IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.3, July 2023, pp. 281~290
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.86291

Application of Extreme Learning Machine Method With Particle Swarm Optimization to Classify of Heart Disease

Adela Putri Ariyanti*[1], Muhammad Itqan Mazdadi[2], Andi Farmadi[3], Muliadi[4], Rudy Herteno[5]
[1,2,3,4,5]Department of Science and Mathematics, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
e-mail: *[1]**adelaputria.yanti@gmail.com**, [2]mazdadi@ulm.ac.id, [3]Andifarmadi@gmail.com,[4]muliadi@ulm.ac.id, [5]rudy.herteno@ulm.ac.id

Abstrak

Penyakit jantung koroner adalah kondisi aliran darah ke otot jantung tersumbat, sehingga terjadi serangan jantung yang disertai nyeri dada atau stroke. Penyakit jantung adalah penyebab utama kematian di seluruh dunia, dengan lebih dari 17 juta kematian dilaporkan oleh Organisasi Kesehatan Dunia. Berbagai faktor risiko berkontribusi terhadap penyakit jantung, termasuk merokok, gaya hidup tidak sehat, kolesterol tinggi, dan hipertensi. Untuk mencegah meningkatnya angka kematian akibat penyakit jantung, prediksi penyakit dapat dilakukan untuk mengidentifikasi individu yang berisiko. Penambangan data, khususnya metode Extreme Learning Machine (ELM), biasanya digunakan untuk tujuan ini. ELM merupakan metode neural network pada kecepatan pelatihan dan tidak membutuhkan backpropagation, menentukan jumlah node tersembunyi yang optimal dan mencapai hasil yang akurat tetap menjadi tantangan. Pada penelitian ini, ELM dengan Particle Swarm Optimization (PSO) diusulkan untuk mengoptimalkan klasifikasi penyakit jantung, yang bertujuan untuk mencapai hasil yang optimal dengan pembelajaran cepat. Hasil dan pembahasan menyajikan keefektifan metode yang diusulkan dengan mengevaluasi akurasi klasifikasi berdasarkan berbagai parameter, seperti ukuran populasi, jumlah node tersembunyi, dan iterasi. Temuan menunjukkan bahwa ELM dengan optimasi PSO dapat memberikan hasil klasifikasi yang akurat untuk diagnosis penyakit jantung, dengan tingkat akurasi yang menjanjikan.

Kata kunci*—Penyakit Jantung, Extreme Learning Machine, Particle Swarm Optimization*

Abstract

Coronary heart disease is the condition where the heart's blood supply is blocked. Heart disease is the leading cause of death worldwide. Various risk factors contribute to heart disease, including smoking, unhealthy lifestyle, high cholesterol, and hypertension. Thus, disease predictions can be made to identify individuals at risk in order to prevent increased deaths from heart disease. Data mining, particularly the Extreme Machine Learning (ELM) method, is typically used for this purpose. ELM is a neural network method in training speed and does not require backpropagation, and determining the optimal number of hidden nodes and achieving accurate results remains a challenge. In this research, ELM with Particle Swarm Optimization (PSO) was proposed to optimize the classification of heart disease, which aimed to achieve optimal results with rapid learning. The research followed systematic processes, including data collection, preprocessing, modeling, and evaluation using confusion matrix analysis. The results and discussion presented the effectiveness of the proposed method by evaluating the accuracy of classification based on a variety of parameters, such as population size, number of hidden nodes, and iteration. Results suggested that ELM with PSO optimization can provide accurate classification results for the diagnosis of heart disease, with promising levels of accuracy.

Keywords*—Heart Disease, Extreme Learning Machine, Particle Swarm Optimization*

Received June 28th,2023; Revised July 26th, 2023; Accepted July 29th, 2023

1. INTRODUCTION

Coronary heart disease, or cardiovascular disease, occurs when blood flow to the heart muscle is narrowed or blocked by blood vessels that cause a heart attack, accompanied by chest pain or stroke. Heart disease is a disease that causes many deaths around the world. According to the World Health Organization (WHO), more than 17 million people have died from heart disease. Smoking, unhealthy lifestyle, high cholesterol, and hypertension are just a few risk factors that contribute to heart disease [1][2][3].

