**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

# *OUTCOME PREDICTION FOR HEART DISEASE WITH ARTIFICIAL NEURAL NETWORK ALGORITHM*

**[1]Hambali Ahmad Aripin, [2]Nawzia Yasmin, [3]Yosan Dionisio Tamayo**
[1]Program Studi Manajemen Informatika, Politeknik Piksi Ganesha
[23]Computer Education Department, Centro Escolar University Philippines
Email : [1]hamham506@yahoo.com ; [2]yasmin@sub.edu.bd ; [3]jtamayo@ceu.edu.ph

***ABSTRACT***

*Heart disease is still the number one killer in the world, the latest discovery, this disease is the trigger of one third of all deaths in the world from year to year is always increasing. This study aims to help make predictions for heart disease early as well as create a model of prediction analysis (outcome prediction) obtained from health data (Healthcare). The method proposed in this study with deep learning techniques that apply artificial neural network algorithms with hidden layer techniques in making predictions of heart disease. In this stage of research, problems were found in improving the accuracy of datasets used by handling problems in pre-processing data such as missing data and determining the form of data correlation. The model that was then tested through a heart disease dataset resulted in 90 % accuracy and with Random Forest result in 85% accuracy. With the creation of this prediction model is expected in addition to helping to make predictions of diseases can also provide the next innovation in data science in the field of health*

***Keywords****: heart disease, outcome, prediction, artificial neural network*

## INTRODUCTION

### A. Problem Identification

Being in life or living in modern big cities impacts our health in many ways. This occurs because: 1) stress associated with urban living, 2) sedentary lifestyle due to working conditions and lack of time, 3) air pollution that occurs, and 4) the number of people living in poverty, the urban population is at increased risk for development of chronic health conditions. Air pollution that occurs can cause respiratory disease, and cancer, in fact, some premature deaths other than air pollution are caused by ischemic heart disease and stroke. There is an increasing percentage of the world's population facing the adverse health effects of urban living. In particular, according to the United Nations [1], 54% of the earth's population resides in urban areas, a percentage that is expected to reach 66% by 2050. It is becoming clear that the health of citizens should be an

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

important priority in everyday life. To that end, smart health services involve the use of e-health and m-health systems, smart and connected medical devices, and the implementation of policies that promote health, health, and well-being [1]. Diseases of civilization or so-called lifestyle diseases, such as diabetes, coronary heart disease and obesity. Heart disease is still the number one killer in the world.

According to recent findings, this disease is the cause of a third of all deaths in the world in 2019. The number of deaths also continues to grow. China had the highest number of deaths from heart disease in recent years. Other countries that are ranked below are India, Russia, the United States, and Indonesia. However, countries such as France, Peru and Japan have the lowest rates of death from heart disease, 6 times lower than in 1990. Such is the discovery of information throughout the last 30 years. In 2019, most deaths from heart disease were related to ischemic heart disease (a heart problem caused by narrowing of the arteries) and stroke. From the death rate from heart disease worldwide obtained in 2019 was 18.6 million, the proportion of men who died was 9.6 million people, on the other hand 8.9 million women [2]. The research topic to be researched is to explore and develop predictive analytics aimed at predicting a leading chronic disease: heart disease. Previous research using data mining techniques using the Naïve Bayes algorithm by Riani, Ade et al [3] in predicting heart disease resulted in an accuracy of 86% for the 303 datasets tested. Derisma [4] conducted a heart disease prediction study by comparing 3 (three) data mining algorithms, namely Naïve Bayes, Random Forest and Neural Network which resulted in an average accuracy of 83%. Alhamad, Apriyanto et al[5] conducted a research on prediction of heart disease using the Machine Learning method based on Ensemble – Weighted Vote which took into account the problem of Missing Value (MV), Data Validation (DV), Unbalanced Class (UC) and Noisy Data (ND) resulting in an accuracy of 85.21 %. The public dataset that researchers commonly use in creating highly accurate Machine Learning

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

models for predicting heart disease is in the UCI Machine Learning Repository.

**B. Research Objectives and Benefits**

1) Creating predictive analysis models obtained from health data (Healthcare);
2) Finding the right algorithm and implementation that is efficient and suitable for analysing predictions in the health sector.

