

A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction

Dedi Saputra¹, Windi Irmayani², Deasy Purwaningtyas³, Juniato Sidauruk⁴, Burcu Gurbuz⁵

^{1,2,3,4}Bina Sarana Informatika of University, Indonesia

⁵Johannes Gutenberg-University of Mainz, Germany; Usk'udar University, Istanbul, Turkey

Article Info

Article history:

Received Jun 26, 2021

Revised Sept 13, 2021

Accepted Oct 22, 2021

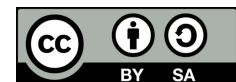
Keywords:

Heart disease
Classification Algorithms
Optimization
PSO

ABSTRACT

Heart disease is a general term for all of types of the disorders which is affects the heart. This research aims to compare several classification algorithms known as the C4.5 algorithm, Naïve Bayes, and Support Vector Machine. The algorithm is about to optimize of the heart disease predicting by applying Particle Swarm Optimization (PSO). Based on the test results, the accuracy value of the C4.5 algorithm is about 74.12% and Naïve Bayes algorithm accuracy value is about 85.26% and the last the Support Vector Machine algorithm is about 85.26%. From the three of algorithms above then continue to do an optimization by using Particle Swarm Optimization. The data is shown that Naïve Bayes algorithm with Particle Swarm Optimization has the highest value based on accuracy value of 86.30%, AUC of 0.895 and precision of 87.01%, while the highest recall value is Support Vector Machine algorithm with Particle Swarm Optimization of 96.00%. Based on the results of the research has been done, the algorithm is expected can be applied as an alternative for problem solving, especially in predicting of the heart disease.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Dedi Saputra,
Department of Information System, Universitas Bina Sarana Informatika.
Email: dedi.dst@bsi.ac.id

1. INTRODUCTION

The soaring number of patients suffering from various diseases, especially heart disease, has an impact and affects the high cost of treatment makes the government or health people seek prevention early on. One of the answers to this problem can be found solutions using artificial intelligence and data mining[1]. The heart is one of the vital organs in the body that serves to pump blood throughout the body of blood vessels. According to Sloane, the Heart is a hollow organ and has four spaces located between the two lungs in the middle of the thoracic cavity [2][3]. More than 36 million people die each year from Non-Communicable Diseases (PTM). Globally PTM is the number one cause of death is cardiovascular disease. Cardiovascular disease (PKV) is a disease caused by impaired heart and vascular function such as coronary heart disease, hypertension, and stroke [4]. In 2008 an estimated 17.3 million deaths were caused by cardiovascular disease. Deaths caused by cardiovascular disease, especially coronary heart disease and stroke are expected to continue to rise to 23.3 million deaths by 2030. Coronary Heart Disease (CHD) is a disorder of heart function due to lack of blood heart muscle due to the narrowing of coronary blood vessels. The fact that a heart attack occurs without obvious symptoms, some existing methods for early detection can

not overcome it.[5]. Clinically, it is characterized by chest pain or discomfort in the chest or chest feels severely depressed when hiking / heavy work or walking in a hurry when walking on a flat road or walking long distances. WHO data onto 2015 shows that 70% of deaths worldwide are caused by Non-Communicable Diseases (39.5 million out of 56.4 deaths). Of all deaths from Non-Communicable Diseases (PTM), 45% were caused by heart and vascular disease, which is 17.7 million out of 39.5 million deaths.[6].

According to the main results of Riskesdas 2018 shows the prevalence of Heart Disease based on the diagnosis of doctors in Indonesia by 1.5%, with the highest prevalence rating : 1) North Kalimantan 2.2%; 2) DIY 2%; 3) Gorontalo 2%. In addition to the three provinces, there are also 8 other provinces with a higher prevalence when compared to the national prevalence. The eight provinces are: 1) Aceh (1.6%); 2) West Sumatra (1.6%); 3) DKI Jakarta (1.9%); 4) West Java (1.6%); 5) Central Java (1.6%); 6) East Kalimantan (1.9%); 7) North Sulawesi (1.8%); and 8) Central Sulawesi (1.9%) [7].

Development of software repositories to gain a better understanding of the data Currently there is an increase. Interest in data mining and education systems, making educational data mining a growing new research community. In the real world, predicting a disease is a challenging task. Data mining is the process of finding patterns in large amounts of data[8]. Trend Data mining is emerging as a promising new workflow that provides a wide range of techniques, methods and tools to perform comprehensive analysis of available data in various fields.

