PREDICTION AND ANALYSIS OF CARDIOVASCULAR DISEASE WITH NEURAL NETWORK ALGORITHM

Ardi Ramadhan Sukma[1]; Riqadri Halfis[2]

Postgraduate Masters Program in Computer Science
STMIK Nusa Mandiri, Jakarta, Indonesia
http://nusamandiri.ac.id/
ardirsukma40@gmail.com[1], arizhboys@gmail.com[2]

Abstract— *In the medical and health world it is very necessary to predict one of his cardiovascular diseases in patients. On this occasion, there were 220 datasets used in the study and 220 attributes. Prediction is done by using a neural network algorithm. Then do a calculation using Rapidminer from the cardiovascular disease dataset. The dataset will be predicted and analyzed with two trials at Rapidminer to determine the optimal level of prediction. The results of the first trial at Rapidminer are 68.81% accuracy with a prescription value of 65.85%, 79.41% recall and, AUC value of 0.745 and the second trial result is 77.27% accuracy, with a precision value of 77.14%, 79.41% recall and, AUC value of 0.764.*

Keywords: Cardiovascular, Machine Learning, Neural Network, Rapidminer.

Intisari—*Dalam dunia medis dan kesehatan sangat perlu untuk melakukan prediksi salah satu nya penyakit kardiovaskular pada pasien. Pada kesempatan kali ini digunakan berupa dataset di dalam penelitian sebanyak 220 record dan beberapa atribut. Prediksi dilakukan dengan menggunkan algoritma neural network. Lalu melakukan perhitungan menggunakan Rapidminer dari dataset penyakit kardiovaskular. Dataset tersebut akan di prediksi dan analisis dengan dua uji coba di Rapidminer untuk menentukan tingakt optimal dari prediksi. Hasil dari uji coba pertama di Rapidminer adalah akurasi 68.81% dengan nilai presese 65.85%, recall 79.41%, dan nilai AUC 0.745 dan hasil uji coba kedua adalah akurasi 77.27%, dengan nilai presisi 77.14%, recall 79.41%, dan nilai AUC 0.764.*

Kata Kunci*: Cardiovascular, Machine Learning, Neural Network, Rapidminer.*

INTRODUCTION

In the medical and health world it is very necessary to know the predictions of an illness and effective decisions in making decisions and analysis of patients who are sick.(Kementrian Kesehatan RI, 2014). According to the cardiovascular definition from WHO, cardiovascular disease is a disease caused by impaired heart and blood vessel function. There are many kinds of cardiovascular disease, but the most common and most well-known are coronary heart disease and stroke. (Kementrian Kesehatan RI, 2014)

Many disease prediction methods are done using Genetic Algorithms (Ratnakar, Rajeswari, & Jacob, 2013), *Naïve Bayes and Decision Tree* (Palaniappan & Awang, 2008), *Neural Network, Multilayer Perceptron*(Rodrigo & Tsokos, 2017). In this study, testing will be performed using neural network algorithms to predict cardiovascular disease in patients who experience several attributes. Neural Network algorithm is an attempt to mimic the function of the human brain. The human brain is believed to consist of millions of small processing units called neurons, which work in parallel. Neurons are connected to each other through neuronal connections. Each individual neuron takes input from a set of neurons. Then it processes the input and passes the output to a set of neurons. The output is collected by other neurons for further processing. (Shukla, 2010). The purpose of this study is to accurately predict cardiovascular disease with the label "0" or "1" or to have a cardiovascular disease "yes" or "no".

MATERIALS AND METHODS

In this study the authors used the collection method:

1. Data Collection

Gather data in the form of a dataset that you want to use in a method later.

2. Data Processing

Processing the dataset in the form of subtracting some attributes.

3. The method used

The method used is the neural network algorithm which is part of deep learning. Deep Learning is part of Machine Learning.

4. Testing Data with the method used

Test the cardiovascular disease dataset with a neural network algorithm using the Rapidminer application.

