ISSN: 2460 – 7223

The Application Of Particle Swarm Optimization Using Naive Bayes Method For Predicting Heart Disease

1st Eko Justino Wahyu
Master in Informatics Engineering
IIB Darmajaya
Bandar Lampung, Indonesia
wahyu.justino.2021211030@mail.darmajaya.ac.id

2nd Chairani
Master in Informatics Engineering
IIB Darmajaya
Bandar Lampung, Indonesia
chairani@darmajaya.ac.id

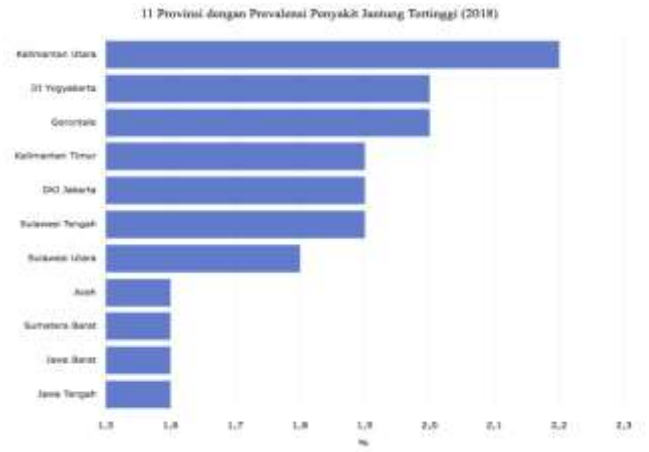
***Abstract*—Heart disease is a disease that has many sufferers and is one of the deadliest diseases in the world. Based on the 2018 Basic Health Research Report, the average prevalence of heart disease in the country (Indonesia) was 1.5% that year. It is recorded that 11 provinces have a prevalence of heart disease above the national average. North Kalimantan has the highest prevalence of heart disease in Indonesia at 2.2%, Yogyakarta and Gorontalo at 2%, East Kalimantan, DKI Jakarta and Central Sulawesi at 1.9%, North Sulawesi at 1.8%, Aceh, West Sumatra, West Java, Central Java by 1.6%, and East Nusa Tenggara by 0.7%. The cause of the increase in death rates every year is due to lack of access to find information about heart attack disease. From this problem, researchers want to develop technology in the health sector, especially using a data mining classification algorithm, namely the Naive Bayes algorithm. In the previous research, namely the journal entitled A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques with the same dataset applying the Naive Bayes algorithm without using feature selection, the accuracy value was 82.17% and the split data accuracy value was 84.28%. In this study, the researcher applyed the Naive Bayes algorithm using the Particle Swarm Optimization feature selection, the accuracy value is 84.16% and the split data accuracy is 85.12%. so it can be concluded that the Particle Swarm Optimization selection feature can be used to optimize the accuracy value of the Naive Bayes algorithm.**

Keywords—Heart Disease, Data Mining, Naïve Bayes and Particle Swarm Optimization

I. INTRODUCTION

Heart disease is a sickness that has many sufferers and is one of the deadliest diseases in the world. Cases in Indonesia, the disease that is often found in adult women is heart disease. Based on the 2018 Basic Health Research Report, the average prevalence of this disease in the country was 1.5% that year.

It reported that 11 provinces have a prevalence of heart disease above the national average. North Kalimantan has the highest prevalence of heart disease in Indonesia at 2.2%. DI Yogyakarta and Gorontalo followed with a prevalence of heart disease of 2% each. Then, East Kalimantan, DKI Jakarta, and Central Sulawesi each had a heart disease prevalence of 1.9%. Then, the prevalence of heart disease in North Sulawesi is 1.8%. Meanwhile, 1.6% of the prevalence of heart disease each came from Aceh, West Sumatra, West Java, and Central Java. The lowest prevalence of heart disease in Indonesia is in East Nusa Tenggara. The prevalence is 0.7% [1].



Source: Riskesdas, 2018

From the data above shows that many people have not responded seriously to this disease, and after a health check, they know the Sickness is in a high stage. There are many alternative ways to prevent and cure the disease, such as chemotherapy and surgery. Due to the lack of access to media/information obtained, is the reason for patients with delays in checking themselves to the doctor [2].

The cause of the increase in death rates every year is due to lack of access to find out the information about heart attack disease. To provide information about heart attack disease requires a classification system experienced by a person by doing an early check. In a system for classifying, the right method is needed to manage knowledge with the sugestion of experts so as to get accurate results [3].