In an effort to prevent an increase in the number of deaths from heart disease, it is possible to predict heart disease in humans. There are many technologies that can be used to manage data that can help determine whether a person has a risk of heart disease or not. Data mining is one of the most widely used technologies. Data mining is the process of gathering important information from large amounts of data. Data Mining is used for the process of collecting important information related to informatics; data mining is also used in the health section for the prediction of diseases, one of which uses classification methods. Classification is a grouping of data in which the data has a label class or a target. There are several methods of classification, including Extreme Machine learning (ELM). This method is found in the nerve network [4] [5] [6].

Extreme Learning Machine (ELM) is a feedforward artificial neural network which has one hidden layer called Single Hidden Layer Feedforward Neural Networks (SLFNs). Feedforward aims to recognize patterns in data so that the identification process has accurate results. Feedforward has the best weight search process by searching from an activation function. In the ELM algorithm the hidden node learning parameters including input weights and input bias can be set randomly, and for network output weights can be determined analytically with a simple general inverse operation[7] Based on research conducted by [8], Extreme Learning Machine is a method of simulated neural networks that has the advantage that its training process does not require reverse propagation, so the training process can be faster because it does not need to update the weight and bias values of each epoch. ELM methods can be used to detect the presence of brain tumors by performing classifications on MRI imaging data. There are two types of tumors: tumors and non-tumors. Previous research was conducted [8] [9] on the comparison of ELM and K-NN algorithms. Test results of both algorithms for the classification of heart disease using the ELM algorithm resulted in the best performance with an average accuracy value of 93.33%, while the K-NN algorithm obtained a lower average accuracy value of 83,52%. This is because the performance of the K-NN algorithm is heavily influenced by the value of k. High value of k results in the K-NN not performing at its maximum. Therefore, the application of ELM methods is expected to produce more effective predictions for the diagnosis of heart disease.

Another study using ELM was carried out by [11] on the classification of tuberculosis diseases (TB). ELM yielded an accuracy of 99.33% using optimal parameters for the tests already obtained. The optimum number of hidden neurons was 20, the binary sigmoid activation function was the optimum function in this research, and the percentage of training and testing data was optimum at 70%:30%. In addition, research conducted by [12] reported that in addition to having the advantages of ELM, it also had a weakness, namely that the number of hidden nodes was determined by trial and error, so it was not possible to know the correct number of hidden nodes to get the correct results using the ELM method. In addition to the problem of hidden nodes in making the selection of input weight, if the bias in the ELM method values is chosen randomly, then this can result in the calculation being less than maximum.

A research was conducted by [13] on the assessment of apartments using the SVM method based on PSO. The research resulted in an accuracy of 82.04%, while using SVM alone produced an accuracy of 79.20%. These results show that the Particle Swarm optimization-based Vector Machine support method is a fairly good method for data classification.

Therefore, based on the abovementioned problem, the current research was conducted on the classification of heart disease utilizing the Extreme Learning Machine (ELM) approach with Particle Swarm Optimization (PSO) optimization with rapid learning to result in optimal outcomes.

2. METHODS

The flow of the research is depicted in Figure 1.

