ISSN: 2338-3070, DOI: 10.26555/jiteki.v8i2.xxxx

The Combination of C4.5 with Particle Swarm Optimization in Classification of Class for Mental Retardation Students

Sausan Hidayah Nova, Budi Warsito, Aris Puji Widodo
Magister of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang 50241, Indonesia

ARTICLE INFO

Article history:

Received December 30, 2022
Revised February 14, 2023
Published February 24, 2023

Keywords:

Mental Retardation;
Classification;
C4.5;
Particle Swarm Optimization;
Feature Selection

ABSTRACT

Mental retardation or brain weakness is a condition of children who experience mental disorders. There are several characteristics to know the child has mental retardation. When entering a school, teachers are expected to be able to determine the right class for mental retardation students according to their category. Data mining is the process of finding patterns in selected data using artificial intelligence and machine learning. Algorithm C4.5 is one of the classification techniques in data mining. C4.5 can be used to create decision trees and classify data that has numeric, continuous, and categorical attributes. But C4.5 has the disadvantage of reading large amounts of data and cannot rank every alternative. PSO is an optimization algorithm for feature selection that can improve performance in data classification. Therefore, this study proposes an algorithm that can overcome the weaknesses of C4.5 by combining PSO. This study aims to classify a class of new mental retardation students using a combination of C4.5 as a classification and PSO as a feature selection to determine the attributes that affect the level of accuracy. The contribution of this research is to make it easier for the school to determine the new class of mental retardation students so that it is appropriate and according to their needs. The classification process in this study uses a combination of C4.5 and PSO. The validation used in this model is 10-fold cross-validation, and the evaluation uses a confusion matrix. This study resulted in an accuracy of C4.5 before using PSO of 91%. While the accuracy of C4.5 uses a PSO of 93%. Of the 20 attributes, there are 6 attributes that affect the level of accuracy. This study shows that PSO can be used to implement feature selection and increase the accuracy value of C4.5 by 2%.

Corresponding Author:

Sausan Hidayah Nova, Master of Information Systems, School of Postgraduate Studies, Diponegoro University, Semarang 50241, Indonesia
Email: novasausan@gmail.com

1. INTRODUCTION

Education is an attempt to help the development of children [1]–[3]. Every human being has the right to get an education, including children with special needs, one of which is mental retardation children [4]–[6]. Mental retardation or what is often referred to as brain weakness, mental retardation, and idiot, is a condition of children who have mental disorders [7]. Mental retardation children have intelligence below the normal average [3]–[6]. Mental retardation children experience difficulties in learning and social adjustment to their surroundings because they cannot achieve full development [8]–[10]. Mental retardation children have deficiencies in terms of adaptive skills, such as the ability to communicate, self-help, social skills, self-direction, self-safety, and academics [4].

One form of special education for children with special needs is special education [5]. This extraordinary educational institution is called an extraordinary school (Sekolah Luar Biasa) [3] where mental retardation children can develop their potential and receive guidance from teachers who understand mental retardation, such as learning about communication, socialization, self-help, and life skills [1], [5]. Because the characteristics of mental retardation children are varied and numerous, a classification method is needed to help and facilitate the school in accepting new mental retardation students according to predetermined criteria quickly and precisely [4], [11].

Data mining is the process of looking for patterns or interesting information in selected data using statistical, mathematical, artificial intelligence, and machine learning techniques [12][13] and using certain methods to extract and identify useful information and knowledge from large databases [14][15]. One of the tasks that can be performed using data mining is classification [16][17]. The purpose of classification is to predict the trend of data that will appear in the future [18]. There are several classification methods such as decision tree classifier, naïve Bayes, neural network, KNN, and others [12][17]. One of the data mining classification techniques is the C4.5 algorithm. The C4.5 algorithm can be used to create decision trees and is an algorithm specifically for supervised learning [19]. C4.5 algorithm can be used to classify data that has numeric, continuous, and categorical attributes [20][21]. The application in the health sector is classifying children with special needs based on the characteristics that exist in children with special needs [14] and classifying self-care problems correctly for the diagnosis of children with disabilities [22]. The application in the education sector is the classification of the student's study period [23].

The C4.5 algorithm has a weakness in reading large amounts of data and cannot rank each alternative [24]–[26], so it is necessary to optimize the algorithm for feature or attribute selection to improve performance in data classification [27]. The purpose of feature selection is to increase the number of features by eliminating redundant features or by selecting the most useful features to increase classification simplification. One of the optimizations used to improve the accuracy of the C4.5 algorithm is Particle Swarm Optimization (PSO) [20], [28]. Each particle in the PSO tends to fly towards a better search area during the search process. C4.5 will be combined with the PSO algorithm which is used as a more effective attribute selection because it uses several parameters so that computation time becomes fast and is expected to produce better accuracy values [29], [30]. One of its applications is in classifying under-five nutrition, the PSO algorithm is used to select attributes from the under-five nutritional status data combined with the C4.5 algorithm to find optimal parameters [30].

This study aims to classify classes for new mental retardation students using a combination of C4.5 as a classification and PSO as feature selection and to find out what attributes affect identifying new mental retardation students. This study will display the final results of class division from the calculation of the combined algorithm, namely class 1C for mild mental retardation students and class 1C1 for moderately mental retardation students. The data used in this research is new student identification data obtained from SLB Negeri 1 Pelaihari. The contribution of this research is to make it easier for the school to determine a new class of mental retardation students by using identification that has been done before so that it is appropriate and according to needs.

2. METHODS

The method used in this study is described using a flowchart. This research has several stages, starting from identifying the problem, data collection, data analysis, design and implementation, and testing. These stages are shown in Fig. 1.

