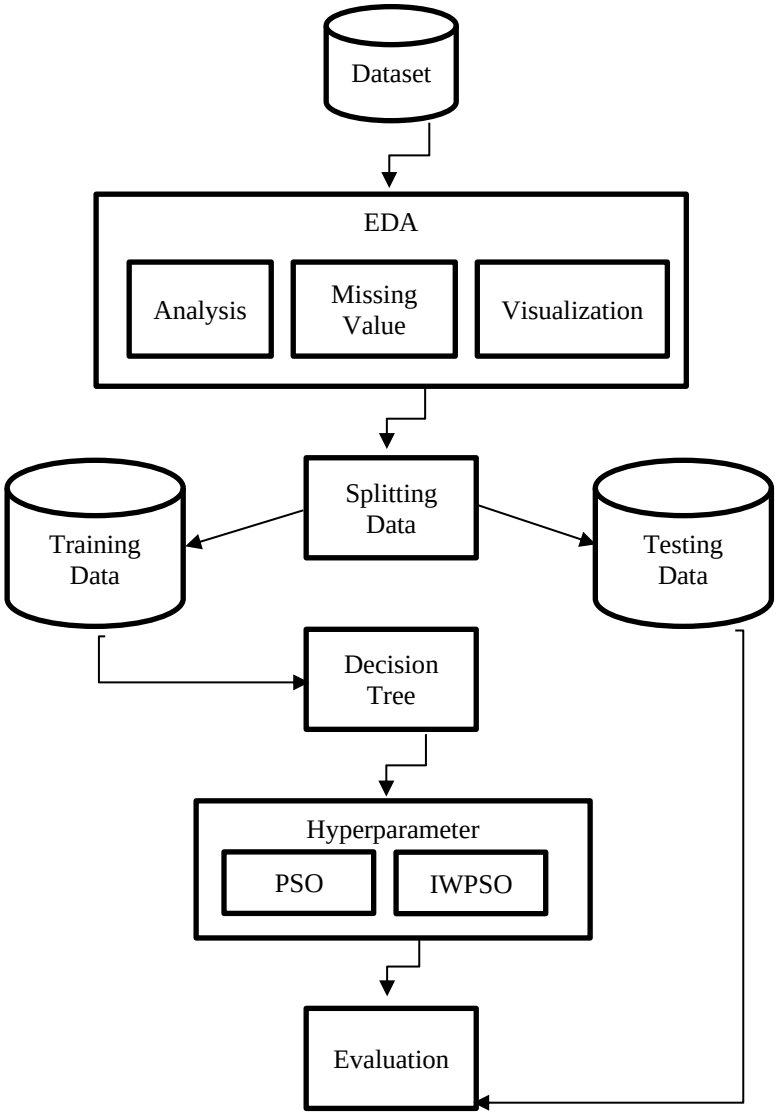

Fig 1. Framework of Research

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13208 p-ISSN : 2541-044X

Dataset

The research collected data from an open source dataset on Kaggle with the title Heart Disease. This data contains data related to heart disease in the form of CSV files containing patient data. The attributes of this data contain string data such as gender, Chest Pain Type, resting electrocardiogram, the slope of the peak exercise and exercise-induced angina. Then there is numeric data, namely age, resting blood pressure, Cholesterol, fasting blood sugar, maximum heart rate achieved, Numeric value measured in depression and finally the label of this data is Heart disease. The patient data in this file 918 data. It contains 11 attributes in the data and 1 label that contains 2 variables, namely normal and heart disease. This amount of data does not contain empty data and has the same amount. The amount of patient data that has different types of data needs to be processed so that it can be analyzed. so that it can be analyzed.

EDA

Exploratory Data Analysis (EDA) is an approach in statistics and data science that aims to understand the structure and characteristics of data in more depth before conducting further statistical analysis or developing predictive models. EDA helps data analysts to formulate questions, identify patterns, and find empty data in the data set. At this stage, researchers analyze the data to find out the relationship between attributes and labels, examine the data starting from the data content, data type, and check for empty data. From this stage, a visualization of the data distribution is also obtained.