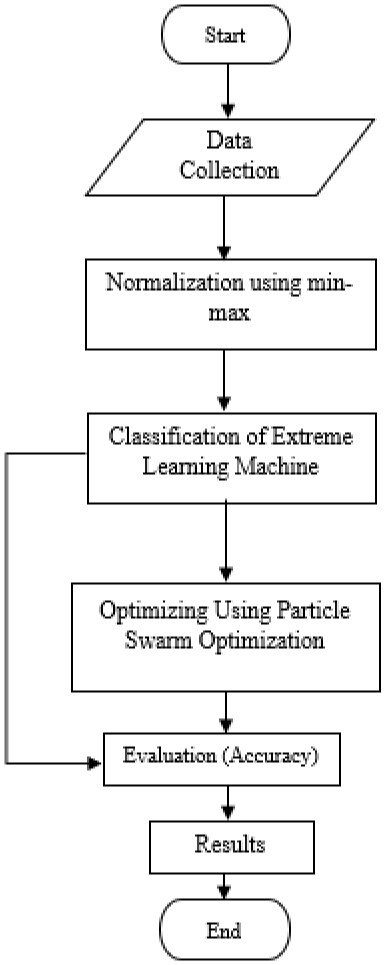

Figure 1 The Research Flow

2.1 Data Collection

This research used heart disease data obtained from the publicly available Kaggle Datasets website. The data used in this research had 14 attributes and consisted of 303 data. The data used had two values: 0 for non-heart disease conditions and 1 for heart disease conditions. This is shown in Table 1.

Table 1 Attributes of the heart disease dataset

Variable	Attribute	Description
X1	*Age*	Age in years

X2	*Sex*	1 = Male ; 0 = Female
X3	*Cp*	*Chest Pain Type* : *Value* 1*: typical angina* *Value* 2: *atypical angina* *Value* 3: *non-anginal pain* *Value* 4: *asymptomatic*
X4	*Trestbps*	Resting blood pressure (in mm Hg)
X5	*Chol*	Serum Cholesterol in mg / dl
X6	*Fbs*	Fasting blood sugar > 120mg / dl 1 = True 0 = False
X7	*Restecg*	Resting electrocardiographic results: - 0 : normal - 1 : Having ST-T wave abnormality - 2: Showing probable or define left ventricular hypertrophy by Estes'criteria
X8	*Thalach*	Maximum heart rate achived
X9	*Exang*	Exercise induced angina : 1 = yes 0 = no
X10	*Oldpeak*	Depression induced by exercise relative to rest
X11	*Slope*	The slope of the peak exercise segment: Nilai 1 : up sloping Nilai 2 : flat Nilai 3 : down sloping
X12	*Ca*	Number of major vessels colored by dluoroscopy that ranged between 0 and 3
X13	*Thal*	3 = normal 6 = fixed defect 7= reversible defect
Y	*num*	Diagnosis classes: 0 = healthy 1 = Patient who is subject to possible heart disease

2.2 Preprocessing

This phase used data normalization using the Min-Max Normalization method. Min-Max Normalization is a method of normalization that involves performing a linear transformation of the original data to produce a balance of comparative values between the data before and after the process. This method can be done using the formula in Equation 1.

$$x\ normal = \frac{\mathrm{x - x\,min}}{\mathrm{x\,max - x\,min}} \tag{1}$$

2.3 Designing the Model

The dataset was divided into two parts using the split data technique, consisting of 70% training data and 30% test data, after going through the preprocessing stage. This is based on a previous research, namely the classification of breast cancer using the Naïve Bayes algorithm. Other research conducted by [11] on the classification of tuberculosis diseases using the Extreme Machine Learning Method produced the highest average accuracy at 70% and 30% percentages. This research was divided into two classification models: Extreme Learning Machine classification and Extreme Machine Learning classification with Particle Swarm optimization.

2.4 Evaluation

Evaluation is a stage to measure the performance of the data mining modelling. In the evaluation stage, accuracy was determined using confusion matrix. This is shown in Table 2 and Equation 2.

Table 2 Confusion Matrix

Actual Class	**Predicted Class**	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

The classification accuracy can be calculated from the confusion matrix as the sum of correct cells in the table (true positives and true negatives) divided by all cells in the table.

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{2}$$

3. RESULTS AND DISCUSSION

3.1 Data Collection

The data collection process was done through Kaggle Datasets that are publicly available. The data consisted of 303 records, consisting of 165 records of ill class label attributes and 138 records of healthy class label attributes.