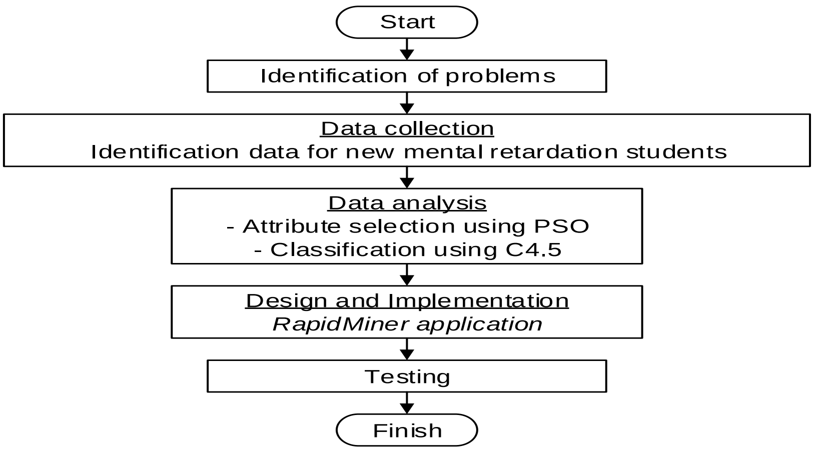

Fig. 1. Research Flowchart

2.1. Identification of Problems

The initial stage of this research is to identify the problem. In this study, we will find out whether the combination of the C4.5 algorithm and Particle Swarm Optimization (PSO) can solve the problem being researched to determine the class for new mental retardation students.

2.2. Data Collection

The data used in this research is secondary data. Secondary data is data that is already available so that finding and collecting data can be obtained more easily and quickly [31][32]. The data used was obtained from SLB Negeri 1 Pelaihari, namely identification data for new mental retardation students totaling 100 data and consisting of 20 categorical type attributes. This data collection conforms to the identification of the characteristics of prospective new mental retardation students. Table 1 contains data on the identification characteristics of mental retardation students.

Table 1. Characteristic Data for Mental Retardation Students

Attribute	Characteristics
A01	IQ 50-70.
A02	Two times in a row did not go to class.
A03	Still able to read, write, and do simple math.
A04	Low imagination.
A05	They are less able to control their feelings.
A06	They are easily influenced.
A07	Inability to think logically.
A08	Unable to think abstractly.
A09	Concentration is less or not for long, about less than 15 minutes.
A10	Personality was less harmonious.
A11	Have an IQ of 25 to less than 50.
A12	Almost no initiative.
A13	They are only able to read single sentences.
A14	They have difficulty in calculating simple calculations.
A15	Can't concentrate (less than 10 minutes) and gets bored quickly.
A16	Weak motor.
A17	Spoken language is weak or unclear or inhibited.
A18	Dirty and lack of understanding of cleanliness.
A19	They do not understand compassion, justice, and manners.
A20	They have clinical features of down syndrome.

2.3. Data Analysis

This research aims to determine the class for new students with mental retardation using PSO to select features at the beginning to get more optimal results and then classify using C4.5 to classify the target class. These stages are shown in Fig. 2.

Using PSO to calculate the transfer velocity [33][34] can be seen in (1), and the position of the particles in PSO can be seen in (2). Where $V_i(t)$ is the velocity of particle i during iteration t, $X_i(t)$ is the position of particle i during iteration t, X_{Pbesti} is the best position of particle i, X_{Gbest} is the global best position, c_1 and c_2 are learning rates for individual ability (cognitive) and social influence (group), then r_1 and r_2 are random numbers with values between 0 and 1. Fig. 3 shows the PSO stages in selecting attributes.

$$V_i(t) = V_i(t-1) + c_1 r_1\,[X_{Pbesti} - X_i(t)] + c_2 r_2\,[X_{Gbest} - X_i(t)] \tag{1}$$

$$X_i(t) = X_i(t-1) + v_i(t) \tag{2}$$

After the attributes are selected, then the classification process is carried out using C4.5 [36][37]. Equation (3) is used to calculate the number of class cases and the entropy value of all cases, where S is the case set, n is the number of partitions S, and pi is the proportion of S_i to S. Then, calculate the entropy value and gain value for each attribute. The highest gain value is the attribute that becomes the root of the decision tree to be made. Equation (4) is used to calculate the gain value, where A is the attribute, $|S_i|$ is the number of cases on the i-th partition, and $|S|$ is the number of cases in S. After the entropy and gain values is calculated for all attributes to get the highest gain value, the process can be repeated for each branch until all cases on the branch have the same class. Fig. 4 shows the classification using the C4.5 algorithm.

$$Entropy(S) = \sum_{i=0}^{n} - pi\ x\ log_2 pi \tag{3}$$

$$Gain(S,A) = Entropy(S) - \sum_{i=0}^{n} \frac{|S_i|}{|S|}\ x\ Entropy\ (S_i) \tag{4}$$