In digital era, the data mining, can be implemented in the health sector. One of the best

known applications at this time is; the use of data mining to predict diseases, one of which is heart disease. Predictions for patients with heart disease can be obtained through the collection of several data on patients with heart disease stored in a database, then processed with a certain pattern so that the results can be used for early diagnosis of heart illness [4]. There are several related studies regarding the prediction of heart disease symptoms using data mining methods, including a journal entitled "*A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques*". The investigation was done by cross-validation without the feature selection process in this study, using 303 records with an accuracy of 82.17% and split data with an accuracy of 84.28% [5]. Furthermore, the journal entitled "Classification Algorithm for naive Bayes data mining based on Particle Swarm Optimization for detection of heart disease". The investigation is done with a data set using the results of a medical check-up recap from 300 people who are processed with the rapid miner tool, and from that data will be divided to 75% for training data, 25% for testing data with a feature selection process then the results are obtained accuracy of 92.86% [6]. Next is the journal entitled "*Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes*" The test devided 303 data into five subsets, 60 records, 120 records, 180 records, 240 records, and 303 records, which would be validated with 5-fold cross-validation without a feature selection process. The results obtained an average value of accuracy of 90.61% [7]. Furthermore, the journal entitled "*Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Penderita Penyakit Jantung Di Indonesia Menggunakan Rapid Miner*". The test with 500 records, divided into 80% training data and 20% testing data using split validation without a feature selection process, results in an accuracy of 70.00% [8].

Based on previous research, researchers want to develop and apply Particle Swarm Optimization to eliminate input attributes in the Naive Bayes Algorithm method with cross-validation to get an algorithm with a high level of accuracy that will be used to predict symptoms of heart disease.

II. THEORETICAL FRAMEWORK

A. Heart Disease

Heart disease is a disorder that occurs in the artery system, causing the heart and blood circulation to not function properly.

Diseases related to the heart and blood vessels include heart failure, coronary heart disease, and rheumatic heart [6]. Coronary heart disease is a disease of the heart and blood vessels caused by the narrowing of the coronary arteries. The narrowing of blood vessels occurs due to the process of atherosclerosis or spasm or a combination of both. Atherosclerosis occurs due to the accumulation of cholesterol and connective tissue in the blood vessels slowly, this is often characterized by complaints of chest pain. When the heart has to work harder, there is an imbalance between oxygen demand and intake, which is what causes chest pain. If the artery is completely blocked, the blood supply to the heart will stop, this is called a heart attack.

B. Data Mining

Data mining is the process of tracing the latest knowledge, patterns and trends selected from large amounts of data and stored in a repository or storage area using pattern recognition techniques as well as statistical and mathematical techniques [10]. Data mining became known as Knowledge-discovery in Database, this is an activity that includes the collection and use of historical data to solve regular patterns or relationships in large data sets. The output of this data mining is can be used to improve future decision-making based on information obtained from past data [11].

C. Naive Bayes

The Naive Bayes was proposed by the British scientist Thomas Bayes. Naive Bayes predicts future opportunities based on past experiences. Naive Bayes is considered to have good potential in classifying documents compared to other classification methods in terms of accuracy [13]. Naive is the basis of Bayes statistics and can be applied in many fields such as science, engineering, economics, medical game theory, law, and so on.

Bayes rule is used to calculate the probability of a class. The Naive Bayes algorithm provides a way to combine previous probabilities with possible conditions into a new formula that is used to calculate the probability of each possibility that will occur. The general form of Bayes' theorem is as follows:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad (2.1)$$

Where:

X : datawith an unknown class
H : Data hypothesis X is a specific class
P(H|X) : *Probability*hypothesis H based on condition X
(posterior probability)
P(H) : *Probability*hypothesis H (prior probability)
P(X|H) : *Probability*X based on the conditions on the
hypothesis H
P(X) : *Probability*from X

Naive bayes is a simplification of the Bayes method. Bayes' theorem after simplification becomes:

$$P(H|X) = P(X|H)P(X) \qquad (2.2)$$

Bayes rule is applied to calculate the posterior and probability from the previous data. In Bayesian analysis, the final classification is generated by combining the two sources of information, namely prior information and posterior information to generate probabilities using Bayes' rules.

D. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is often used in research because Particle Swarm Optimization has similar properties to Genetic Algorithm (GA). The advantage of PSO is that it is easy to implement and there are several parameters to adjust it [6]. Particle Swarm Optimization is a population-based stochastic optimization technique proposed in 1995 by Kennedy and Eberhart. The development of PSO is based on the metaphor of social interaction and communication of the movement of a flock of birds or fish (bird flocking or fish schooling) [14]. In the Particle Swarm Optimization algorithm, the search for a solution is carried out by a population consisting of several particles. The article looks for the optimal solution by traversing the search space using the related particles making adjustments to the best position of each particle or called the local best and the best particle position from the entire swarm or called the global best while traversing the search space. Thus the information dissemination occurs within the particle itself and between a particle and the best particle of the entire swarm during the search for a solution. After that, a search process is carried out to find the best position of each particle in a certain number of iterations until a steady relative position is obtained or reaches a predetermined iteration limit. In each iteration (t), each solution is represented by particle I, its performance is evaluated by entering the solution into the fitness function value.

Each particle requires a point in a certain dimension of space then 2 factors characterize the status of the particle in the search space, namely at the particle position (X) and at the particle velocity (Y). The mathematical formula that describes the position and speed of particles in a certain dimensional space is as follows:

$$X_i(t) = X_{i1}(t), X_{i2}(t), \ldots\ldots, X_{in}(t) \quad (2.3)$$

$$V_i(t) = V_{i1}(t), V_{i2}(t), \ldots\ldots, V_{in}(t) \quad (2.4)$$

The above equation (2.4) is used to describe the new particle velocity based on the velocity at the previous velocity, the distance between the current position and the local best, and the distance between the current position and the global best position. Then the particle flies to a new position according to the equation (2.5).

$$V_i(t) = v_{i1}(t-1) + c_1 r_1 (X_i^L - X_i(t-1) + c_2 r_2 (X_i^G - X_i(t-1)) \quad (2.5)$$

$$X_{i(t)} = v_i(t) + X_i(t-1) \quad (2.6)$$

Where :

$V_i(t)$: the velocity of the i-th particle in the i-th iteration

$X_{i(t)}$: the current position of the i-th particle in the i-th iteration

t : iteration

X_i^L : local best of the i-th particle

X_i^G : global best of the whole flock

$c_1 c_2$: constant acceleration or learning rate

$r_1 r_2$: a random or random number with a value between 0 up to 1

III. METHODOLOGY

At this stage the researcher will conduct a literature study using the Naïve Bayes classification method and Particle Swarm Optimization (PSO). This method is the application of machine learning that is used to classify data. In testing this study using parameters, namely: accuracy, precision, recall and time later resulting from the application of the Naïve Bayes method and Particle Swarm Optimization (PSO) in the classification of heart attack disease.

At this stage, the flow contains a sequence of process diagrams from the research flow in detail and detail which includes; algorithms, modelling, and design of the system. Below is shown Figure 3.1 as a research flow classification of heart disease.

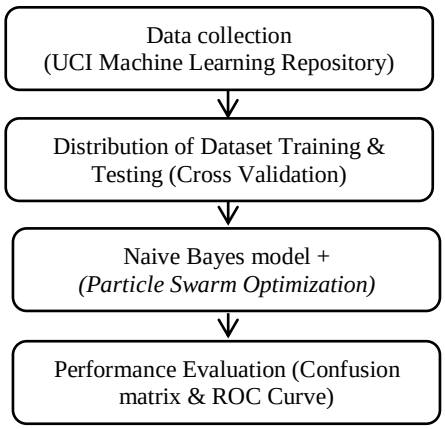


Figure 1. Research Method Design

Based on the diagram above, the research starts from data collection which is the data used is open source data or public data. After that, the next step is data analysis, namely data preprocessing. Then tested the prediction model of the naive Bayes algorithm using rapidminer tools. The next stage is the distribution of the dataset between training data and testing data using cross-validation which is optimized

with the Particle Swarm Optimization (PSO) feature. After that, the performance evaluation was carried out using the confusion matrix and the ROC curve. And the last stage is the selection of the most effective algorithm.

A. Data collection

In this study, the data source is a dataset obtained from the UCI Machine Learning Repository with the name heart disease. Here is the URL link to the dataset.https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/, there are 303 data on patients with heart disease and 14 attributes including age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, num, where one attribute is used as the target attribute. Each of these attributes can be seen in table 3.1.