## THEORY REVIEW

### 1) Outcome Prediction

The main purpose of research in the field of computational biomedical is to predict the disease suffered by the patient. The following are two differences from the type of outcome prediction [6], namely:

a) Static or one-time prediction: predicts only one form of output for example heart disease resulting in the absence or presence of heart disease.
b) Temporal outcome prediction: predicting by involving the future time or history data sequentially for example absence or presence of heart disease obtained from history data sequentially.

### 2) Machine Learning

Machine Learning [7] is defined as an automated process that extracts patterns from data and to create models that are used in predictive analytics applications. The prediction model is obtained from the results of the machine learning algorithm that processes the dataset used. Below is Figure 1 Machine Learning Model Learning.

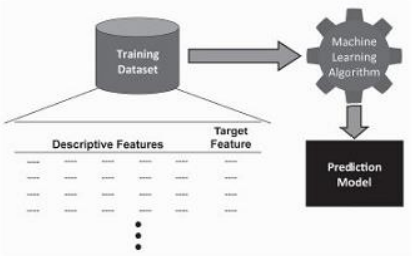



Figure 1 Machine Learning Model Learning [7]

Machine learning algorithms automate the process of studying models that capture the relationship between descriptive features and target features in a data set (dataset). One of the algorithms used in machine learning is Random Forest[8], [9] Random Forest can be described as a non-parametric model used for regression and classification cases.

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

The Random Forest algorithm is based on the bagging method which represents the concept of combining learning models to improve performance (higher accuracy or some other metric), below is Figure 2 Random Forest Simplified [10]

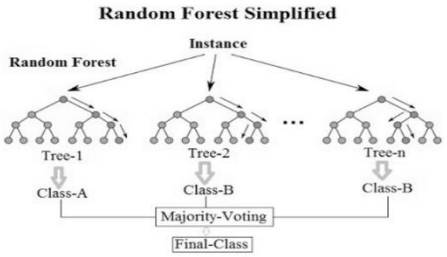



Figure 2 Random Forest Simplified

**3) Deep Learning**

Deep Learning [11] is a subfield of machine learning (Machine Learning) in artificial intelligence (Artificial Intelligence) in dealing with algorithms inspired by the biological structure and function of the brain to assist machines with intelligence.

This realization has led to a different order of fields based on the data. The new rule of thumb is that Machine Learning will not be able to improve performance by increasing training data after a certain threshold, while Deep Learning is able to utilize surplus data more effectively for performance improvement. The following chart is an illustration to represent the overall idea of model performance with data measures for the three areas.

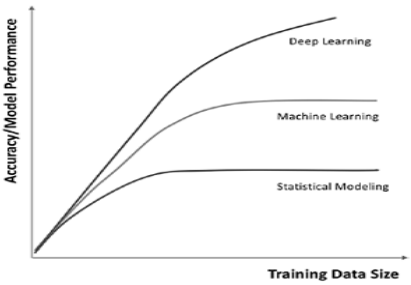



Figure 3. Performance Model [11]

Broadly speaking, the position of Deep Learning can be described as follows:

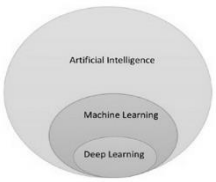



Figure 4. Deep Learning Position Diagram [11]

**4) Artificial Neural Network**

An artificial neural network [12] is a computational network (system of nodes and interconnections between nodes) inspired by biological networks, namely neural networks which are complex networks of neurons in the human brain. Nodes created in an Artificial Neural

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

Network are supposed to be programmed to behave like real neurons. Figure 3 and Figure 4 show the network of nodes (artificial neurons) that make up the artificial neural network.