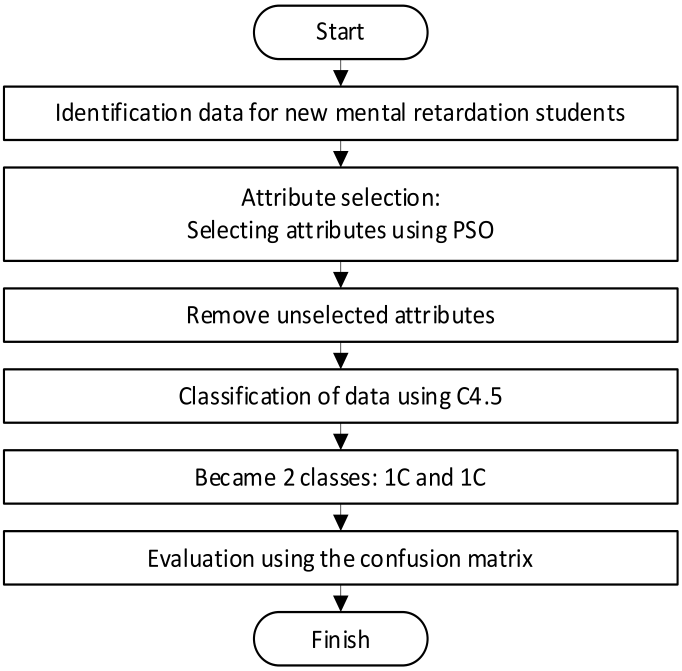

Fig. 2. Algorithm Used in Research Flowchart

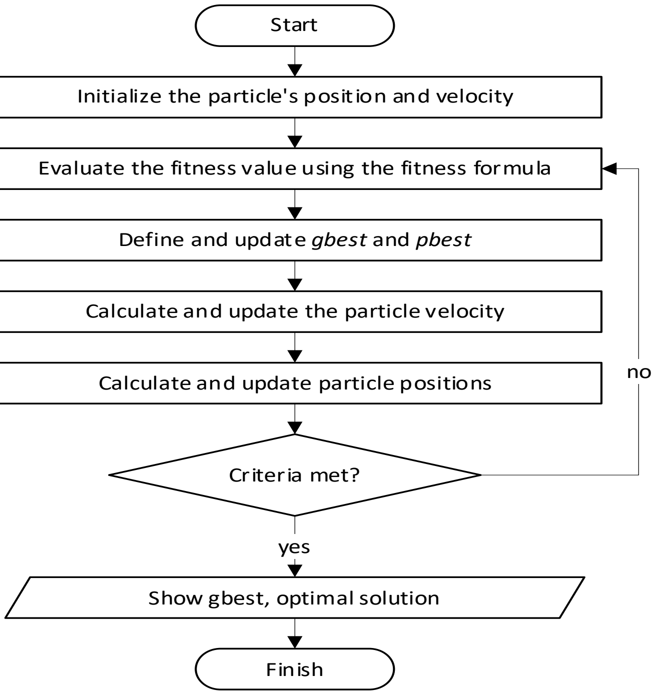

Fig. 3. PSO Attribute Selection Flowchart [35]

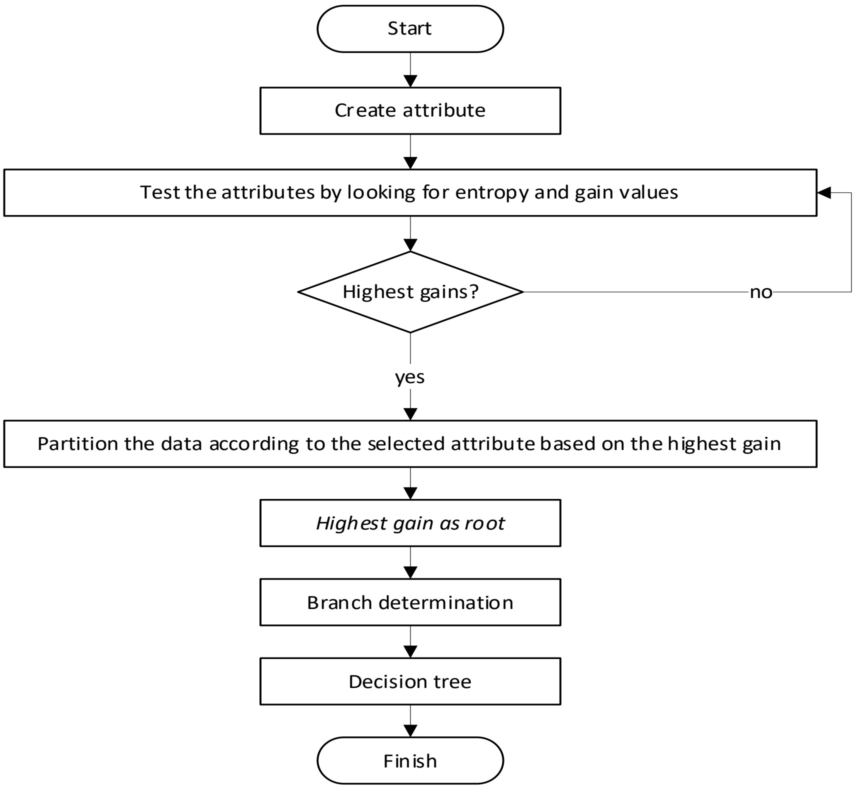

Fig. 4. Flowchart of Classification Using C4.5 [38]

2.4. Design and Implementation

At this stage, a model is implemented to determine classes for new mental retardation students using a combination of the C4.5 algorithm and the Particle Swarm Optimization (PSO) algorithm and will be implemented using RapidMiner tools.

2.5. Testing

At this stage, algorithm testing will be carried out. The data that has been collected is then classified using the C4.5 algorithm and the Particle Swarm Optimization (PSO) algorithm. Algorithm testing is carried out using the confusion matrix model to determine the level of classification accuracy of the data being tested.

3. RESULTS AND DISCUSSION

This research aims to build a class classification model for new mental retardation students using a combination of C4.5 and Particle Swarm Optimization (PSO) attribute selection using the RapidMiner tool.

3.1. Data Classification Using C4.5

The initial data classification in this study was carried out using the C4.5 algorithm. The "class" attribute is a label and the evaluation process uses k-fold cross-validation 10 times which will be separated into training data and testing data with sampling type using shuffled sampling [35][39]. The application of this classification uses several parameters because the use of parameters will affect the results of the accuracy of C4.5 [37][40]. The C4.5 parameters used are criterion, maximum depth, minimum leaf size, minimum size for split, and minimum gain. The parameters to be used can be seen in Table 2.

Table 2. Parameters C4.5

Parameters	Values
Criterion	Information gain
Maximal depth	5
Minimum leaf size	2
Minimum size for split	4
Minimum gain	0

After determining the parameters to be used, then data processing is carried out as shown in Fig. 5. Then, the performance results of the C4.5 algorithm can be seen in Tabel 3.