Splitting Data

Data splitting is the process of dividing a dataset into two or more subsets referred to as training set and testing set. The main purpose of splitting data is to train a machine learning model on a portion of the data and test its performance on another portion that has never been seen before. In this research, the split data is divided into training and testing where the training data is 80% and testing is 20%.

Decision Tree

Decision Tree is a predictive model that takes the form of a decision tree, where each node in the tree represents a decision or prediction based on the features of the input data. The decision tree variation used in this study is the default variation of the library, namely CART. CART (Classification and Regression Trees) is one type of Decision Tree algorithm used for classification and regression problems (Widiyati et al., 2018).

$$G(t) = 1 - \sum_{i=1}^{K}(p_i)^2 \quad (1)$$

PSO

Particle Swarm Optimization (PSO) is an optimization algorithm inspired by the behavior of particle swarms. Each particle in PSO represents a potential solution in the search space and moves through that space with the goal of finding the optimal solution (Purwaningsih, 2019).

$$v_{ij} = w * v_{ij} + c_1 * r_1 * \left(Pbest_{ij} - x_{ij}\right) + c_2 * r_2 * (Gbest_j - x_{ij}) \quad (2)$$

IWPSO

Inertia Weighted Particle Swarm Optimization (IWPSO) is a variation of the Particle Swarm Optimization (PSO) algorithm that introduces a weighting element to the particles in the colony. This weighting aims to improve the performance of the PSO algorithm. IWPSO has the same formula and flow as PSO but the particle weights are updated for each weight. (Sekyere et al., 2023).

$$v_{ij} = w_i * v_{ij} + c_1 * r_1 * \left(Pbest_{ij} - x_{ij}\right) + c_2 * r_2 * (Gbest_j - x_{ij}) \quad (3)$$

Evaluation

In this last stage, the results will be obtained in the form of 3 results, the first is the performance of the decision tree before hyperparameter, the second is the result of the decision tree when optimized with PSO and the third is the result of the decision tree when optimized with IWPSO.

RESULT

In the initial stage, insert the data that has been obtained from Kaggle and then import the library to carry out the program, here researchers use hyponic to call PSO and IWPSO. After the library is installed the data will be checked. The calculated data amounted to 918 data. Before modeling the data obtained, it was analyzed first. This process consists of several parts, namely checking the data type, checking for missing or empty data, and visualizing it. Clean data is divided into 80% training data and 20% testing data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13208
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 Raw Data

Age	Sex	CPT	RBP	CL	FBP	RECG	MaxHR	Execise A	Oldpeak	ST	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
…	…	…	…	…	…	…	…	…	…	…	…

The data is checked for data types and null data in order to classify the model.

Table 2Missing Value Data

Age	False
Sex	False
ChestPainType	False
RestingBP	False
Cholesterol	False
FastingBS	False
RestingECG	False
MaxHR	False
ExerciseAngina	False
Oldpeak	False
ST_Slope	False
HeartDisease	False

Table 3 Data Type

Column	Non-Null Count	Dtype
Age	918 non-null	int64
Sex	918 non-null	object
ChestPainType	918 non-null	object
RestingBP	918 non-null	int64
Cholesterol	918 non-null	int64
FastingBS	918 non-null	int64
RestingECG	918 non-null	object
MaxHR	918 non-null	int64
ExerciseAngina	918 non-null	object
Oldpeak	918 non-null	float64
ST_Slope	918 non-null	object
HeartDisease	918 non-null	int64
Dtypes : float64(1), int64(6), object(5)		

After checking the data researcher check function in summary data, and the next step is data visualization to analyze the distribution of the data in it. Label 0 is for normal patients and 1 for heart disease patients.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13208
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 4 Summary Data

	count	mean	std	min	25%	50%	75%	max
Age	918	53.51	9.43	28	47	54	60	77
RestingBP	918	132.396	18.514	0	120	130	140	200
Cholesterol	918	198.799	109.384	0	173.25	223	267	603
FastingBS	918	0.233	0.423	0	0	0	0	1
MaxHR	918	136.809	25.46	60	120	138	156	202
Oldpeak	918	0.887	1.066	-2.6	0	0.6	1.5	6.2
HeartDisease	918	0.553	0.497	0	0	1	1	1