3.2 Preprocessing

The stage of data preprocessing in this research was to transform smaller distance values (0-1) using the min-max normalization method. Normalization was used for data standardization so that both input range and output range are in the same range between 0 and 1. This is shown in Figures 2 and 3.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

Figure 2 Data Before the Normalization

```
          age  sex        cp  trestbps      chol  fbs  restecg   thalach  \
0    0.708333  1.0  1.000000  0.481132  0.244292  1.0      0.0  0.603053
1    0.166667  1.0  0.666667  0.339623  0.283105  0.0      0.5  0.885496
2    0.250000  0.0  0.333333  0.339623  0.178082  0.0      0.0  0.770992
3    0.562500  1.0  0.333333  0.245283  0.251142  0.0      0.5  0.816794
4    0.583333  0.0  0.000000  0.245283  0.520548  0.0      0.5  0.702290
..        ...  ...       ...       ...       ...  ...      ...       ...
298  0.583333  0.0  0.000000  0.433962  0.262557  0.0      0.5  0.396947
299  0.333333  1.0  1.000000  0.150943  0.315068  0.0      0.5  0.465649
300  0.812500  1.0  0.000000  0.471698  0.152968  1.0      0.5  0.534351
301  0.583333  1.0  0.000000  0.339623  0.011416  0.0      0.5  0.335878
302  0.583333  0.0  0.333333  0.339623  0.251142  0.0      0.0  0.786260

     exang   oldpeak  slope    ca      thal
0      0.0  0.370968    0.0  0.00  0.333333
1      0.0  0.564516    0.0  0.00  0.666667
2      0.0  0.225806    1.0  0.00  0.666667
3      0.0  0.129032    1.0  0.00  0.666667
4      1.0  0.096774    1.0  0.00  0.666667
..     ...       ...    ...   ...       ...
298    1.0  0.032258    0.5  0.00  1.000000
299    0.0  0.193548    0.5  0.00  1.000000
300    0.0  0.548387    0.5  0.50  1.000000
301    1.0  0.193548    0.5  0.25  1.000000
302    0.0  0.000000    0.5  0.25  0.666667
```

Figure 3 Data After the Normalization

3.3 Testing of Populations

The test aimed to identify the number of populations that can provide optimal evaluation values for cases of diagnosis of heart disease using the ELM method optimized using PSO. In this test, an inertia weight (w) of 0.5 was used with c1 and c2 which had a value of 1, the number of iterations was 2, and the number of hidden nodes for ELM was 5. The division of training data and test data was 70%:30%. The size of population variations started at 100 and multiplies to 500. For each variation, the number of populations was tested five times, and then the average for each evaluation was obtained. The results of the test are provided in Table 3.

Table 3 Results of Population Testing

Size of Population	Accuracy of trial no.					Average
	1	2	3	4	5	
100	79,12%	83,52%	81,32%	82,42%	78,02%	80,88%
200	81,32%	82,42%	82,42%	81,32%	83,52%	82,20%
300	84,62%	85,71%	80,22%	81,32%	84,62%	83,30%
400	82,42%	83,52%	82,42%	81,32%	82,42%	82,42%
500	82,42%	81,32%	82,42%	81,32%	84,62%	82,42%

Based on Table 3 above, optimal evaluation resulted for the population of 300 with an average of 83.30%.

3.4 Testing of the Number of Hidden Node

This test aimed to find out how many hidden nodes could produce an optimal evaluation. This test employed the parameter values of the optimal evaluation results that have been performed previously, namely the Population test with a total population of 300, an inertia weight of 0,5, and constants 1 (c1) and 2 (c2) of 1. This test was performed five times with the number of hidden nodes being 3, 5, 7, 9, and 11, and then the average for each evaluation was obtained. The number of hidden node testing is given in Table 4.