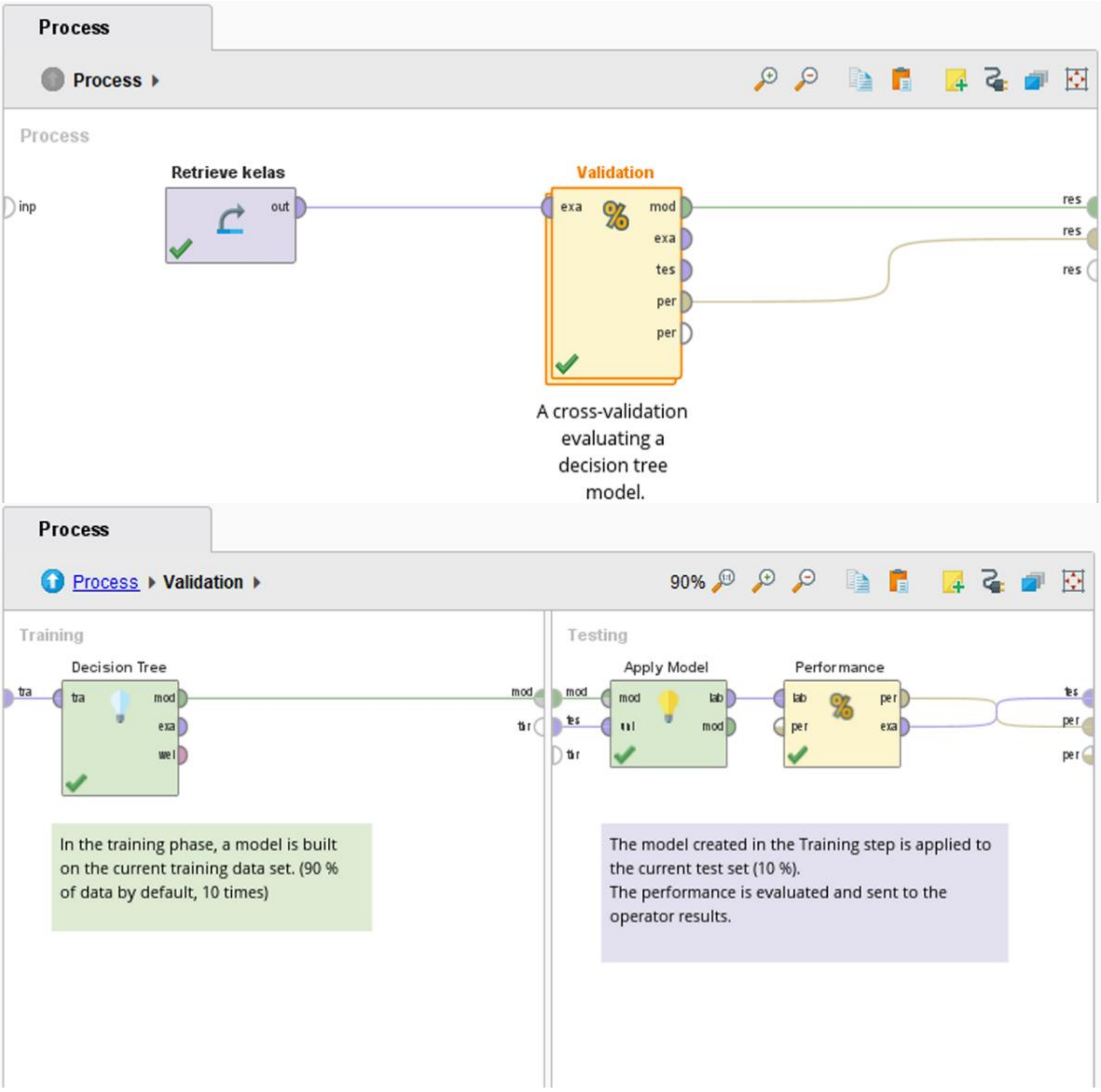

Fig. 5. C4.5 Algorithm Implementation Model

Table 3. Performance Results From The C4.5 Model

	True C	True C1	Class Precision
Pred. C	39	4	90.70%
Pred. C1	5	52	91.23%
Class Recall	88.64%	92.86%	
Accuracy : 91%			
Precision : 90.06%			
Recall : 92.14%			

From the application of the C4.5 model above and the parameters used, an accuracy of 91%, a precision of 90.06%, and a recall of 92.14% are obtained. Based on the data in Tabel 3, it is known that class C data is correct and predicted by class C as many as 39 data, class C data predicted by class C1 is 4 data, class C1 data is predicted by class C as many as 5 data, and class C1 data is correct and Class C1 predictions total 52 data.

After that, a decision tree is obtained from the application of the C4.5 algorithm model. The decision tree can be seen in Fig. 6.