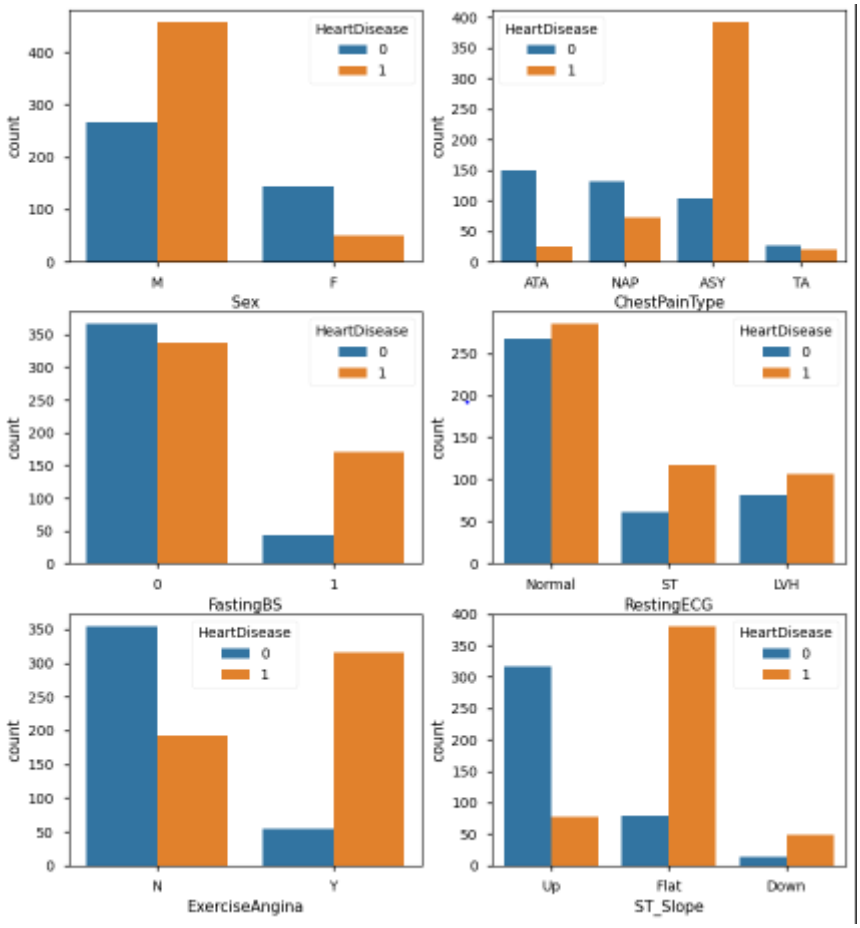

Fig 2 Distributions Data

After EDA is completed, the data will be modeled by Decision tree Classifier. At this stage, researchers use 80% training data and 20% testing data. The data is separated into x and y. In the x variable, the data used are all attributes in the dataset, int attributes are not encoded while those with object types are encoded so that they can be used in the model. For variable y contains label data, namely HeartDisease. Tests taken to evaluate the model include accuracy, precision, and F1-Score.

Table 5 Result Modelling Decision Tree

Decision Tree	
Accuracy	71.74%
Precision	79.2%
F1_Score	75.47%

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13208
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The table shows the performance of the decision tree before optimization. Then at this stage PSO and IWPSO optimization is carried out on this algorithm. This optimization also compares the speed between these hyperparameters to see the best performance of both and include hyperparameters from PSO, IWPSO, and Decision tree.