Table 4 Results of Testing the Number of Hidden Node

Number of Hidden Nodes	Accuracy of trial no.					Average
	1	2	3	4	5	
3	79,12%	83,52%	83,52%	81,32%	82,42%	81,98%
5	83,52%	82.42%	85,71%	82,42%	80,22%	82,86%
7	85,71%	82,42%	85,71%	83,52%	82,42%	83,96%
9	86,81%	82,42%	83,52%	85,71%	82,42%	84,18%
11	83,52%	81,32%	82,42%	82,42%	84,62%	82,86%

Table 2 shows that the number of hidden nodes that resulted in the best evaluation results is 9 hidden nodes with an average accuracy of 84.18%.

3.5 Testing of the number of iterations

Testing based on the number of iterations aimed to obtain the number of iterations that can produce the best evaluation value in of heart disease cases. In this test, the number of iterations used was 3.5, 10, 15, and 20. In each variation, the number of iterations was performed five times. The number of hidden nodes was 9, the value of inertia weight was 0.5, and the constants 1 (c1) and 2 (c2) are 1. The number of iterations can be seen in Table 5.

Table 5 Results of Testing the Number of Iterations

Number of Iterations	Accuracy of trial no.					Average
	1	2	3	4	5	
3	82,42%	83,52%	82,42%	83,52%	81,32%	83,64%
5	83,52%	83,52%	85,71%	81,32%	83,52%	83,51%
10	82,42%	82,42%	84,62%	85,71%	81,32%	83,30%
15	83,52%	82,42%	82,42%	82,42%	82,42%	82,64%
20	81,32%	82,42%	82,42%	83,52%	82,42%	82,42%

The decrease in average fitness was consistent. It can be concluded from Table 5 that the average value of fitness is influenced by the number of iterations. The more the iterations, the lower the value of fitness. The best fitness score is of iteration 3 with an average accuracy of 83.64%.

3.6 ELM testing

This test aimed to find out the evaluation value of the application of the ELM method with PSO-ELM in conducting classifications in the diagnosis of heart disease. The ELM test without PSO used nine hidden nodes. The program was ran 5 times. Then, an average of the evaluation score was obtained. Table 6 shows the average of the evaluation results on the application of ELM for the diagnosis of heart disease.

Table 6 Results of the ELM Test

ELM	Accuracy of trial no.					Average
	1	2	3	4	5	
	67,00%	54,90%	53,80%	54,90%	56,00%	57,32%

From Table 6, it can be seen that the average evaluation score is 57,32%.

3.7 ELM testing with PSO

This test aimed to find out the best evaluation value in cases of heart disease using an ELM optimized with PSO. This test used the parameter value of the optimal evaluation result that has been performed previously: the number of hidden nodes of 9 pieces with the number of iterations of 3. The inertia weight was 0,5, constants 1 (c1) and 2 (c2) are 1. This test was also done five times, and the average of the evaluation value was obtained. The ELM testing with PSO can be seen in Table 7.

Table 7 Results of ELM test with PSO

ELM with PSO	Accuracy of trial no.					Average
	1	2	3	4	5	
	83,52%	83,52%	84,62%	84,62%	82,42%	83,74%

From Table 7, it can be seen that the result of ELM testing using the optimal parameters was established. The average accuracy value is 83.74%.

4. CONCLUSIONS

Based on the research that has been conducted using Extreme Learning Machine and Particle Swarm Optimization for the classification of heart disease, it can be concluded that the evaluation results on the diagnosis process of coronary heart disease using the ELM method produced an accuracy value of 57,32%, while by performing the optimization of ELM parameters using the PSO method, the best parameter obtained from the test results was the number of hidden nodes of 9, the total population of 300, the best number of iterations of 3, the inertia weight of 0.5, and the constants 1 (c1) and 2 (c2) of 1. The accuracy of the parameter was 83.74%. It can be concluded that the PSO method used to obtain ELM parameters such as input weight and bias influenced the results of the evaluation on the case of classification for the diagnosis of heart disease using the ELM method. The addition of the PSO algorithm to the ELM classification aimed to obtain better classification results than the conventional ELM. This has been demonstrated in this research with conventional ELM accuracy results that are much smaller than the accuracy of ELM & PSO. The accuracy obtained from the conventional ELM algorithm was only 57.32%. For further development of the research, there are some suggestions from the author, among others, to use algorithms or other optimization methods to obtain optimal ELM parameters so that it can increase the evaluation value of the classification for the diagnosis of coronary heart disease. Oher methods of classification to find out which classification methods can provide better evaluation values in the case of a diagnosis of heart disease can also be used.