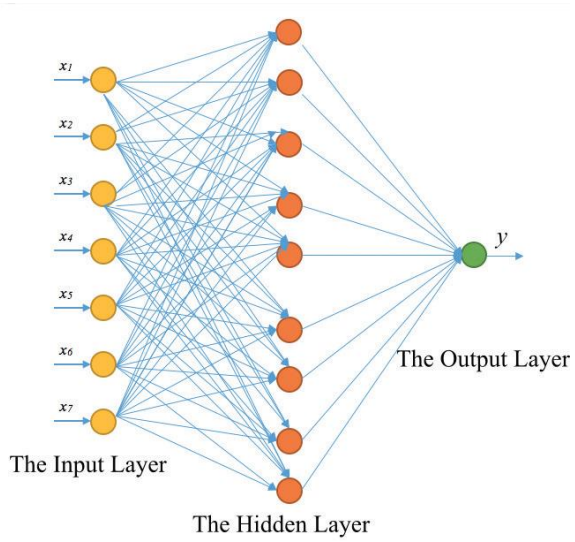
5. Testing and Validation Results

After testing the meal in the prediction validation results can be from cardiovascular disease.

Machine Learning is one branch of Artificial Intelligence that focuses on developing a system based on data. Machine Learning has four main things that are learned namely directed learning (supervised learning), unsupervised learning (unsupervised learning), semi-directed learning (semi-supervised learning), reinforcement learning.(Rodrigo & Tsokos, 2017)

The problem taken from the study is how to implement neural network algorithm to predict disease accuracy, precision, recall, AUC in the dataset which can be https://www.kaggle.com/sulianova/cardiovascular-disease-dataset that is a dataset in the form of disease data Cardiovascular that will be predicted with age, gender, height, weight, an_hi, an_lo, cholesterol, gluc, smok, alco, active, and cardiovascular patterns. The label on the dataset is in cardiovascular ie "0" or "1" can also be called a patient having cardiovascular disease or not. These labels will test the accuracy of cardiovascular disease of a patient with Rapidminer.

Neural Network Algorithm is an attempt to mimic the function of the human brain. The human brain is believed to consist of millions of tiny processing units called neurons, which work in parallel. Neurons are connected to each other through neuronal connections. Each individual neuron takes input from a set of neurons. Then it processes the input and passes the output to a set of neurons. Output collected by other neurons for further processing. (Shukla, 2010)



Source:(Prathama, Aminullah, Saputra, Teknik, & Mada, 2017)

Image 1. Neural Network Architecture

The architecture above is usually referred to as Multi-Layer Perceptron (MLP) or Fully-Connected Layer. The first architecture has 7 neurons in the Input Layer and 1 Output Layer. Between Input and Output, there is a Hidden Layer with 9 neurons. This research uses the dataset with 220 records, with 12 attributes. Table 1 below is the dataset used.

Table 1. Cardiovascular Dataset

Id	Age	Gender	Height	Weight	Ap_hi	Ap_lo	cholesterol	Gluc	Smoke	Alco	Active	cardio
1	20228	1	156	85	140	90	3	1	0	0	1	1
2	18857	1	165	64	130	70	3	1	0	0	0	1
3	17623	2	169	82	150	100	1	1	0	0	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...
299	21179	1	164	67	120	80	1	1	0	0	1	1

Source: (Sukma & Halfis, 2019)

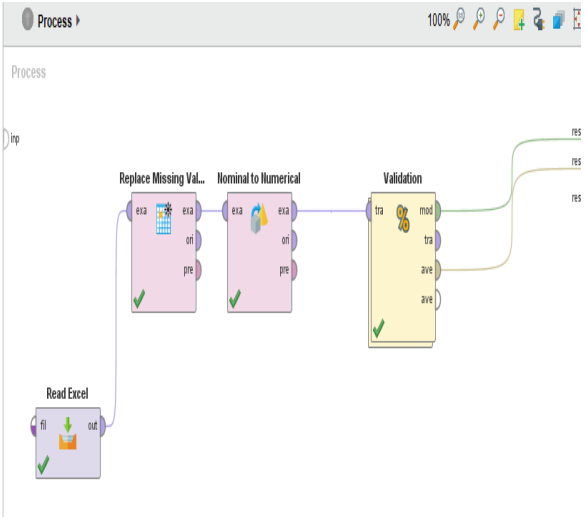
From the dataset above, the output from the cardio attributes "0" and "1" will be used to predict whether the patient is sick or not.

The research will use Microsoft Excel 2010 as a dataset and use the Rapidminer application as a predictive measurement tool.