Hyperparameter Decision Tree Clasifier = {criterion': ['gini', 'entropy'], 'splitter': ['best', 'random'], 'max_depth': range(1, 20), 'min_samples_split': range(2, 20), 'min_samples_leaf': range(1, 20)}

Hyperparameter [PSO,IWPSO] = {'epoch': 50, 'population_size': 50,}

Table 6 Result Modeling After PSO

Decision Tree + PSO	
Accuracy	78.26%
Precision	81.44%
F1_Score	79.79%
Time Taken	7.51s

Table 7 Result Modeling After IWPSO

Decision Tree + IWPSO	
Accuracy	84.23%
Precision	88.29%
F1_Score	85.13%
Time Taken	9.14s

DISCUSSIONS

From the results that have been obtained, there are differences in each result that can be compared between the results before hyperparameter and when hyperparameter is performed as follows.

Table 8 Comparison Results Modelling

	Accuracy	Precision	F1_Score	Time Taken
Decision Tree	71.74%	79.2%	75.47%	-
==========	==========	==========	==========	==========
Decision Tree + PSO	78.26%	81.44%	79.79%	7.51s
Increased	**6.52%**	**2.24%**	**4.32%**	
==========	==========	==========	==========	==========
Decision Tree + IWPSO	84.23%	88.29%	85.13%	9.14s
Increased	**12.49%**	**9.09%**	**9.66%**	

From the table above, it can be seen that the model has improved performance after optimization. Decision tree when PSO optimization has increased slightly but at a faster speed than IWPSO, while Decision tree when optimized with IWPSO has increased much more than PSO but the time required is slightly longer than PSO. So the best performance results are produced by IWPSO with an accuracy of 84.23% greater than previous research examining heart disease data using C4.5 of 80.43% (Agus Oka Gunawan et al., 2023).

The limitation of this research is that it has not compared with other algorithms and is only limited to heart disease datasets.

*name of corresponding author

CONCLUSION

Researchers concluded that the results of PSO and IWPSO have their respective advantages. The results produced by PSO are very fast than IWPSO but the resulting performance is lower than IWPSO. The best performance result obtained by IWPSO is 84.23%. Then the best speed time result is obtained by PSO of 7.51s. Classification results before and after optimization of the decision tree algorithm have improved performance when using GridSearch. performance is different. It can be It can be concluded from the PSO and IWPSO comparison that this optimization is very useful in the case of heart disease dataset classification but with different advantages. At In previous research, the same research has been conducted on the classification of heart disease using the C4.5 decision tree. The results obtained by the previous researcher showed a C4.5 decision tree performance of 80.43%. In this research there has been no application of optimization, so this research uses a similar dataset. This study uses the same dataset with different variants of the decision tree algorithm and applies optimization in the form of PSO and IWPSO hyperparameters to the model. With the best results, namely IWPSO of 84.23%.

REFERENCES

Agus Oka Gunawan, I. M., Indah Saraswati, I. D. A., Riswana Agung, I. D. G., & Eka Putra, I. P. (2023). Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(2), 73–83. https://doi.org/10.47233/jteksis.v5i2.775

Al-Taie, R. R. K., Saleh, B. J., Saedi, A. Y. F., & Salman, L. A. (2021). Analysis of WEKA data mining algorithms Bayes net, random forest, MLP and SMO for heart disease prediction system: A case study in Iraq. *International Journal of Electrical and Computer Engineering*, *11*(6), 5229–5239. https://doi.org/10.11591/ijece.v11i6.pp5229-5239

Alfaris, L., Siagian, R. C., Muhammad, A. C., Nyuswantoro, U. I., Laeiq, N., & Mobo, F. D. (2023). Classification of Spiral and Non-Spiral Galaxies using Decision Tree Analysis and Random Forest Model: A Study on the Zoo Galaxy Dataset. *Scientific Journal of Informatics*, *10*(2), 139–150. https://doi.org/10.15294/sji.v10i2.44027

Aniamarta, T., Salsabilla Huda, A., & Lizariani Aqsha, F. (2022). Causes and Treatments of Heart Attack. *Biologica Samudra*, *4*(1), 22–31. https://doi.org/10.33059/jbs.v4i1.3925

Aziz, F., & Lawi, A. (2022). Increasing electrical grid stability classification performance using ensemble bagging of C4.5 and classification and regression trees. *International Journal of Electrical and Computer Engineering*, *12*(3), 2955–2962. https://doi.org/10.11591/ijece.v12i3.pp2955-2962