Classification is a simple process to find the type of data analysis that can help people predict sample class labels or data concepts, with the aim of being able to use an approach in predicting object classes whose class labels are not yet known. Another goal is to be able to predict distinct and unordered labels in very large data sets. Like a classification, a test setting is used to build a predictive tool while still using an independent test setting to be used to assess its accuracy. Another fact in a classification approach, there is a decrease in classification for category and numeric attributes. Of the twenty classification methods, Bayes Net, Naïve Bayes, Classification through Regression, Logistic Regression and Random Forest are the best classification methods. For mixed attribute datasets, Naïve Bayes, Bayes Net and Random Forest are the best classification methods. For Regression Classification numerical attribute datasets, NBTree and multiclass Classifier are the best methods. For the categorical attributes of the NB-Tree dataset, Classification via Regression and Bayes Net methods is best[9]. Of these above five rules classification method based on PART and Decision Tree method is the best. However, several methods have shown the best results where the right technique is chosen for the right data. No particular classifier performs best for all datasets.

The quality health service is one of the basic necessities of any person or customer. To predict the large potential for heart disease, it is necessary to have research related to the detection of heart disease [10]. Ideally data mining is based on several disciplines, such as machine learning, artificial intelligence, probability, and statistics[11]. In a study on the Naïve Bayes data mining method based on Particle Swarm Optimization for the detection of deep heart disease [12], measurement of Naives Bayes resulted in an accuracy of 82.14%, while with Naives Bayes based on Particle Swarm Optimization, the accuracy increased to 92.86%. In another study that discusses the C4.5 algorithm as the simplest classification and is easy to implement, however, the C4.5 Algorithm still has weaknesses for handling high-dimensional data. The research which aims to apply genetic algorithms with attribute selection so as to reduce the dimensions of the data, and identify features of the data set using the C4.5 algorithm method has already a fairly good accuracy of 76.66% with the attribute selection process by a genetic algorithm, the model formed can be increased again to 77.40% in the classification of heart disease [13].

Support Vector Machine is one of the classical machine learning techniques that can still help solve big data classification problems. Especially, it can help the multidomain applications in a big data environment. Support Vector Machine is one way of machine learning whose process is based on the structural principle of risk minimization (SRM) has the goal of getting the best hyperplane in separating two classes in the scope of input[14][15].

2. RESEARCH METHOD

This study uses data based on the population where the research subjects are people with heart disease by entering test data derived from training data, which serves to provide research information[16]. Each study on heart disease prediction uses a different method, not yet known the right algorithm to predict it. For this reason, a model of approach with a research framework as in figure 1, so that from this study can compare several classification algorithms namely Decision Tree (C4.5), Naïve Bayes and Support Vector Machine (SVM) with PSO to solve problems then test the performance of some of these methods. The data used is based on data from heart disease sufferers by entering test data derived from training data. Where the final data is obtained after the preprocessing process. It then tested six different algorithms using rapid miner 9.1 data mining software to test C4.5, Naïve Bayes, Support Vector Machine, C4.5+PSO, Naïve Bayes+PSO and Support Vector Machine+PSO. The results of the algorithm with the best accuracy will later be used in predicting heart disease.