Table 3.1 Attributes of the Heart Disease Dataset

Attribute	Information
(Age)	Age,[29-77]
(Sex)	gender (1 = male; 0 = female)
(cp)	Type of chest pain: a. Grade 1: typical angina, b. Grade 2: atypical angina, c. Grade 3: non-angina pain, d. Grade 4: no symptoms
(trestbps)	Tresting blood pressure (when the heart muscle is resting) (in mm Hg on admission to the hospital),[94-200]
(chol)	serum cholesterol (the total amount of cholesterol in the blood) in m/dl[126--564]
(fbs)	(fasting blood sugar / before eating > 120m/dl) (1 = true; 0 = false)
(restecg)	resting results of electrocardiography (heart muscle examination tool): a. Value 0: normal, b. Grade 1: have ST-T wave abnormalities (T and/or ST wave inversion, c. Grade 2: left ventricular hypertrophy by Estes' criteria
(thalach)	maximum heart rate is reached[71-202]
(exang)	angina-induced exercise (1 = yes; 0 = no)
(oldpeak)	ST depression caused by exercise relative to rest,[0--6.2]
(slope)	the slope of the ST segment peak exercise: a. Grade 1: uphill, b. Value 2: flat, c. Value 3: downsloping
(ca)	number of blood vessels (0-3) described by fluoroscopy
(thal)	3 = normal; 6 = permanent disability; 7 = reversible defect
(num)	(predicted attribute) diagnosis of heart disease (angiographic disease status) a. Value 0: <50% diameter narrowing b. Value 1:> 50% diameter narrowing

B. Dataset Sharing

The next stage after analyzing the data is the distribution of the dataset. The dataset is divided into training data and testing data. The function of the training data is to get the prediction results while the testing data is used to see the performance of the resulting prediction model. In this study, the training data and testing data were tested using the cross-validation technique (k fold) to obtain maximum accuracy results.

C. Nave Bayes Model Test UsingParticle Swarm Optimization (PSO)

At this stage the prediction model test uses the Naïve Bayes algorithm and Particle Swarm Optimization (PSO), the dataset between training data and testing data uses cross validation which is optimized with the Particle Swarm Optimization (PSO) feature. rapidminer tools.

D. Performance Evaluation

The next stage is the performance evaluation stage using the Confusion Matrix and ROC curve. Next, the most effective algorithm is selected based on the results using cross validation which is optimized with the Particle Swarm Optimization (PSO) feature.

IV. Result And Discussion

The results of the research will be divided into two experiments, the first is investigation is the Naive Bayes algorithm with a cross-validation technique without feature selection, and the second investigation is the Naive Bayes algorithm with a cross-validation technique with Particle Swarm Optimization (PSO) feature selection optimization using the rapidminer tool.

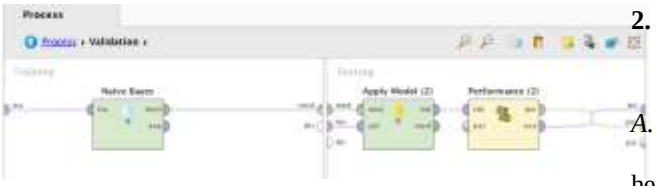
1. First Test of Cross Validation Technique

A. Operator Stage

For the completion of the heart disease dataset or heart disease classification, it is shown in the following figure.

Picture 2. Completion Stage

The first process is to read the disease dataset file or heart disease classification in CSV format, then connect it to the Cross Validation operator. This experiment was carried out on a dataset using rapidminer tools with 10 validations, the next step can be seen in the image below.

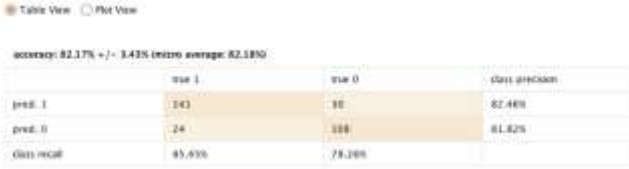


Picture 3. Second Process Stage

To be able to display this stage by clicking 2 x on the Cross Validation operator, then 2 windows will appear, the first is the training window, in this window fill it with the Naïve Bayes operator, then in the testing window fill it with the Apply model and Performance operator.