Hashim, N., Ismail, N. F. N., Johari, D., Musirin, I., & Rahman, A. A. (2022). Optimal population size of particle swarm optimization for photovoltaic systems under partial shading condition. *International Journal of Electrical and Computer Engineering*, *12*(5), 4599–4613. https://doi.org/10.11591/ijece.v12i5.pp4599-4613

Hussein, A. A. (2018). Improve The Performance of K-means by using Genetic Algorithm for Classification Heart Attack. *International Journal of Electrical and Computer Engineering (IJECE)*, *8*(2), 1256. https://doi.org/10.11591/ijece.v8i2.pp1256-1261

Iqbal, M., Herliawan, I., Ridwansyah, & Gata, W. (2020). Implementation of Particle Swarm Optimization Based Machine Learning Algorithm for Student Performance Prediction. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *6*(2), 195–204. https://doi.org/10.33480/jitk.v6i2.1695.IMPLEMENTATION

Juliane, C., & Technology, I. (2023). *Comparison Of The C . 45 And Naive Bayes Algorithms To Predict Diabetes*. *8*(4), 2641–2650.

Korzhakin, D. A., & Sugiharti, E. (2021). Implementation of Genetic Algorithm and Adaptive Neuro Fuzzy Inference System in Predicting Survival of Patients with Heart Failure. *Scientific Journal of Informatics*, *8*(2), 251–257. https://doi.org/10.15294/sji.v8i2.32803

Kristiyanti, D. A., & Normah, N. (2019). Optimising the Particle Swam Optimazion Usage for Predicting Indonesia Presidential Election Result Period 2019-2024. *SinkrOn*, *4*(1), 32. https://doi.org/10.33395/sinkron.v4i1.10149

Murinto, M., & Rosyda, M. (2022). Logarithm Decreasing Inertia Weight Particle Swarm Optimization

*name of corresponding author

Algorithms for Convolutional Neural Network. *JUITA: Jurnal Informatika*, *10*(1), 99. https://doi.org/10.30595/juita.v10i1.12573

Phan, T. M., Ha, P. T., Duong, T. L., & Nguyen, T. T. (2020). Improved particle swarm optimization algorithms for economic load dispatch considering electric market. *International Journal of Electrical and Computer Engineering*, *10*(4), 3918–3926. https://doi.org/10.11591/ijece.v10i4.pp3918-3926

Purwaningsih, E. (2019). Application of the Support Vector Machine and Neural Network Model Based on Particle Swarm Optimization for Breast Cancer Prediction. *SinkrOn*, *4*(1), 66. https://doi.org/10.33395/sinkron.v4i1.10195

Riansyah, M., Suwilo, S., & Zarlis, M. (2023). Improved Accuracy In Data Mining Decision Tree Classification Using Adaptive Boosting (Adaboost). *SinkrOn*, *8*(2), 617–622. https://doi.org/10.33395/sinkron.v8i2.12055

Santoso, H., & Musdholifah, A. (2019). Case Base Reasoning (CBR) and Density Based Spatial Clustering Application with Noise (DBSCAN)-based Indexing in Medical Expert Systems. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, *5*(2), 169–178. https://doi.org/10.23917/khif.v5i2.8323

Sekyere, Y. O. M., Effah, F. B., & Okyere, P. Y. (2023). Hyperbolic Tangent - Based Adaptive Inertia Weight Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro*, *2*. https://doi.org/10.25077/jnte.v12n2.1095.2023

Setiawan, A., Santoso, L. W., & Adipranata, R. (2019). Penerapan Algoritma Particle Swarm Optimization (PSO) untuk Optimisasi Pembangunan Negara dalam Turn Based Strategy Game. *Jurnal Infra*, *7*(1), 249–255.

Widiyati, D. K., Wati, M., & Pakpahan, H. S. (2018). Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *2*(2), 125. https://doi.org/10.30872/jurti.v2i2.1864

*name of corresponding author