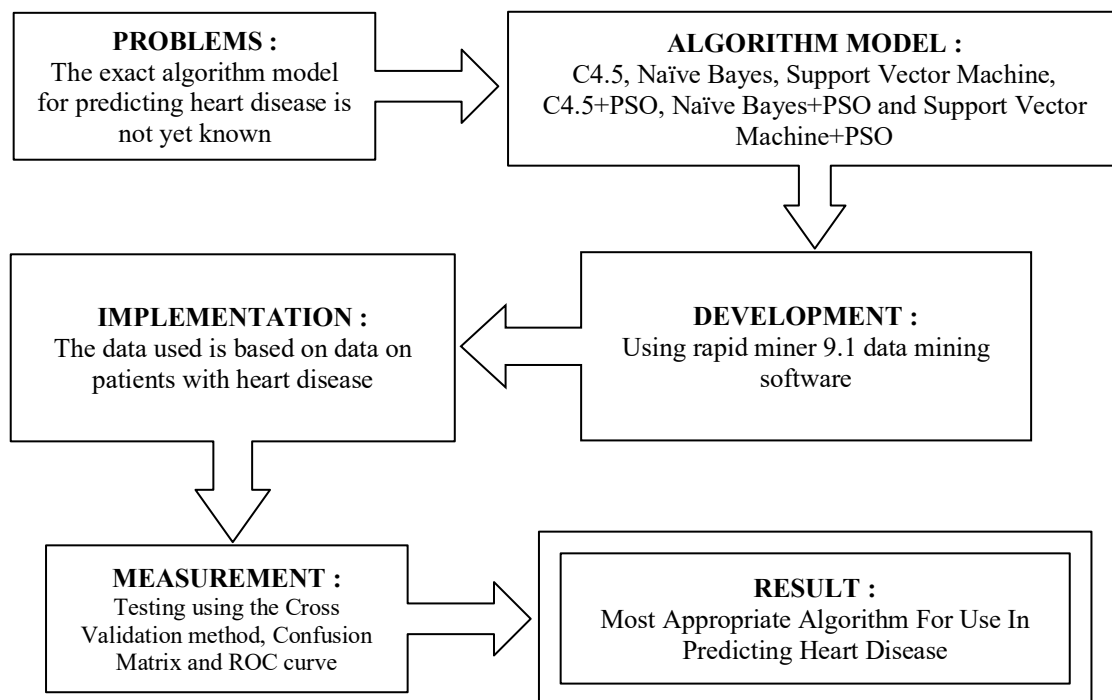


Figure 1. Review of Literature

2.1 Data Mining

Data mining is the activity of studying the collection, use of historical data to find regularity, patterns or relationships in large data sets. The output of this data mining can be used to improve decision making in the future. Data Mining is an automated process of existing data. The data to be processed is a large data, with the aim of obtaining relationships or patterns that may give a useful indication [17]. According to [18] in his book entitled "Data mining for Classification and Clustering of Data", mentioning that Data mining is an analysis of the process of discovery of knowledge in databases or knowledge discovery in databases abbreviated as KDD. According to Sumathi and Sivandham (2009), data mining is also defined as part of the process of extracting knowledge in a database known as Knowledge Discovery in Database (KDD) which consists of several stages, namely Cleaning and Integration, Selection and Transformation, Data mining, and Evaluation and Interpretation [19].

From some of the above definitions it can be concluded that data mining is a method of digging up valuable information that is buried or hidden in a very large data set so that it is found an interesting form that was previously unknown. The word mining itself means an attempt to obtain a small amount of valuables from a large amount of basic materials. Therefore, data mining actually has a long root in the fields of science such as artificial intelligence, machine learning, statistics and databases.

2.2 C4.5 Algorithm

C4.5 algorithm is one of the very effective decision tree algorithms to do organization. The decision tree method turns a very large reality into a decision tree that reflect the rules. A decision tree consists of internal nodes that determine tests on individual input variables or attributes that divide data into smaller set of sections, and a series of leaf nodes assign a class to each observation in the resulting segment [20].

C4.5 algorithm is one of the algorithms used in decision tree, especially in machine learning area [21]. According to Larose, The C4.5 algorithm is Quinlan's extension of his own ID3 algorithm for generating decision trees. C4.5 is a method and prediction that is very strong and very widely used. C4.5 used information gain to select the attributes that will be used for object separation [22].

2.3 Naive Bayes

Naive Bayes is a classification algorithm with a formula that is simple and easy to apply [23]. Naive Bayes is included in the supervised learning, so at the learning stage it takes preliminary data in the form of training data to be able to make decisions [24]. Naive bayes is one of the statistical methods for classification that makes it possible to capture uncertainty about a model in a principled way on defining probability outcomes [25][26].

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) was first presented by Vapnik in 1992 as a compatible series of excellent concepts in the field of pattern recognition. Support Vector Machine is a machine learning method that works on the principle of structural risk minimization (SRM) with the aim of finding the best hyperplane that separates two classes in the input space [14]. The optimum hyperplane is a hyperplane located in the central in two sets of objects of two classes. The optimum separator hyperplane between the two classes can be found by measuring the hyperplane's margin and finding its maximum point. The margin is the distance in the hyperplane and the closest system of each class. The closest system is referred to as root vector [27].

According to [28][29] Support Vector Machine (SVM) is defined as a set of related learning methods that analyze data and recognize system, which are then used for classification and regression analysis. SVM takes a set of input data and predicts for each given input, which comes from two classes that are then classified by looking for the best hyperplane value.

According to [30] Support Vector Machine (SVM) is a classification mode to find the best hyperplane value competent of invention best global solutions. So that the value of accuracy is not easily capricious.

According to [31] Support Vector Machine (SVM) is a knowledge that guides to quadratic programming with linear barrier. Based on the risk minimization of structured principles, SVM strives to minimize the upper limit of error generalizations rather than empirical errors, so that new prediction models to good purpose avoid over-fitting problems. In extension, SVM models perform in high-dimensional attribute spaces formed by nonlinear mapping of N-dimensional vector input x into K-dimensional attribute space ($K > N$) through the use of nonlinear ϕ functions (x).

SVM has a basic principle of linear classifier that is a classification case that can be linearly separated, but SVM has been developed to work on non-linear problems by inserting kernel concepts in high-resolution workspaces [32].

Based on the above understanding can be concluded Support Vector Machine (SVM) is a classification method that maximizes the limit of hyperplane (maximal margin hyperplane). If in ANN all practice data will be studied for the practice process, while in SVM it is different because only a selected amount of data contributes to forming the model used in the classification to be

studied [33]. This is an advantage of SVM because not all practice data will be seen to be bound in each interaction of the practice. The provided data is seen to as support vector so the method is called Support Vector Machine.

2.5 Particle Swarm optimization (PSO)

Particle Swarm Optimization (PSO) is a global heuristic optimization technique introduced by Doctors Kennedy and Eberhart in 1995 inspired by the social behavior of flocks of birds trying to achieve unintroduced goals [34]

According to [28] Particle Swarm Optimization (PSO) is a type of intelligence algorithm capable of optimizing a related variable that is most effective.