RESULTS AND DISCUSSION

Implementation with Rapidminer

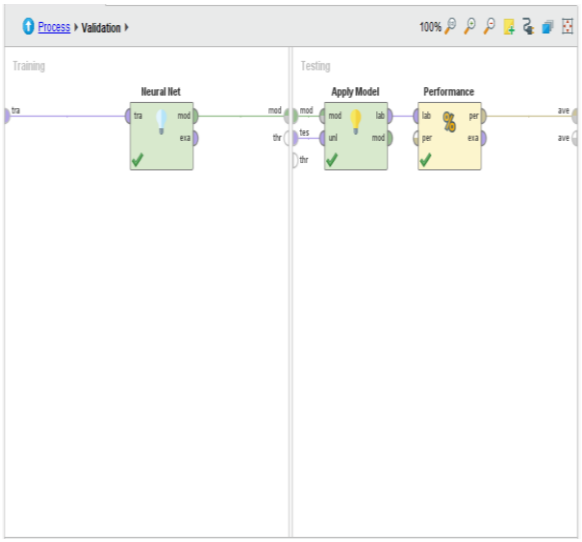
In this calculation, the prediction measurements will be made from the Dataset using the Neural Network Aglorithmetic using the Rapidminer tool. The following is the neural network algorithm model in the Rapidminer framework:



Source: (Sukma & Halfis, 2019)

Figure 2. Main Process Rapidminer

In the test process in Rapidminer, the first dataset is inputted to Read Excel and then the process of replacing the missing value, nominal to numeric, and cross-validation is carried out. The replace missing value process is used when data contains duplication. After doing the process, the nominal to numeric process is continued to change the non-numeric attribute type to the numeric type and also map all attribute values to the numeric value where the attribute values are 0 and 1. Then for the test, cross-validation is performed.



Source: (Sukma & Halfis, 2019)

Picture3. Neural Network Validation Process

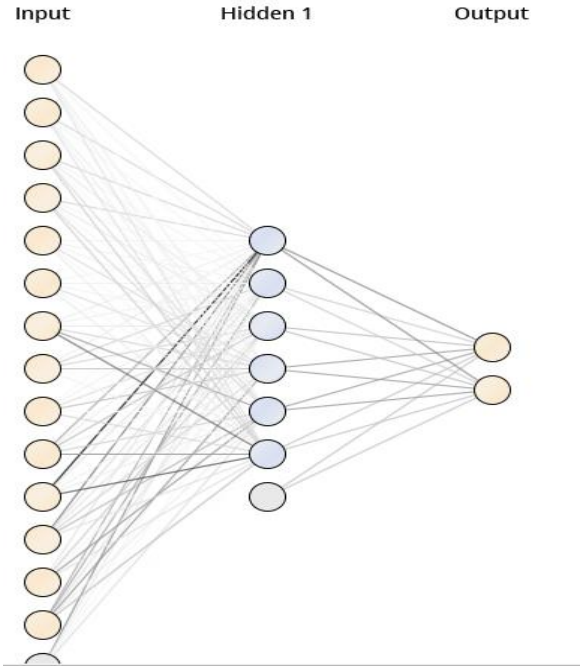
In the cross-validation process there is a neural network operation where the first try and the second try with a training cycle, learning rate, and momentum different from the output in the form of neural network architecture. Apply Model and Performance are used to get accuracy, precision, recall, and AUC values.

The neural network algorithm model is carried out with a training model process by providing 2 trials i.e.:

a. First Trial
   1. *Training cycles* : 200
   2. *Learing rate* : 0.5
   3. *Momentum* : 0.1

b. Second trial
   1. *Training cycles* : 1500
   2. *Learing rate* : 0.5
   3. *Momentum* : 0.8

From the first trial research above, it can be a neural network architecture by producing a hidden layer with two outputs:



Source: (Sukma & Halfis, 2019)

Figure 4. First Trial Cardiovascular Neural Network Architecture

In the first trial, the neural network architecture can be in the form of Input and Output Layer. Between Input and Output, there is a Hidden Layer. Input Layer has 14 neurons, 1 bias and Output Layer has 2 neurons of which there are 6 hidden layers and 1 bias. Each has a different value from the first trial training cycle: 200, learning rate: 0.5, and momentum: 0.1. With the value of each of the neural network algorithm nodes obtained from the Rapidminer process:

Table 2. First Layer Hidden Layer Input Node Values

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Smoke = 0	-2.759	0.693	0.137	-1.129	0.958	1.880
Smoke =1	2.763	-0.715	-0.156	1.086	-0.952	-1.953
Alco = 0	2.899	-0.757	-1.469	0.489	-2.280	3.089
Alco =1	-2.855	0.702	1.407	-0.484	2.366	-3.123
Active = 0	-0.640	0.898	-0.835	-2.227	-0.497	-1.271
Active = 1	0.656	-0.820	0.761	2.264	0.497	1.318
Age	1.126	-0.994	-1.208	-2.752	-6.178	9.533
Gender	-1.150	-3.111	-2.690	-3.603	-0.241	0.751
Height	-0.437	0.993	2.359	4.055	-0.263	-2.672
Weight	5.823	-1.187	-4.275	1.057	-2.991	6.300
Ap_hi	16.375	1.782	0.096	-0.901	-1.153	10.751
Ap_lo	8.813	-2.228	-1.699	-2.864	-4731	2.034
Cholesterol	-1.792	0.114	3.328	0.632	6.720	-2.059
Gluc	-5.518	5.823	1.904	6.887	1.166	4.282
Bias	-7.424	-1.476	-0.771	-0.509	-2.155	-0.216

Source: (Sukma & Halfis, 2019)

From the hidden layer node values the output obtained by the neural network algorithm is as follows:

Table 3. First Trial Output Node Values

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Threshold
Class = 0	-7.545	-3.261	-4.126	5.947	-6.199	-3.875	3.889
Class =1	7.544	3.260	4.169	-5.950	6.195	3.874	-3.889

Source: (Sukma & Halfis, 2019)

From the analysis of the analysis using the neural network algorithm the accuracy, prescription, recall, and AUC values were obtained. Here are the accuracy, precision, recall, and AUC values:

Table 4 First Test Neural Network Algorithm Value

	Neural Network
Accuracy	68.81%
Presicion	65.85%
Recall	79.41%
AUC	0.745

Source: (Sukma & Halfis, 2019)

accuracy: 68.18%

	true 0	true 1	class precision
pred. 0	18	7	72.00%
pred. 1	14	27	65.85%
class recall	56.25%	79.41%	

Source: (Sukma & Halfis, 2019)
Figure 5. Results of the First Test Accuracy Output

precision: 65.85% (positive class: 1)

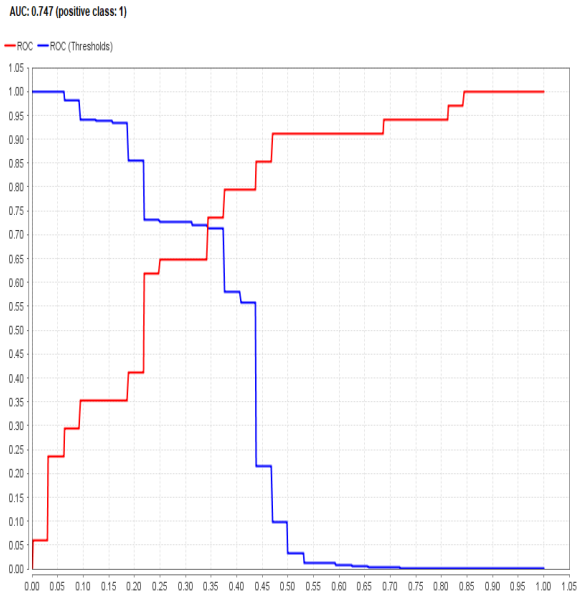
	true 0	true 1	class precision
pred. 0	18	7	72.00%
pred. 1	14	27	65.85%
class recall	56.25%	79.41%	

Source: (Sukma & Halfis, 2019)
Figure 6. First Test Precision Output Results

recall: 79.41% (positive class: 1)

	true 0	true 1	class precision
pred. 0	18	7	72.00%
pred. 1	14	27	65.85%
class recall	56.25%	79.41%	