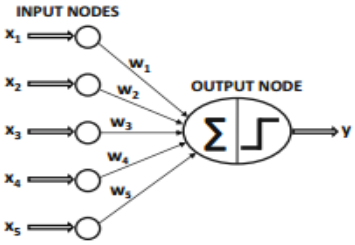



Figure 4. Perceptron without bias

With the following equation formula:

$$\hat{y} = sign\left\{\sum_{j=1}^{d} w_j\, x_j\right\}$$

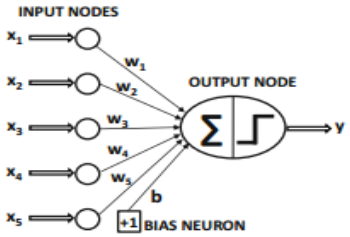



Figure 5. Perceptron with bias

With the following equation formula:

$$\hat{y} = sign\left\{\sum_{j=1}^{d} w_j\, x_j + b\right\}$$

**5) TensorFlow**

TensorFlow [11] has the unique ability to perform computational partial subgraphs thus enabling distributed training with the help of later neural networks. In other words, TensorFlow enables model parallelism and data parallelism.

**6) Keras**

Keras [11], [13] are compact and easy-to-learn high-level Python libraries for deep learning that can run on top of TensorFlow. This development focuses more on key deep learning concepts, such as creating layers for neural networks, TensorFlow is the back end for Keras. Keras is used for deep learning applications without interacting with complex TensorFlow. There are two main types of frameworks: sequential APIs and functional APIs. Sequential APIs are based on the notion of layer order; this is the most common use of Keras and the easiest part of Keras. Sequential models can be thought of as stacks of linear layers. There are 4 (four) parts of the deep learning deep learning model.

**7) Heart Disease**

Heart disease [14] is the formation of a disturbance in the balance between

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

blood supply and demand that occurs due to blockage of blood vessels. Deaths due to heart disease reached 959,227 patients, namely 41.4% of all deaths or every day 2600 people die from heart disease.

## METHOD

Clearer research steps can be seen in Figure 5 Outcome Prediction Model Methodology

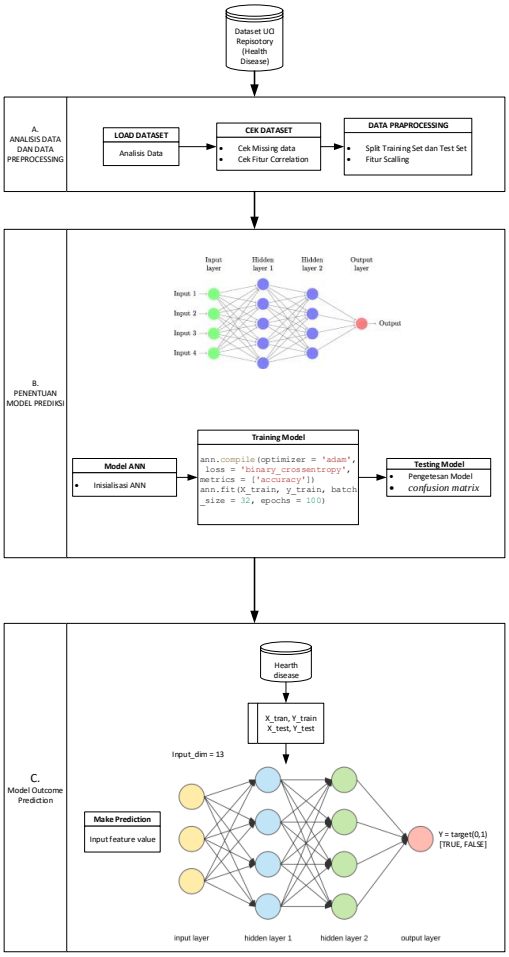



Figure 6. Outcome Prediction Model Methodology

### A. Data Analysis and Data Pre-processing

The data analysis stage analyses the dataset used to find out the names of the features and to find out the correlation between the features, the dataset used is obtained from the UCI Repository. Data pre-processing performs the Splitting dataset stage into Training sets and Test sets and performs Feature Scaling with the aim of checking variables that have very varied and random values so that at this stage the numerical data in the dataset has the same range of values (scale).

### B. Determination of Prediction Model

This stage performs model selection activities in producing outcome predictions using deep learning techniques, namely artificial neural networks (ANN). This activity performs ANN Initialization, Model Training and tests the model that has been made. This modelling scripting is assisted by the Keras library, a Python library intended for developing and evaluating deep learning models.

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

### C. Outcome Prediction Model

This stage produces a model for making predictions by testing the model that has been made by testing predictions by filling in the values of the dataset features which will produce True or False values that identify the value of absence or the presence of heart disease.