According to [35] Particle Swarm Optimization (PSO) is an evolutionary calculation technique. Similar to genetic algorithms, PSO is an optimization tool. It is inspired by social behavior among individuals. Particles (individuals) that represent potential solutions to problems move through n-dimensional search space. Each particle i maintains a record of the best performance position in the vector called p_{best} .

According to [28] Particle Swarm Optimization (PSO) is a computational method that optimizes problems with iterative solutions to improve candidates with regard to a certain size of quality. The movement of each particle is influenced by the local position guided towards the most known position in the search for space, which is updated as a better position than other particles.

According to [36] Particle Swarm Optimization (PSO) is an evolutionary computing technique capable of producing globally optimal solutions in search space through individual interactions in a swarm of particles. Each particle conveys information in the form of its best position to the other particles and adjusts the position and speed of each based on the information received about the best position.

Particle Swarm Optimization (PSO) is a tool to deal with optimization problems [37]. Although relatively new, many have applied PSO algorithms, because it is quite simple and has faster computing speed compared to other optimization algorithms such as Genetic Algorithm (GA). Each particle in the PSO is also associated with the speed at which particles fly through the search space with dynamic speeds adjusted for their historical behavior. Therefore, particles have a tendency to fly towards better search areas during the search process.

Based on the above understanding, it can be concluded that Particle Swarm Optimization (PSO) is an optimization method that is able to optimize the nearest variable to achieve maximum accuracy.

3. RESULTS AND DISCUSSION

In this study, the method used was experimental research where experimental research involved investigating causal relationships using tests controlled by the researcher himself. In this study on predicting heart disease using data based on data on people with heart disease by entering test data derived from training data where the final data is obtained after carrying out the preprocessing process. Then, six different algorithms were tested using the Rapid Miner 9.1 data mining software to test the C4.5 algorithm, Naïve Bayes, Support Vector Machine, C4.5 + PSO, Naïve Bayes + PSO and Support Vector Machine + PSO. The algorithm results with the best accuracy will be used to predict heart disease.

Algorithm C4.5

In this study, there are two kinds of variables, namely the dependent variable (dependent) and the independent variable (free), among others: Based on the test results from data on patients with heart disease that have been preprocessed, the accuracy of the testing data using the C4.5 Algorithm is known as following.

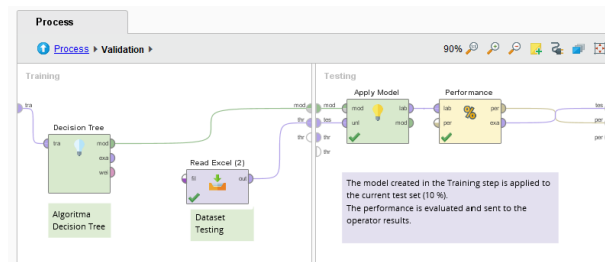


Figure 2. C4.5 Algorithm Testing

Table View Plot View

accuracy: 74.12% +/- 14.55% (micro average: 74.07%)

	true present	true absent	class precision
pred. present	54	25	68.35%
pred. absent	17	66	79.52%
class recall	76.06%	72.53%	

Figure 3. the accuracy value of the C4.5 algorithm

TP = 54; FP = 25; TN = 66; FN= 17

Accuracy = $((TP+TN) / (TP+TN+FN+FP)) * 100\%$

Accuracy = $((54+66) / (54+66+17+25)) * 100\%$

Accuracy = 74,12%

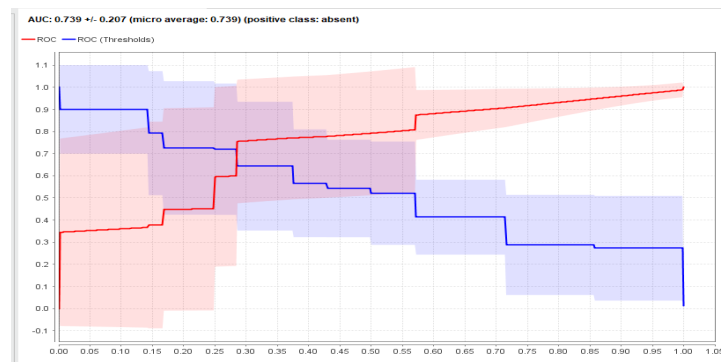


Figure 4. the AUC value of the C4.5 algorithm

Naïve Bayes Algorithm

Based on the test results from data on patients with heart disease who have been preprocessed, the calculation of the accuracy of the testing data using the Naïve Bayes Algorithm shows the level of accuracy as follows.