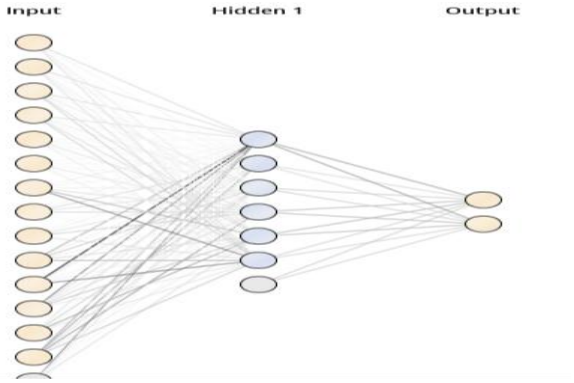
Source: (Sukma & Halfis, 2019)
Figure 7 First Test Recall Output Results



Source: (Sukma & Halfis, 2019)
Figure 8. Results of the First Test AUC Output

The following is the Second Trial conducted with different training cycles, learning rates, and

momentum:



Source: (Sukma & Halfis, 2019)

Figure 9 Cardiovascular Neural Network Architecture Second Trial

In the second trial, the neural network architecture can be in the form of Input and Output Layer. Between Input and Output, there is a Hidden Layer. Input Layer has 14 neurons, 1 bias and Output Layer has 2 neurons of which there are 6 hidden layers and 1 bias. Each has a different value with the second trial of the training cycle: 1500, learning rate: 0.5, and momentum: 0.8.

From the results of the second trial we get the nodes of the 14 existing neurons as follows:

Table 5. The value of the Second Layer Hidden Layer Input Node

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Smoke = 0	3.704	4.052	-2.852	-8.311	2.009	8.306
Smoke =1	-3.700	-4.704	2.833	8.268	-2.004	-8.380
Alco = 0	-7.830	-3.558	-2.499	1.288	-11.217	-7.216
Alco =1	7.874	3.503	2.436	-1.282	11.303	7.182
Active = 0	-0.526	5.030	-1.314	2.325	-3.533	-5.223
Active = 1	0.542	-4.953	1.241	-2.288	3.533	5.270
Age	-12.285	-3.528	0.210	14.758	-6.184	27.720
Gender	3.859	-6.174	0.823	-23.025	8.384	18.956
Height	-5.499	-2.699	4.469	9.010	0.589	-34.999
Weight	-9.425	-6.989	-18.075	-23.025	-17.029	-14.849
Ap_hi	0.205	11.948	62.911	-15.314	13.218	13.441
Ap_lo	-8.907	0.439	-1.733	41.201	16.183	7.438
Cholesterol	23.376	5.041	9.582	-22.617	3.698	-20.498
Gluc	4.381	18.612	-18.661	-15.600	21.012	31.701
Bias	-3.187	-9.873	-15.415	-32.010	21.596	-11.787

Source: (Sukma & Halfis, 2019)

From the hidden layer node values the output obtained by the neural network algorithm is as follows:

Table 6. First Trial Output Node Values

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Thereshold
Class = 0	-15.624	-6.831	-15.133	-3.828	8.763	-4.108	2.529
Class =1	15.624	6.831	15.133	3.828	-8.763	4.108	-2.529

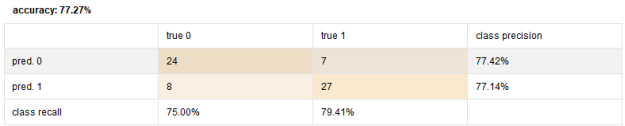
Source: (Sukma & Halfis, 2019)

From the analysis of testing with the neural network algorithm the accuracy, precision, recall, and AUC values are obtained. Here are the accuracy, precision, recall, and AUC values:

Table 7. Value of the Second Trial Neural Network Algorithm

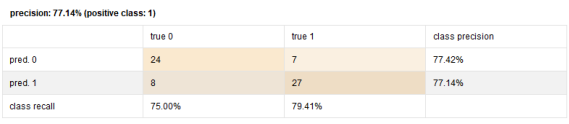
	Neural Network
Accuracy	77.27%
Presicion	77.14%
Recall	79.41%
AUC	0.764

Source: (Sukma & Halfis, 2019)



Source: (Sukma & Halfis, 2019)

Figure 10. Results of Accuracy Output of Second Trial



Source: (Sukma & Halfis, 2019)