## DISCUSSION

A. Data Analysis

This stage explores data analysis of the dataset that is used to find out the names of the features and find out the correlation between the features that will be used to make predictions.

The identification of the features used can be seen from Table I below.

TABLE I

DESCRIPTION OF HEART UCI REPOSITORY DATASET

| Feature | Description |
|---|---|
| Age | age in years |
| Sex | sex (1 = male; 0 = female) |
| cp | The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic) |
| trestbps | The person's resting blood pressure (mm Hg on admission to the hospital) |
| Chol | The person's cholesterol measurement in mg/dl |
| fbs | The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false) |
| restecg | Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria) |
| thalach | The person's maximum heart rate achieved |
| exang | Exercise induced angina (1 = yes; 0 = no) |
| oldpeak | ST depression induced by exercise relative to rest |
| slope | the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping) |
| ca | The number of major vessels (0-3) |
| thal | A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect) |

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

| | |
|---|---|
| target(the predicted attribute) | Heart disease (0 = no, 1 = yes) |

Missing data in datasets originating from the EHR can be due to lack of collection or lack of documentation, so they need to be checked before making predictions. Data Cleaning process ensures data is correct, consistent, and usable. In cleaning data by identifying errors or damage, correcting or deleting them, or processing data manually if necessary to prevent errors in determining predictive variables.

The dataset used in this study uses an open dataset from the UCI Repository [15], using the data.shape() function from Python to find out the number of records and their features, namely 303 records and 14 features in the heart attack dataset. In checking the missing data using the data.info() function the results are as shown in Figure 7 below:

```
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Age                 303 non-null    int64
 1   Sex                 303 non-null    int64
 2   Chest Pain Type     303 non-null    int64
 3   Rest BP             303 non-null    int64
 4   Cholestrol          303 non-null    int64
 5   FBS                 303 non-null    int64
 6   RestECG             303 non-null    int64
 7   Max Heart Rate      303 non-null    int64
 8   Exer Angina         303 non-null    int64
 9   Prev Peak           303 non-null    float64
 10  Slope               303 non-null    int64
 11  No of Major Vessels 303 non-null    int64
 12  Thal Rate           303 non-null    int64
 13  Target              303 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

Figure 7. Checking lost data

From the results of the image above, it can be concluded that there is no empty or missing data by knowing that each feature contains 303 records.

This stage conducts a search on the features that affect the target variable. The results of determining the relationship between features and targets can be seen in Figure 8 below:

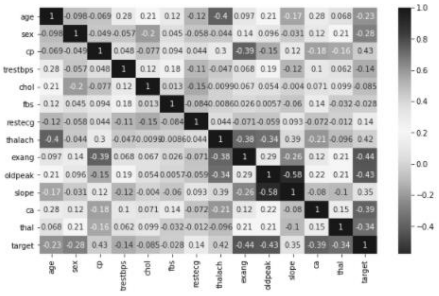



Figure 8. Correlation Matrix

To make it clearer in determining the features that are related to each other, you can calculate the correlation of each feature and sort it from the highest correlation using the df.corr()['target'].sort function, which results as shown in Figure 9 below:

```
target      1.000000
cp          0.433798
thalach     0.421741
slope       0.345877
restecg     0.137230
fbs        -0.028046
chol       -0.085239
trestbps   -0.144931
age        -0.225439
sex        -0.280937
thal       -0.344029
ca         -0.391724
oldpeak    -0.430696
exang      -0.436757
Name: target, dtype: float64
```

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

Figure 9. Correlation feature with corr() function

Then choose the feature that has the greater correlation because this feature will provide more information. The minimum threshold is 0.2, which aims to make it easier to choose features that have a high correlation with the diagnostic target variable. Figure 10 below is the result of threshold > 0.2 with the correlations[abs(correlations) function > 0.2.