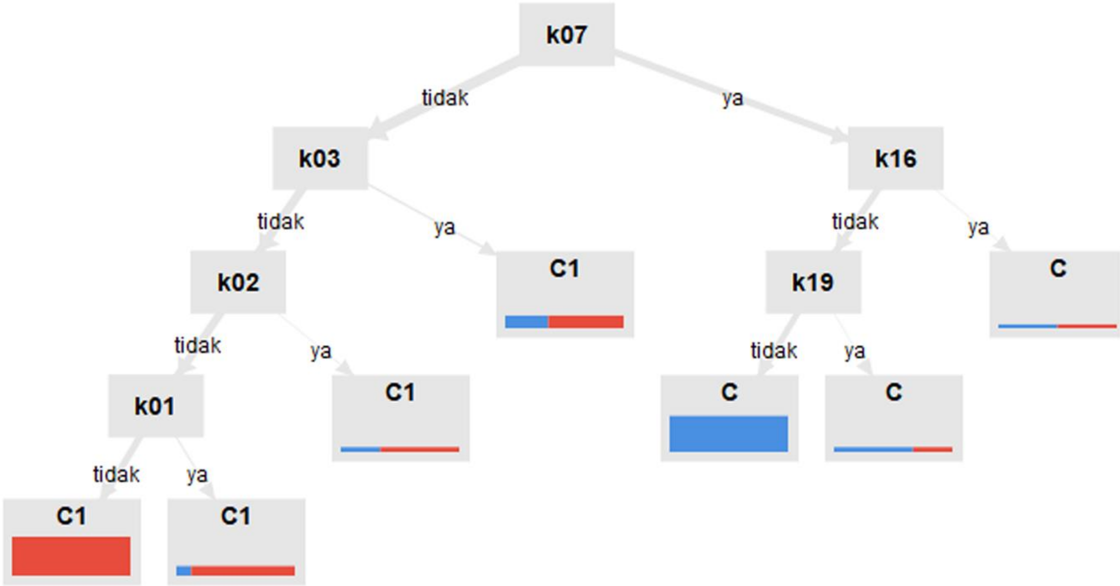

Fig. 6. Decision Tree from C4.5 Algorithm

3.2. Data Classification Using C4.5 with PSO

Data classification begins by using k-fold cross-validation 10 times as an evaluation process which will be separated into training data and testing data with sampling type using shuffled sampling. The C4.5 parameter used is the same as before, which can be seen in Table 2. The PSO parameters used are population size, the maximum number of generations, inertial weight, local best weight, and global best weight. The PSO parameters used can be seen in Table 4. After determining the parameters to be used, then data processing is carried out as shown in Fig. 7.

Table 4. Parameters PSO

Parameters	Values
Population size	5, 10, 15, 20, 25, 30, 35
Maximum number of generations	30, 35, 40, 45, 50, 55
Inertial weight	1
Local best weight	1
Global best weight	1

After implementing the C4.5 and PSO combination model, PSO was tested 9 times using predetermined parameters. Table 5 below shows that the performance of the C4.5 algorithm with PSO produces the best accuracy value of 93% when the population size is 30 and the maximum number of generations is 50 with an execution time of 19 seconds.

Table 5. Test Results Using Different Population Sizes and Maximum Number Of Generations

Test	PSO Parameters	Accuracy	Execution Time
1	Posize=5, generate=30	92%	2 second
2	Posize=10, generate=30	92%	4 second
3	Posize=15, generate=35	92.%	7 second
4	Posize=15, generate=40	92%	7 second
5	Posize=20, generate=40	93%	10 second
6	Posize=25, generate=45	92%	14 second
7	Posize=25, generate=50	92%	14 second
8	Posize=30, generate=50	93%	19 second
9	Posize=35, generate=55	92%	28 second