B. Accuracy Results

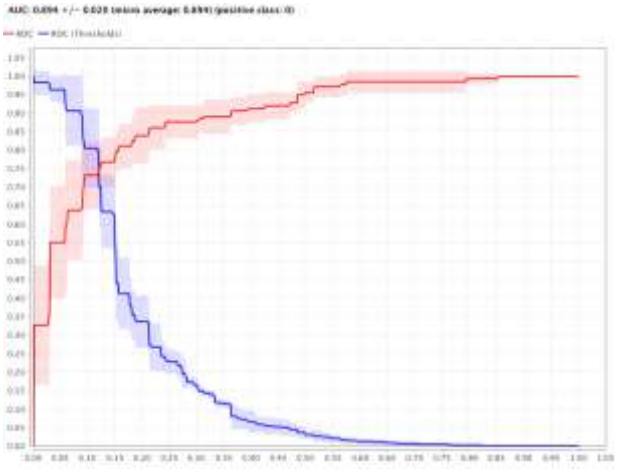
From processing a dataset of 303 records from 14 attributes, the results obtained are:



Picture 4. Accuracy Results

The results from the calculation of accuracy are 82.17%, and the Class Precision results from prediction 1 are 82.46%, and the Class Precision results from prediction 0 are 81.82%. For Class Recall results from a True 1 value of 85.45% and for a Class Recall value of a True 0 value of 78.26%.

C. AUC Results



Picture 5. AUC Graphic

The picture above is a graphic image of the results of the AUC calculation.

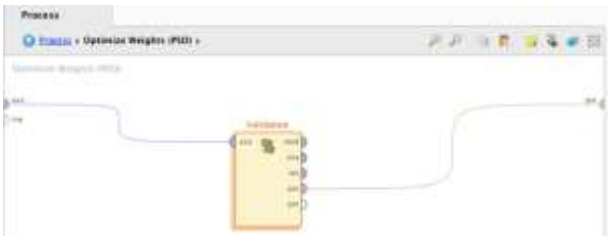
2. Second Test of Cross Validation Technique with Particle Swarm Optimization (PSO) Feature Selection Optimization

A. Operator Stage

For the completion of the heart disease dataset or heart disease classification with the Naïve Bayes algorithm using the Particle Swarm Optimization (PSO) feature selection, as shown in the following figure.

Picture 10. Completion Stage

The first process is to read the disease dataset file or heart disease classification in CSV format, then connect it to the Particle Swarm Optimization (PSO) feature selection. This experiment was carried out on a dataset using rapidminer tools with a Cross Validation operator, the next step can be seen in the image below.



Picture 6. Stages of the Second Process *Cross Validation*

The second process is the Particle Swarm Optimization (PSO) feature selection process which is connected to the Cross Validation operator. This experiment was carried out on a dataset using 10 validations, the next step can be seen in the image below.

Picture 7. Stages of the Third Process

To be able to display this stage by clicking 2x on the Validation operator, then 2 windows will appear, the first is the training window, in this window fill it with the Naïve Bayes operator, then in the testing window fill it with the Apply model and Performance operator.

B. Accuracy Results

From processing a dataset of 303 records from 14 attributes, the results obtained are:

Picture 8. Accuracy Results

The results of the calculation accuracy are 84.16%, and the Class Precision results from prediction 1 are 85.03%, and the Class Precision results from prediction 0 are 83.09%. For Class Recall results from a True 1 value of 86.06% and for a Class Recall value of a True 0 value of 81.88%.

C. AUC Results

Picture 9. AUC Graphic

The picture above is a graphic image of the results of the AUC calculation.

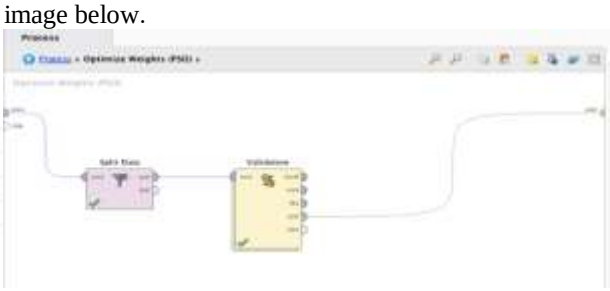
3. Second Test of Cross Validation Technique With Particle Swarm Optimization (PSO) Feature Selection Optimization

A. Operator Stage

For solution *heart disease dataset* or heart disease classification with algorithm *Naive Bayes* use feature selection *Particle Swarm Optimization (PSO)*, is shown by the following figure.