ACKNOWLEDGEMENTS

The author would like to thank the supervisors for their guidance during the accomplishment of this present research. The author also thank Universitas Lambung Mangkurat who supported her to conduct this research.

REFERENCES

[1] A. N. Sari and S. Alfionita, “Klasifikasi Penyakit Jantung Menggunakan Metode Naïve Bayes,” *AMRI (Analisa Metod. Rekayasa Inform.*, vol. 1, no. 1, pp. 22–26, 2022, doi: 10.12487/AMRI.v1i1.xxxxx.

[2] A. B. Wibisono and A. Fahrurozi, “Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner,” *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.

[3] A. A. Syafitri Hidayatul AA, Yuita Arum S, “Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 2546–2554, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[4] M. K. Nasution, R. R. Saedudin, and V. P. Widartha, “Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9765–9772, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15759

[5] B. Gbadamosi, R. O. Ogundokun, E. A. Adeniyi, S. Misra, and N. F. Stephens, “Medical Data Analysis for IoT-Based Datasets in the Cloud Using Naïve Bayes Classifier for Prediction of Heart Disease,” *Internet of Things*, no. September, pp. 365–386, 2022, doi: 10.1007/978-3-031-05528-7_14.

[6] N. A. Sugianto, I. Cholissodin, and A. W. Widodo, “Klasifikasi Keminatan Menggunakan Algoritme Extreme Learning Machine dan Particle Swarm Optimization untuk Seleksi Fitur (Studi Kasus: Program Studi Teknik Informatika FISugianto, N. A., Cholissodin, I., & Widodo, A. W. (2018). Klasifikasi Keminatan Mengg,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 5, pp. 1856–1865, 2018.

[7] V. U. M. Maksum, *Klasifikasi Data Citra X-Ray Covid-19 Meggunakan Metode Glcm Dan Extreme Learning Machine (Elm)*. 2021.

[8] R. R. Wahid, F. T. Anggraeny, and B. Nugroho, “Implementasi Metode Extreme Learning Machine untuk Klasifikasi Tumor Otak pada Citra Magnetic Resonance Imaging,” *Pros. Semin. Nas. Inform. Bela Negara*, vol. 1, pp. 16–20, 2020, doi: 10.33005/santika.v1i0.45.

[9] I. Larasati, “Analisa Perbandingan Data Mining Pada Klasifikasi Penyakit Jantung Menggunakan Algoritma Extreme Learning Machine (Elm) Dan K-Nearest Neighbor (K-NN),” p. 1, 2021.

[10] A. A. Altae and A. Ehsani Rad, “Diagnosing heart disease by a novel hybrid method: Effective learning approach,” *Informatics Med. Unlocked*, vol. 40, no. March, p. 101275, 2023, doi: 10.1016/j.imu.2023.101275.

[11] V. V. Nurdiansyah, I. Cholissodin, and P. P. Adikara, “Klasifikasi Penyakit Tuberkulosis (TB) menggunakan Metode Extreme Learning Machine (ELM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 5, pp. 1387–1393, 2020, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7237

[12] F. A. B. Darmayanti Eka Yuni, Budi Dharma Setiawan, “Particle Swarm Optimization Untuk Optimasi Bobot Extreme Learning Machine Dalam Memprediksi Produksi Gula Kristal Putih Pabrik Gula,” vol. 2, no. 11, pp. 5096–5104, 2018.

[13] L. Nilawati and Y. E. Achyani, “Optimasi Metode Particle Swarm Optimization (PSO) Pada Prediksi Penilaian Apartemen,” vol. 21, no. 2, pp. 227–234, 2019, doi: 10.31294/p.v20i2.