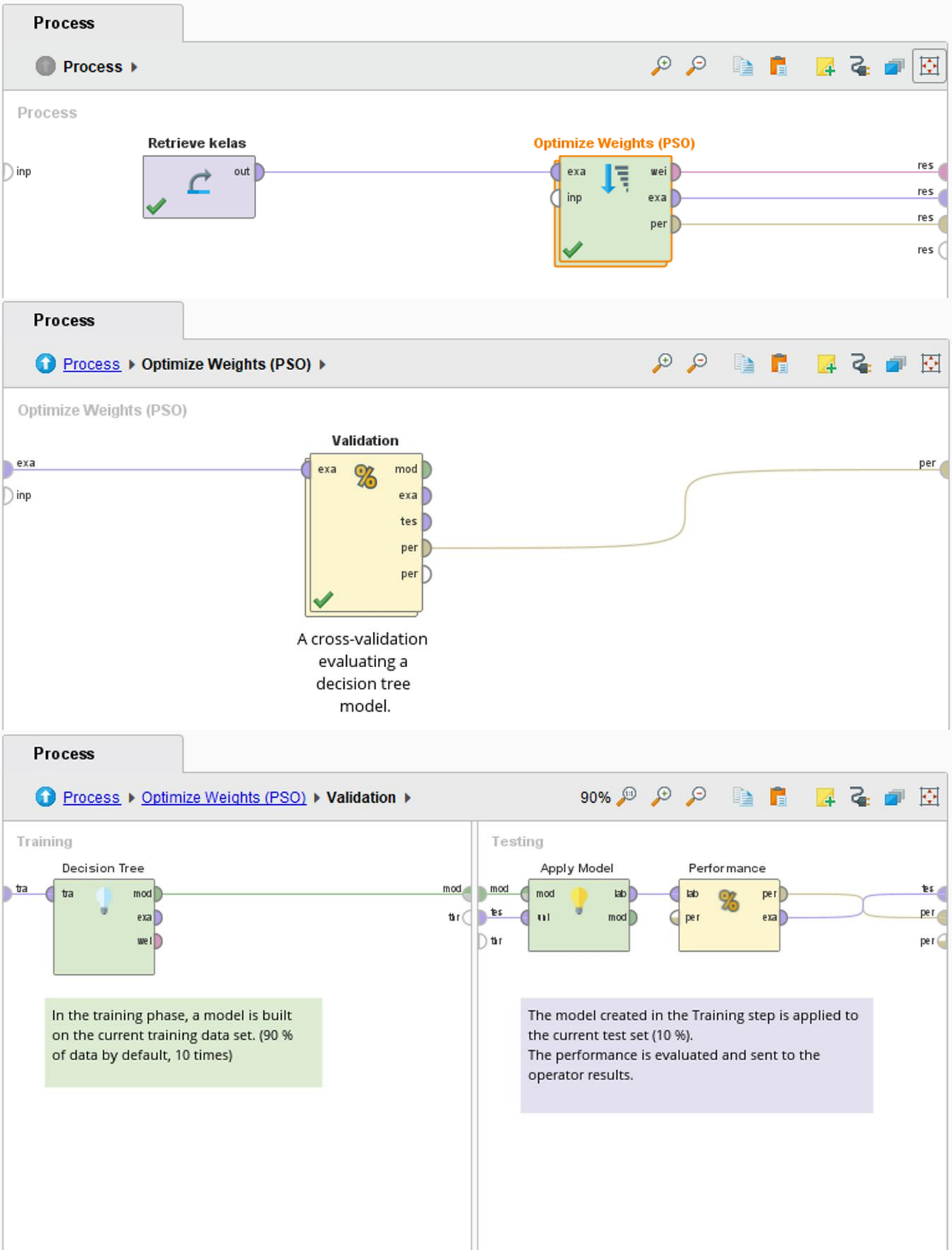

Fig. 7. Implementation Model Using C4.5 with PSO

And also produces a precision of 92,46% and a recall of 96%. Based on the data in Table 6, it is known that class C data is correct and predicted by class C is 39 data, class C data is predicted by class C1 is 2 data, class C1 data is predicted by class C is 5 data, and class C1 data is a correct and predictable class C1 amounted to 54 data. The performance results of the C4.5 model with PSO can be seen in Table 6.

Table 6. Performance Results From The C4.5 with PSO Model

	True C	True C1	Class Precision
Pred. C	39	2	95.12%
Pred. C1	5	54	91.53%
Class Recall	88.64%	96.43%	
Accuracy : 93%			
Precision : 92.46%			
Recall : 96%			

In this study, a feature selection algorithm is needed because of the 20 attributes used, it turns out that not all attributes can affect the resulting level of accuracy. The results of attribute weighting using PSO at the best accuracy can be seen in Table 7.

Table 7. PSO Feature Selection Weighting Results

Attributes	Weight
A01	0
A02	0
A03	0.050
A04	0
A05	0.754
A06	0.966
A07	1
A08	0
A09	1
A10	0.650
A11	0
A12	1
A13	1
A14	1
A15	0.505
A16	0
A17	1
A18	0.705
A19	0
A20	0.631

Based on the data in Table 7, there are 20 attributes used. Attributes that have a weight value of 1 are 6 attributes, there are A07, A09, A12, A13, A14, and A17. While the attributes that have a weight value of 0 are 7 attributes, there are A01, A02, A04, A08, A11, A16, and A19. An attribute with a value of 1 allows it to significantly influence the accuracy results, while an attribute with a value of 0 does not affect the accuracy results.

3.3. Analysis Evaluation and Model Validation

From the results of algorithm C4.5 with PSO, and algorithm C4.5, a comparison of classification performance is obtained in Table 8 below. Based on the data in Table 8 it is known that the classification using C4. 5 with PSO produces a better accuracy rate of 93% compared to classification using C4.5 without optimizing feature selection using PSO, namely 91%. As well as producing precision from 90.06% to 92.46%, and recall from 92.14% to 96%. Applying PSO to feature selection increases accuracy by 2%, precision by 2.4%, and recall by 3.86%.

Table 8. Performance Comparison

Performance	Accuracy	Precision	Recall
C4.5	91%	90.06%	92.14%
C4.5 with PSO	93%	92.46%	96%

By comparing previous research on classification using algorithm C4.5 [23] and self-care diagnosis for children with special needs in the health sector [22] without using any optimization methods. The case in this study also shows that the C4.5 algorithm can work well for a classification model by combining PSO as a feature selection, namely determining classes for new mental retardation students using.

In this study, PSO worked well for selecting existing attributes and increasing the accuracy of the model applied. PSO removes redundant features and selects the most useful features to increase classification simplification. When compared with previous research [22], the data used in this study has advantages in the input data used.

4. CONCLUSION

This study uses the C4.5 algorithm and a combination of C4.5 and PSO. The data used in this research is the identification data of new mental retardation students obtained from SLB Negeri 1 Pelaihari. This study focuses on the application of the PSO algorithm as feature selection in the C4.5 data mining classification method. The validation used in this model is 10-fold cross-validation, while the evaluation of the model in this application uses a confusion matrix.

The results of this study show that the accuracy produced by C4.5 before using PSO for C4.5 optimization is 91%. The results of the accuracy of C4.5 using a PSO of 93% using parameters, namely a population of 30 and a maximum number of generations of 50 with an execution time of 19 seconds. Of the 20 attributes used, there are 6 attributes resulting from the PSO feature selection that can affect the results of accuracy, namely A07, A09, A12, A13, A14, and A17. This study shows that the PSO algorithm is proven to be used for feature selection implementation and increases the accuracy value of the C4.5 algorithm by 2%.

From this research, it is hoped that it will be useful and can be used by the school to prepare classes for new mental retardation students. For further research, you can try using different datasets, more attributes, and using other optimization algorithm methods so that they can produce a higher level of accuracy.

Acknowledgments

The author would like to thank SLB Negeri 1 Pelaihari who has contributed in the form of research data.

REFERENCES

[1] G. Sampogna *et al.*, “Undergraduate psychiatric education: current situation and way forward,” *BJPsych Int.*, vol. 19, no. 2, pp. 34–36, 2022, https://doi.org/10.1192/bji.2021.48.

[2] E. Elihami, “Supporting about ‘education’ in elementary School: A review of literature,” *Mahaguru J. Pendidik. Guru Sekol. Dasar*, vol. 3, no. 1, pp. 42–48, 2022, https://doi.org/10.33487/mgr.v3i1.3248.

[3] P. Astuti, A. M. Huda, and R. N. Setyowati, “Opportunities and Challenges of Universitas Negeri Surabaya in Fulfillment of the Rights of Education for People with Mentally Retardation,” *Proc. Int. Jt. Conf. Arts Humanit. 2021 (IJCAH 2021)*, vol. 618, pp. 915–918, 2022, https://doi.org/10.2991/assehr.k.211223.158.

[4] M. E. Wood, K. P. Brown, A. R. Bitting, C. Slobogin, and B. Bowerman, “Legal Admissibility of the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR),” *J. Pers. Assess.*, vol. 104, no. 2, pp. 289–301, 2022, https://doi.org/10.1080/00223891.2021.1951742.