Picture 10. Completion Stage

The first process is to read the disease dataset file or heart disease classification in CSV format, then connect it to the Particle Swarm Optimization (PSO) feature selection. This experiment was carried out on a dataset using rapidminer tools with a Cross Validation operator, the next stage can be seen in the image below.



Picture 11. Second Process Stage *Split Data + Cross Validation*

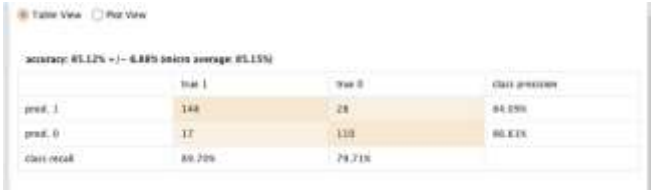
The second process is the Particle Swarm Optimization (PSO) feature selection process which is connected to the Split Data and Cross Validation operators. This experiment was conducted on a dataset using 10 validations, the next step can be seen in the image below.

Picture 12. Stages of the Third Process

To be able to display this stage by clicking 2x on the Validation operator, then 2 windows will appear, the first is the training window, in this window fill it with the Naïve Bayes operator, then in the testing window fill it with the Apply model and Performance operator.

B. *Accuracy Results*

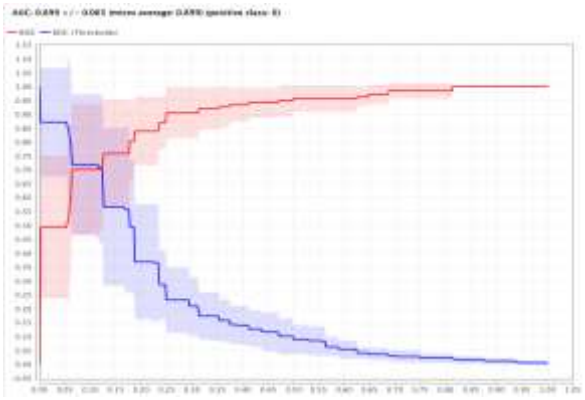
From processing a dataset of 303 records from 14 attributes, the results obtained are:



Picture 13. Accuracy Results

The results of the calculation accuracy are 85.12%, and the Class Precision results from prediction 1 are 84.09%, and the Class Precision results from prediction 0 are 86.61%. For Class Recall results from a True 1 value of 89.70% and for a Class Recall value of a True 0 value of 79.71%..

C. AUC Results



Picture 14. AUC Graphic

The picture above is a graphic image of the results of the AUC calculation.

D. Experiment Results

At this stage, the prediction model test using the Naïve Bayes algorithm is tested with a cross-validation technique and then optimization is done using the Particle Swarm Optimization (PSO) feature selection using rapidminer tools. The following are the results of the prediction model test which are presented in tabular form.

Table 4.1 Comparison of Performance Methods

Method	Method *Naive Bayes*	Accuracy	Precision	Recall	AUC
Previous researchers : without feature selection optimization (El Hamdaoui, H., Boujraf, S., Chaoui, NEH, & Maaroufi, M., 2020)	*Cross-Validation*	82.17%	-	-	-
	Cross-Validation + Split Data	84.28%	-	-	-
Current Researcher : *feature selection optimization* using method *Particle Swarm Optimization (PSO)*	*Cross-Validation*	82.17%	81.73%	78.25%	0.894
	Cross-Validation + PSO	84.16%	84.35%	81.98%	0.908
	Cross-Validation + Split Data + PSO	85.12%	87.12%	79.51%	0.899

From the table above, it can be seen that the optimization feature can increase the accuracy value. Where high accuracy is obtained in the Naïve Bayes algorithm using the cross-validation technique which is optimized with the Particle Swarm Optimization (PSO) feature with an accuracy value of 84.16% and the test using Split Data results in an accuracy value of 85.12%.

V. CONCLUSION

In this research, the optimization of the naive Bayes algorithm was done by using the feature selection or optimization method, where the previous researcher, namely the journal entitled A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques, had similarities, namely regarding the dataset and algorithm used, but had not implemented features. selection or optimization, so that applying the feature selection or optimization method, it is expected to increase the accuracy value of the model's performance.