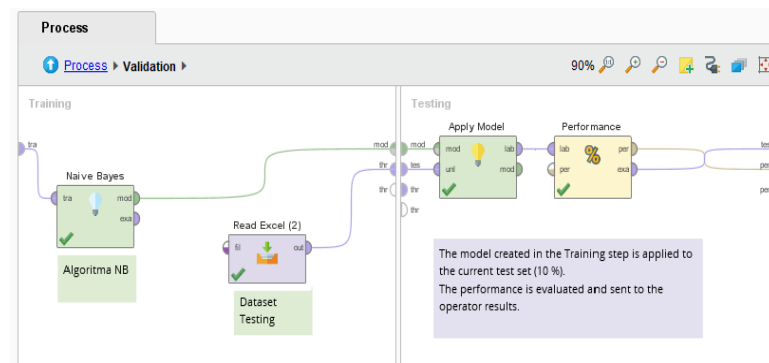


Figure 5. Naïve Bayes Testing

Table View Plot View

accuracy: 85.26% +/- 8.60% (micro average: 85.19%)

	true present	true absent	class precision
pred. present	59	12	83.10%
pred. absent	12	79	86.81%
class recall	83.10%	86.81%	

Figure 6. accuracy value naïve Bayes

TP = 59; FP = 12; TN = 79; FN= 12

Accuracy = $((TP+TN) / (TP+TN+FN+FP)) * 100\%$

Accuracy = $((59+79) / (59+79+12+12)) * 100\%$

Accuracy = 85,26%

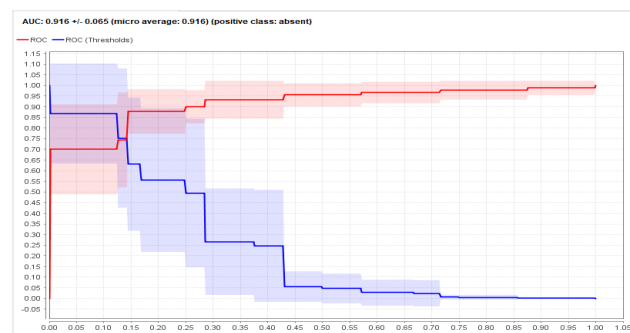


Figure 7. The AUC value of the Naïve Bayes algorithm

Support Vector Machine Algorithm

Based on test results from data on patients with heart disease who have been preprocessed, the calculation of the accuracy of the testing data using the Support Vector Machine algorithm is known for the level of accuracy as follows.

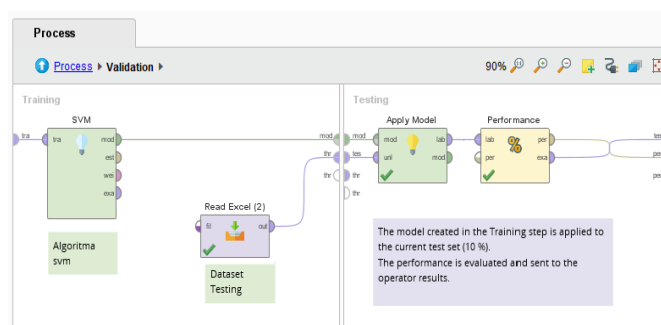


Figure 8. Support Vector Machine Testing

Table View Plot View

accuracy: 85.26% +/- 8.60% (micro average: 85.19%)

	true present	true absent	class precision
pred. present	59	12	83.10%
pred. absent	12	79	86.81%
class recall	83.10%	86.81%	

Figure 9. support vector machine accuracy value

$TP = 59; FP = 12; TN = 79; FN = 12$
 $Accuracy = ((TP+TN) / (TP+TN+FN+FP)) * 100\%$
 $Accuracy = ((59+79) / (59+79+12+12)) * 100\%$
 $Accuracy = 85,26\%$

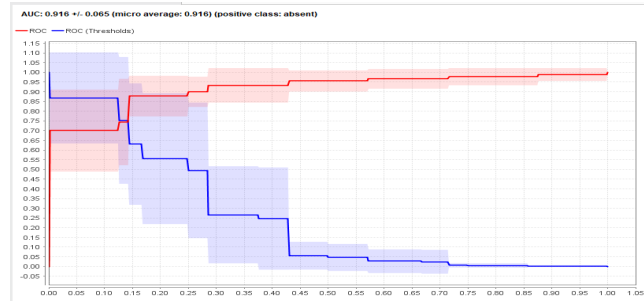


Figure 10. the AUC value of the support vector machine algorithm

Algorithm C4.5 with Particle Swarm Optimization

Based on test results from data on patients with heart disease who have been preprocessed, the calculation of the accuracy of the testing data uses the C4.5 Algorithm. with Particle Swarm Optimization, the level of accuracy is known as follows.