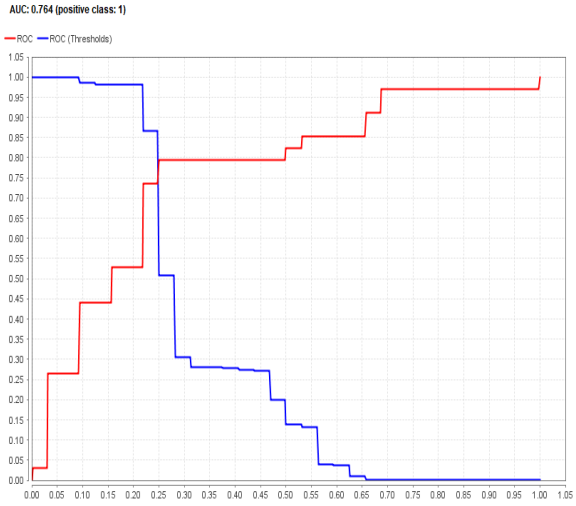
Figure 11. Results of the Second Precision Trial Output

recall: 79.41% (positive class: 1)

	true 0	true 1	class precision
pred. 0	24	7	77.42%
pred. 1	8	27	77.14%
class recall	75.00%	79.41%	

Source: (Sukma & Halfis, 2019)

Figure 12. Second Test Outcome Recall Results



Source: (Sukma & Halfis, 2019)

Figure 13 AUC Outcome of Second Trial Output

The results of the study for the application of cardiovascular disease prediction using the network algorithm with 2 trials where the first trial obtained 68.81% accuracy with precision value of 65.85%, recall 79.41%, and, AUC value of 0.745 and the second trial obtained accuracy of 77.27% with precision value 77.14 %, 79.41% recall, and, AUC value of 0.764.

CONCLUSION

The neural network algorithm is used in this study to predict cardiovascular disease based on several attributes using the Rapidminer application to test the dataset where the results of the prediction of cardiovascular disease prediction are compared with two trials. The results of the Rapidminer test were obtained in the first trial in the form of an accuracy of 68.81% with a precision value of 65.85%, a recall of 79.41%, and an AUC value of 0.745 and in the second trial an accuracy of 77.27% with a precision value of 77.14%, a recall of 79.41%, and an AUC value 0.764. Thus the results of testing with neural network algorithms are still not optimal and are expected to use other algorithms. The results of the study were prediction and analysis of cardiovascular disease in patients with two trials in Rapidminer using a neural network algorithm. More optimal results in the second trial with different training cycles, learning rates, and momentum.

REFERENCES

Kementrian Kesehatan RI. (2014, September). Situasi Kesehatan Jantung. *Pusat Data Dan Informasi Kementrian Kesehatan RI*, 1–8. Retrieved from https://www.kemkes.go.id/resources/download/pusdatin/infodatin/infodatin-jantung.pdf

Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. *AICCSA 08 - 6th IEEE/ACS International Conference on Computer Systems and Applications*, (December), 108–115. https://doi.org/10.1109/AICCSA.2008.4493524

Prathama, A. Y., Aminullah, A., Saputra, A., Teknik, D., & Mada, U. G. (2017). UNTUK PENENTUAN PROSENTASE BOBOT PEKERJAAN PADA RUMAH SAKIT PRATAMA, *7*. https://doi.org/10.22146/teknosains.30139

Ratnakar, S., Rajeswari, K., & Jacob, R. (2013). Prediction of Heart Disease Using Genetic Algorithm. *International Journal of Advanced Computational Engineering and Neworking*, *1*(2), 51–55.

Rodrigo, H., & Tsokos, C. P. (2017). Artificial Neural Network Model for Predicting Lung Cancer Survival. *Journal of Data Analysis and Information Processing*, *05*(01), 33–47. https://doi.org/10.4236/jdaip.2017.51003

Shukla, A. (2010). *Real Life Applications of Soft Computing*. *Real Life Applications of Soft Computing*. https://doi.org/10.1201/ebk1439822876

Sukma, A. R., & Halfis, R. (2019). PREDIKSI DAN ANALISIS PENYAKIT KARDIOVASKULAR DENGAN, *16*(2), 1–6.