Based on the results of experiments conducted on the nave Bayes algorithm using feature selection with the Particle Swarm Optimization (PSO) method, it shows an increase in accuracy from previous researchers who have not applied feature selection or optimization methods, where in previous researchers the accuracy obtained was 82.17%, while in this research The accuracy obtained is 84.16% and the test using Split Data results in an accuracy value of 85.12% with Good Classification criteria by maintaining 14 attributes.

REFERENCES

[1] Kemenkes RI, "Hasil Riset Kesehatan Dasar Tahun 2018," Kementrian Kesehat. RI, vol. 53, no. 9, pp. 1689–1699, 2018.

[2] Santoso, M., & Setiawan, T. (2005). Penyakit Jantung Koroner. Cermin Dunia Kedokteran, 147, 5-9.

[3] Majid, A. (2007). Penyakit jantung Koroner: Patofisiologi, pencegahan dan pengobatan terkini. Universitas Sumatera Utara.

[4] Mutiara, E., No, J. K. R., Barat, R., & Cengkareng, J. B. (2020). Algoritma Klasifikasi Naive Bayes Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Tuberculosis (Tb).

[5] El Hamdaoui, H., Boujraf, S., Chaoui, N. E. H., & Maaroufi, M. (2020, September). A clinical support system for prediction of heart disease using machine learning techniques. In 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-5). IEEE.

[6] Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi data mining naïve bayes berbasis Particle Swarm Optimization untuk deteksi penyakit jantung. Pseudocode, 1(1), 11-14.

[7] Bianto, M. A., Kusrini, K., & Sudarmawan, S. (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes. Creative Information Technology Journal, 6(1), 75-83.

[8] Maulana, D., & Yahya, R. (2019). Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Penderita Penyakit Jantung Di Indonesia Menggunakan Rapid Miner. Jurnal SIGMA, 10(2), 191-197.

[9] Maimon, O., & Rokach, L. (2009). Introduction to knowledge discovery and data mining. In Data mining and knowledge discovery handbook (pp. 1-15). Springer, Boston, MA.

[10] Ardiyansyah, A., Rahayuningsih, P. A., & Maulana, R. (2018). Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. Jurnal Khatulistiwa Informatika, 6(1).

[11] Setyo, J. S., & Sudradjat, A. (2017). Penerapan Metode C4. 5 Terhadap Penyakit Tuberkulosis Paru. Jurnal Kajian Ilmiah, 17(3), 111-118.

[12] Gorunescu, F. (2011). Data Mining: Concepts, models and techniques (Vol. 12). Springer Science & Business Media.

[13] [Dewi, S. (2016). Komparasi 5 metode algoritma klasifikasi data mining pada prediksi keberhasilan pemasaran produk layanan perbankan. Techno Nusa Mandiri: Journal of Computing and Information Technology, 13(1), 60-65.

[14] Delice, Y., Kızılkaya Aydoğan, E., Özcan, U., & İlkay, M. S. (2017). A modified particle swarm optimization algorithm to mixed-model two-sided assembly line balancing. Journal of Intelligent Manufacturing, 28(1), 23-36.

[15] Wajhillah, R. (2014). Optimasi Algoritma Klasifikasi C4. 5 Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Jantung. Swabumi, 1(1), 26-36

[16] Nugroho, D., & Murdiansyah, D. T. (2016). Prediksi Penyakit Menggunakan Genetic Algorithm (GA) dan Naive Bayes untuk Data Berdimensi Tinggi. eProceedings of Engineering, 3(2).

[17] Ramanda, K. (2015). Penerapan Particle Swarm Optimization Sebagai Seleksi Fitur Prediksi Kelahiran Prematur Pada Algoritma Neural Network. Jurnal Teknik Komputer AMIK BSI, 1(2), 178-183.

[18] Nusa, D. C. P. B. S. (2016). Optimasi algoritma naïve bayes dengan menggunakan algoritma genetika untuk prediksi kesuburan (fertility). EVOLUSI: Jurnal Sains dan Manajemen, 4(1).

[19] Aronson, J. E., Liang, T. P., & MacCarthy, R. V. (2005). Decision support systems and intelligent systems (Vol. 4). Upper Saddle River, NJ, USA:: Pearson Prentice-Hall.

[20] Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4. 5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS/BMT Al-Fadhila. Jurnal Teknologi Informasi Magister, 2(01), 52-62.

[21] Annisa, R. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. JTIK (Jurnal Teknik Informatika Kaputama), 3(1), 22-28.

[22] Lestari, M. E. I. (2015). Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung. Faktor Exacta, 7(4), 366-371.