[5] A. Marsyaelina, S. Sudiyatno, and R. Iskandar, “Appropriate learning media for mild mentally impaired students at inclusive vocational schools: A literature review,” *J. Pendidik. Vokasi*, vol. 12, no. 1, pp. 93–99, 2022, https://doi.org/10.21831/jpv.v12i1.47717.

[6] M. Marlborough, A. Welham, C. Jones, S. Reckless, and J. Moss, “Autism spectrum disorder in females with fragile X syndrome: a systematic review and meta-analysis of prevalence,” *J. Neurodev. Disord.*, vol. 13, no. 1, pp. 1–19, 2021, https://doi.org/10.1186/s11689-021-09362-5.

[7] A. B. Osman *et al.*, “Examining Mental Disorder/Psychological Chaos through Various ML and DL Techniques: A Critical Review,” *Ann. Emerg. Technol. Comput.*, vol. 6, no. 2, pp. 61–70, 2022, https://doi.org/10.33166/aetic.2022.02.005.

[8] K. Y. Kim, “Literature Review in Preparation for Professional Life of the People with Developmental Disability,” *Asia-pacific J. Converg. Res. Interchang.*, vol. 7, no. 8, pp. 103–112, 2021, https://doi.org/10.47116/apjcri.2021.08.09.

[9] N. A. Rizqianti, E. Ediyanto, and M. Efendi, “Completeness of Shopping Skills Media Features for Students with Mental Retardation at Special Junior High School: A Literature Review,” *J. Disruptive Learn. Innov.*, vol. 3, no. 1, p. 8, 2021, https://doi.org/10.17977/um072v3i12021p8-13.

[10] N. N. P. Lutfianah, “Literature Review: Utilization of Social Media To Increase Social Relationship Children With Mild Mentally Retarded At Inclusive School,” *Int. Conf. Humanit. Soc. Sci.*, vol. 1, no. 1, pp. 523–528, 2022, https://doi.org/10.1234/ichss.v1i1.60.

[11] Z. Wang, X. Xu, Q. Han, Y. Chen, J. Jiang, and G. X. Ni, “Factors associated with public attitudes towards persons with disabilities: a systematic review,” *BMC Public Health*, vol. 21, no. 1, pp. 1–15, 2021, https://doi.org/10.1186/s12889-021-11139-3.

[12] P. Kaur, J. K. Chahal, and T. Sharma, “A data mining approach for crop yield prediction in agriculture sector,” *Adv. Math. Sci. J.*, vol. 10, no. 3, pp. 1425–1430, 2021, https://doi.org/10.37418/amsj.10.3.32.

[13] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J.*

Appl. Sci. Technol. Trends, vol. 2, no. 1, pp. 20–28, 2021, https://doi.org/10.38094/jastt20165.

[14] S. Gupta, N. Mehndiratta, S. Sinha, S. Chaturvedi, and M. Singla, "Data Mining Techniques and Algorithms in Psychiatric Health: A Systematic Review," *Biomed. Data Min. Inf. Retr.*, pp. 263–291, 2021, https://doi.org/10.1002/9781119711278.ch10.

[15] N. Werghi and F. Kamoun, "A decision-tree-based system for student academic advising and planning in information systems programmes," *Int. J. Bus. Inf. Syst.*, vol. 5, no. 1, pp. 1–18, 2010, https://doi.org/10.1504/ijbis.2010.029477.

[16] C. Sirichanya and K. Kraisak, "Semantic data mining in the information age: A systematic review," *Int. J. Intell. Syst.*, vol. 36, no. 8, pp. 3880–3916, 2021, https://doi.org/10.1002/int.22443.

[17] M. Nabeel, S. Majeed, M. J. Awan, H. Muslih-Ud-din, M. Wasique, and R. Nasir, "Review on effective disease prediction through data mining techniques," *Int. J. Electr. Eng. Informatics*, vol. 13, no. 3, pp. 717–733, 2021, https://doi.org/10.15676/IJEEI.2021.13.3.13.

[18] M. Effendi, R. D. Ariani, and R. Astuti, "Determination of Work Schedule Based on Employee Data Classification Using the Decision Tree Algorithm C4.5 Method," *Ind. J. Teknol. dan Manaj. Agroindustri*, vol. 10, no. 3, pp. 249–259, 2021, https://doi.org/10.21776/ub.industria.2021.010.03.6.

[19] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, p. 100071, 2022, https://doi.org/10.1016/j.dajour.2022.100071.

[20] D. Saputra, W. Irmayani, D. Purwaningtias, and J. Sidauruk, "A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 2, pp. 84–95, 2021, https://doi.org/10.25008/ijadis.v2i2.1221.

[21] M. N. Ashtiani and B. Raahemi, "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 72504–72525, 2022, https://doi.org/10.1109/access.2021.3096799.

[22] A. Dardzinska-Głebocka and M. Zdrodowska, "Analysis children with disabilities self-care problems based on selected data mining techniques," *Procedia Comput. Sci.*, vol. 192, pp. 2854–2862, 2021, https://doi.org/10.1016/j.procs.2021.09.056.

[23] Y. A. Gerhana, I. Fallah, W. B. Zulfikar, D. S. Maylawati, and M. A. Ramdhani, "Comparison of naive Bayes classifier and C4.5 algorithms in predicting student study period," in *Journal of Physics: Conference Series*, vol. 1280, no. 2 https://doi.org/10.1088/1742-6596/1280/2/022022.

[24] M. Ahmad, M. Abrori, A. Syukur, and M. A. Soeleman, "Improving C4 . 5 Algorithm Accuracy With Adaptive Boosting Method For Predicting Students in Obtaining Education Funding," *Journal of Development Research*, vol. 6, no. November, pp. 137–140, 2022, https://doi.org/10.28926/jdr.v6i2.205.