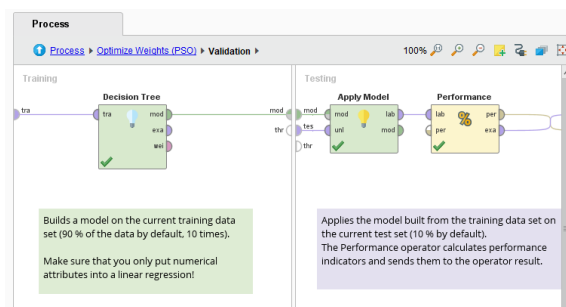


Figure 11. Testing C4.5 + Particle Swarm Optimization

Table View Plot View

accuracy: 80.00% +/- 5.84% (micro average: 80.00%)

	true present	true absent	class precision
pred. present	95	29	76.61%
pred. absent	25	121	82.88%
class recall	79.17%	80.67%	

Figure 12. the accuracy value of the C4.5 algorithm + particle swarm optimization

$TP = 95; FP = 29; TN = 25; FN = 121$
 $Accuracy = ((TP+TN) / (TP+TN+FN+FP)) * 100\%$
 $Accuracy = ((95+121) / (95+121+25+29)) * 100\%$
 $Accuracy = 80,00\%$

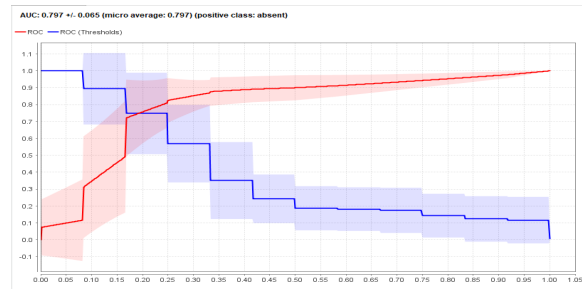


Figure 13. AUC value for the C4.5 algorithm + particle swarm optimization

Naïve Bayes Algorithm with Particle Swarm Optimization

Based on the test results from data on patients with heart disease who have been preprocessed, the calculation of the accuracy of the testing data using the Naïve Bayes Algorithm with Particle Swarm Optimization is known for the level of accuracy as follows.

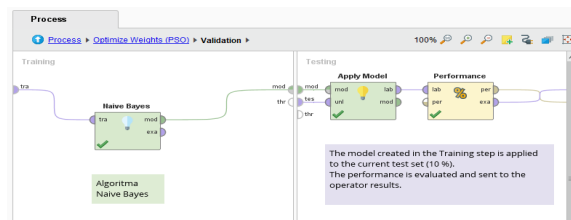


Figure 14. Naïve Bayes Testing + Particle Swarm Optimization

Table View Plot View

accuracy: 86.30% +/- 5.80% (micro average: 86.30%)

	true present	true absent	class precision
pred. present	99	16	86.09%
pred. absent	21	134	86.45%
class recall	82.50%	89.33%	

Figure 15. the accuracy value of the naïve Bayes algorithm + particle swarm optimization

TP = 99; FP = 16; TN = 134; FN= 21

Accuracy = $((TP+TN) / (TP+TN+FN+FP)) * 100\%$

Accuracy = $((99+134) / (99+134+21+16)) * 100\%$

Accuracy = 86,30%

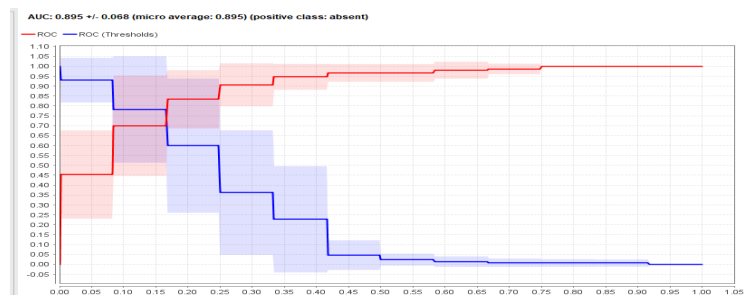


Figure 16. the AUC value of the naïve Bayes algorithm + particle swarm optimization

Support Vector Machine Algorithm with Particle Swarm Optimization

Based on the test results from data on patients with heart disease who have been preprocessed, the accuracy of the testing data using the Support Vector Machine Algorithm with Particle Swarm Optimization is known as the following.

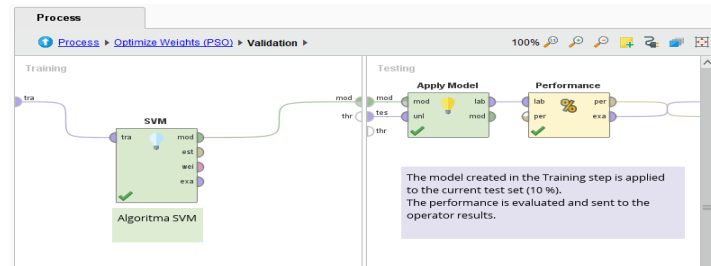


Figure 17. Testing the Support Vector Machine algorithm + particle swarm optimization

Table View Plot View

accuracy: 84.81% +/- 6.64% (micro average: 84.81%)

	true present	true absent	class precision
pred. present	85	6	93.41%
pred. absent	35	144	80.45%
class recall	70.83%	96.00%	

Figure 18. the accuracy value of the Support Vector Machine algorithm + particle swarm optimization

TP = 85; FP = 6; TN = 144; FN= 35

Accuracy = $((TP+TN) / (TP+TN+FN+FP)) * 100\%$

Accuracy = $((85+144) / (85+144+35+6)) * 100\%$

Accuracy = 84,81%

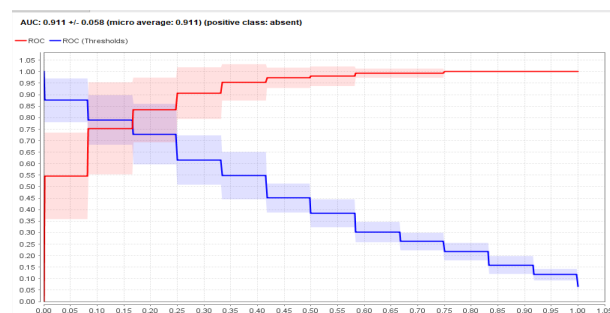


Figure 19. AUC value for the Support Vector Machine algorithm + particle swarm optimization

Table 1. Comparison of all algorithm values

Dataset	C4.5	NB	SVM	C4.5+PSO	NB+PSO	SVM+PSO
Accuracy	74,1%	85,26%	85,26%	80,00%	86,30%	84,81%
Precision	80,38%	86,88%	86,88%	83,22%	87,01%	80,87%
Recall	72,67%	86,67%	86,67%	80,67%	89,33%	96,00%
AUC	0,379	0,916	0,916	0,797	0,895	0,911

From the discussion above obtained Naïve Bayes algorithm combined with Particle Swarm Optimization algorithm turns out to have the highest value based on accuracy value of 86.30%, AUC of 0.895 and precision of 87.01% while the highest recall value obtained by Support Vector Machine algorithm combined with Particle Swarm Optimization is 96.00%.

4. CONCLUSION

Based on the results of the above research discussion can be concluded that: 1). To know which algorithms are good at predicting heart disease; 2) The compared algorithms are modeling algorithms C4.5, NB, and SVM, where the data used are data of heart disease patients; 3). This research also applies PSO optimization as weighting attributes of data mining classification techniques C4.5, NB and SVM; 4) From this research it was obtained that the use of PSO optimization can improve the accuracy of the algorithms C4.5, NB and SVM; 5). It is concluded that Naïve Bayes algorithm method with Particle Swarm Optimization is the best method of six modeling algorithms that have been tested and can be used as an alternative solution in predicting heart disease.

ACKNOWLEDGEMENTS

The authors gratitude is conveyed to all parties, especially the Ministry of Health the Republic of Indonesia who have published data on Heart disease in Indonesia in 2018. The authors also expressed his appreciation and gratitude to all colleagues who have helped contribute thought suggestions for the perfection of this research.

REFERENCES

- [1] S. I. Novichasari and I. S. Wibisono, "Particle Swarm Optimization For Improved Accuracy of Disease Diagnosis," *J. Appl. Intell. Syst.*, vol. 5, no. 2, pp. 57–68, 2020.
- [2] R. Wajhillah, "Optimasi Algoritma Klasifikasi C4. 5 Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Jantung," *Swabumi*, vol. 1, no. 1, pp. 26–36, 2014.
- [3] F. Babič, J. Olejár, Z. Vantová, and J. Paralič, "Predictive and descriptive analysis for heart disease diagnosis," in *2017 federated conference on computer science and information systems (fedcsis)*, 2017, pp. 155–163.
- [4] M. Martiningsih and A. Haris, "Risiko Penyakit Kardiovaskuler pada Peserta Program Pengelolaan Penyakit Kronis (Prolanis) di Puskesmas Kota Bima: Korelasinya dengan Ankle Brachial Index dan Obesitas," *J. Keperawatan Indones.*, vol. 22, no. 3, pp. 200–208, 2019.
- [5] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *Biomed Res. Int.*, vol. 2020, 2020.
- [6] P2PTM Kemenkes RI, "Hari Jantung Sedunia (HJS) Tahun 2019 : Jantung Sehat, SDM Unggul," 2018.
- [7] R. I. Kemenkes, "Hasil utama RISKESDAS 2018," *Online*) http://www.depkes.go.id/resources/download/info-terkini/materi_rakorpop_2018/Hasil%20Risikesdas, vol. 202018, 2018.
- [8] P. S. Patel, J. Undavia, and D. Bhatti, "Master Course Selection Prediction Model Using Modify Hybrid Neuro-fuzzy Inference System," *ICTACT J. Soft Comput.*, vol. 11, no. 1, pp. 2205–2212, 2020.
- [9] S. Saha and D. Nandi, "Data Classification based on Decision Tree, Rule Generation, Bayes and Statistical Methods: An Empirical Comparison," *Int. J. Comput. Appl.*, vol. 129, no. 7, pp. 36–41, 2015.
- [10] R. Rachmat and S. Suhartono, "Comparative Analysis of Single Exponential Smoothing and Holt's Method for Quality of Hospital Services Forecasting in General Hospital," *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 2 SE-Articles, pp. 80–86, Aug. 2020, doi: 10.25008/bcsee.v1i2.8.
- [11] S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction," *IEEE Access*, vol. 6, pp. 29637–29647, 2018.
- [12] N. A. Widiastuti, S. Santosa, and C. Supriyanto, "Algoritma Klasifikasi data mining naïve bayes berbasis Particle Swarm Optimization untuk deteksi penyakit jantung," *Pseudocode*, vol. 1, no. 1, pp. 11–14, 2014.
- [13] N. Iriadi and N. Nuraeni, "kajian penerapan metode klasifikasi data mining algoritma C4. 5 untuk prediksi kelayakan kredit pada bank mayapada jakarta," *J. Tek. Komput.*, vol. 2, no. 1, pp. 132–137, 2016.
- [14] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1 SE-, pp. 36–44, Apr. 2021, doi: 10.25008/ijadis.v2i1.1216.
- [15] S. Suthaharan, "Support vector machine," in *Machine learning models and algorithms for big data classification*, Springer, 2016, pp. 207–235.
- [16] D. Saputra, F. Akbar, and A. Rahman, "Decision Support System For Providing Customer Reward Using Profile Matching Method: A Case Study at PT. Atlas Jakarta," *Bull. Comput. Sci. Electr. Eng.*, vol. 2, no. 1, pp. 28–37, 2021.
- [17] D. Nofriansyah, S. Kom, and M. Kom, *Konsep Data Mining Vs Sistem Pendukung Keputusan*.