[25] B. A. C. Permana, R. Ahmad, H. Bahtiar, A. Sudianto, and I. Gunawan, "Classification of diabetes disease using decision tree algorithm (C4.5)," *J. Phys. Conf. Ser.*, vol. 1869, no. 1, 2021, https://doi.org/10.1088/1742-6596/1869/1/012082.

[26] M. I. Ahmadi, F. Apriani, M. Kurniasari, S. Handayani, and D. Gustian, "Sentiment Analysis Online Shop on the Play Store Using Method Support Vector Machine (SVM)," *Semin. Nas. Informatika (SEMNASIF)*, vol. 1, no. 1, pp. 196–203, 2020, https://doi.org/10.1088/1742-6596/1819/1/012030.

[27] O. O. Akinola, A. E. Ezugwu, J. O. Agushaka, R. A. Zitar, and L. Abualigah, "Multiclass feature selection with metaheuristic optimization algorithms: a review," *Neural Computing and Applications*, vol. 34, no. 22, 2022. https://doi.org/10.1007/s00521-022-07705-4.

[28] S. Siswanto, A. Abdussomad, W. Gata, N. K. Wardhani, G. Gata, and B. H. Prasetyo, "The Feasibility of Credit Using C4.5 Algorithm Based on Particle Swarm Optimization Prediction," *Proceeding Electr. Eng. Comput. Sci. Informatics*, vol. 6, 2019, https://doi.org/10.23919/eecsi48112.2019.8977074.

[29] Tim Dapodikbud, "Dapodikbud SLB Negeri 1 Pelaihari," *Tim Dapodikbud*, 2022. https://sekolah.data.kemdikbud.go.id/index.php/chome/profil/6065d253-30f5-e011-b32d-a93a3a8e6354.

[30] A. Nazir, A. Akhyar, Y. Yusra, and E. Budianita, "Toddler Nutritional Status Classification Using C4.5 and Particle Swarm Optimization," *Sci. J. Informatics*, vol. 9, no. 1, pp. 32–41, 2022, https://doi.org/10.15294/sji.v9i1.33158.

[31] A. Miklosik and N. Evans, "Impact of Big Data and Machine Learning on Digital Transformation in Marketing: A Literature Review," *IEEE Access*, vol. 8, pp. 101284–101292, 2020, https://doi.org/10.1109/access.2020.2998754.

[32] W. Bolander, N. N. Chaker, A. Pappas, and D. R. Bradbury, "Correction to: Operationalizing salesperson performance with secondary data: aligning practice, scholarship, and theory," *J. Acad. Mark. Sci.*, vol. 49, no. 6, pp. 1267–1268, 2021, https://doi.org/10.1007/s11747-021-00789-9.

[33] Normah, I. Yulianti, D. Novianti, M. N. Winnarto, A. Zumarniansyah, and S. Linawati, "Comparison of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program," in *Journal of Physics: Conference Series*, vol. 1641, no. 1, 2020, https://doi.org/10.1088/1742-6596/1641/1/012006.

[34] A. A. Nagra *et al.*, "Hybrid self-inertia weight adaptive particle swarm optimisation with local search using C4.5 decision tree classifier for feature selection problems," *Conn. Sci.*, vol. 32, no. 1, pp. 16–36, 2020, https://doi.org/10.1080/09540091.2019.1609419.

[35] A. S. Talita, O. S. Nataza, and Z. Rustam, "Naïve Bayes Classifier and Particle Swarm Optimization Feature Selection Method for Classifying Intrusion Detection System Dataset," *J. Phys. Conf. Ser.*, vol. 1752, no. 1, 2021, https://doi.org/10.1088/1742-6596/1752/1/012021.

[36] F. Tempola, M. Muhammad, A. K. Maswara, and R. Rosihan, "Rule Formation Application based on C4.5 Algorithm for Household Electricity Usage Prediction," *Trends Sci.*, vol. 19, no. 3, 2022, https://doi.org/10.48048/tis.2022.2167.

[37] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, "Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, https://doi.org/10.1088/1757-899x/874/1/012024.

[38] N. A. Prahastiwi, R. Andreswari, and R. Fauzi, "Students Graduation Prediction Based on Academic Data Record Using the Decision Tree Algorithm C4.5 Method," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 8, no. 3, pp. 295–304, 2022, https://doi.org/10.33330/jurteksi.v8i3.1680

[39] X. Wang, C. Zhou, and X. Xu, "Application of C4.5 decision tree for scholarship evaluations," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 179–184, 2019, https://doi.org/10.1016/j.procs.2019.04.027.

[40] X. Meng, P. Zhang, Y. Xu, and H. Xie, "Construction of decision tree based on C4.5 algorithm for online voltage stability assessment," *Int. J. Electr. Power Energy Syst.*, vol. 118, p. 105793, 2020, https://doi.org/10.1016/j.ijepes.2019.105793.

BIOGRAPHY OF AUTHORS

Sausan Hidayah Nova is a postgraduate student of Magister Information Systems at Universitas Diponegoro. She graduated with a bachelor of computer degree from the program study of Information Technology at Universitas Lambung Mangkurat in 2020. Her research interests include data mining and information system. Email: novasausan@gmail.com

Budi Warsito obtained his S.Si degree at Universitas Gadjah Mada in 1998. Then he obtained his M.Si degree at Universitas Gadjah Mada in 2005. Continuing doctoral studies at Universitas Gadjah Mada in 2017. Currently a lecturer at Universitas Diponegoro. His research interests include neural networks, data mining, and information systems. Email: budiwarsitoundip@gmail.com

Aris Puji Widodo obtained his S.Si degree at Universitas Diponegoro in 1997. Gadjah Mada University. Continuing his master of engineering studies from Institut Teknologi Bandung in 2002. Doctoral degree at Universitas Gadjah Mada in 2015. Currently a lecturer at Universitas Diponegoro.
Email: apwcourses2010@gmail.com