- Deepublish, 2015.
- [18] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung, 2017.
 - [19] T. Kristanto and R. Arief, "Analisa Data Mining Metode Fuzzy Untuk Customer Relationship Management Pada Perusahaan Tour & Travel," *SESINDO 2013*, vol. 2013, 2013.
 - [20] R. Annisa, "ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI PENDERITA PENYAKIT JANTUNG," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 3, no. 1, pp. 22–28, 2019.
 - [21] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.
 - [22] D. Saputra, W. S. Dharmawan, M. Wahyudi, W. Irmayani, J. Sidauruk, and Martias, "Performance Comparison and Optimized Algorithm Classification," in *Journal of Physics: Conference Series*, 2020, vol. 1641, no. 1, pp. 12087–12093, doi: 10.1088/1742-6596/1641/1/012087.
 - [23] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C. 45 Dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 455–464, 2018.
 - [24] D. Sartika and D. I. Sensuse, "Perbandingan algoritma klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada studi kasus pengambilan keputusan pemilihan pola pakaian," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 3, no. 2, pp. 151–161, 2017.
 - [25] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 2, no. 2, pp. 152–161, 2018.
 - [26] P. A. Rahayuningsih, R. Maulana, W. Irmayani, D. Saputra, and D. Purwaningtias, "Feature Dependent Naïve Bayes For Network Intrusion Detection System," *J. Phys. Conf. Ser.*, vol. 1641, p. 12023, 2020, doi: 10.1088/1742-6596/1641/1/012023.
 - [27] I. Aydin, M. Karakose, and E. Akin, "A multi-objective artificial immune algorithm for parameter optimization in support vector machine," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 120–129, 2011.
 - [28] Y. Yin, D. Han, and Z. Cai, "Explore Data Classification Algorithm Based on SVM and PSO for Education Decision," *J. Conver. Inf. Technol.*, vol. 6, no. 10, pp. 122–128, 2011, doi: 10.4156/jcit.vol6.issue10.16.
 - [29] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra, and W. E. Jayanti, "Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain," *J. Phys. Conf. Ser.*, vol. 1641, p. 12060, 2020, doi: 10.1088/1742-6596/1641/1/012060.
 - [30] H. Yin, X. Jiao, Y. Chai, and B. Fang, "Scene classification based on single-layer SAE and SVM," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3368–3380, 2015.
 - [31] G. Li, J. You, and X. Liu, "Support vector machine (SVM) based prestack AVO inversion and its applications," *J. Appl. Geophys.*, vol. 120, pp. 60–68, 2015.
 - [32] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang," *J. Gaussian*, vol. 3, no. 4, pp. 811–820, 2014.
 - [33] E. Prasetyo, "Data mining konsep dan aplikasi menggunakan matlab," *Yogyakarta Andi*, 2012.
 - [34] T.-J. Hsieh, H.-F. Hsiao, and W.-C. Yeh, "Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm," *Neurocomputing*, vol. 82, pp. 196–206, 2012.
 - [35] C.-L. Huang and J.-F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1381–1391, 2008.
 - [36] F. Zhao, Y. Liu, C. Zhang, and J. Wang, "A self-adaptive harmony PSO search algorithm and its performance analysis," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7436–7455, 2015.
 - [37] L.-Y. Hsu et al., "Mutual funds trading strategy based on particle swarm optimization," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7582–7602, 2011.