```
target     1.000000
cp         0.433798
thalach    0.421741
slope      0.345877
age       -0.225439
sex       -0.280937
thal      -0.344029
ca        -0.391724
oldpeak   -0.430696
exang     -0.436757
Name: target, dtype: float64
```

Figure 10. Feature correlation with threshold is 0.2

Based on the results of the selection of these features will be seen the effect of the parameters Max Heart Rate (thalach), Cholesterol (chol), Resting Blood Pressure (trestbps) and ST depression (oldpeak) on age (Age). Below is Figure 11 which shows the effect of feature parameters on age.

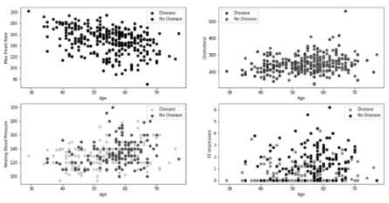


Figure 11. Feature correlation with threshold is 0.2

B. Data Pre-processing

Based on the results of data analysis, it can be concluded that the features of the dataset are interdependent on the target variable.

1. Splitting the dataset into Training sets and Test sets: This stage divides the dataset into training data and Test data into 80% training data and 20% test data. Below is table II Splitting dataset.

TABLE II
SPLITTING DATASET

Feature Scaling: In raw datasets, some variables have very varied and random values, so it is very important to feature scaling these features, Feature Scaling is a way to make numerical data in the dataset have the same range of values (scale) . There is no longer one data variable that dominates the other data variables. In carrying out this process, we use the functions in the sklearn library, namely the from sklearn.preprocessing import function StandardScaler.

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

C. Model Selection

Selection of the model to produce outcome prediction by using deep learning technique, namely artificial neural network (ANN). Keras is a Python library intended for developing and evaluating deep learning models. This library houses a numeric compute library as well as TensorFlow which can train neural network models as they are built.

1. Initialization of Artificial Neural Networks: This network has two layers, hidden layer and output layer. Hidden Layer will use the sigmoid function for activation. The output layer has only one node and is used for regression, the output node is the same as the input node. That is, the activation function. Activation Function is a function that takes an input signal and produces an output signal, but takes into account the threshold. Below is Table III initialization of the Artificial Neural Networks model.

TABLE III

ANN INITIALIZATION

| **Inisialisasi ANN** |
|---|
| ann = tf.keras.models.Sequential()<br># the first hidden layer<br>ann.add(tf.keras.layers.Dense(units=13, activation='relu'))<br># the second hidden layer<br>ann.add(tf.keras.layers.Dense(units=13, activation='relu'))<br># the third hidden layer<br>ann.add(tf.keras.layers.Dense(units=13, activation='relu'))<br># the output layer<br>ann.add(tf.keras.layers.Dense(units=1, activation='sigmoid')) |

2. Training Model: epochs are used for the number of times the data set will pass through the network, and each time updating the weights. As the number of epochs increases, the network gets better and better at predicting targets in a defined training set. In the selection of epochs, it must be adjusted to train the network well and hopefully not too many because it will result in overfitting. In this study, epochs=100 were chosen to predict targets in the training set. The following is Table IV Training Model

INFOKOM
Jurnal Informatika dan Komputer
**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

TABLE IV
TRAINING MODEL

In complile the ANN Model by conducting several experiments for the epochs values and Table V below describes the filled epochs values and the results.

TABLE V
EPOCHS VALUE EXPERIMENT

| Nilai Epochs | Confusion Matrix | Accuracy |
|---|---|---|
| 30 | [[19 8]<br>[ 4 30]] | 0.8032786885245902 |
| 50 | [[20 7]<br>[ 4 30]] | 0.819672131147541 |
| 80 | [[23 4]<br>[ 3 31]] | 0.8852459016393442 |
| 100 | [[23 4]<br>[ 2 32]] | 0.9016393442622951 |

To find out the model formed is not overfitting, it can be seen in the form of visualization of the loss model and the accuracy model from Figure 6 Loss Model and Figure 12 Accuracy Model

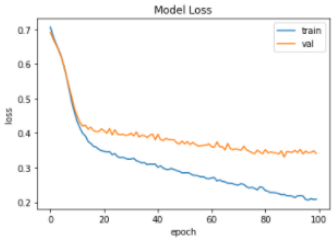



Figure 12 Model Loss

Based on the visualization image of the loss model, it can be seen that if the validation loss decreases, it means that the data is not overfitting. On the other hand, if the validation loss increases, it means that the data is overfitting.

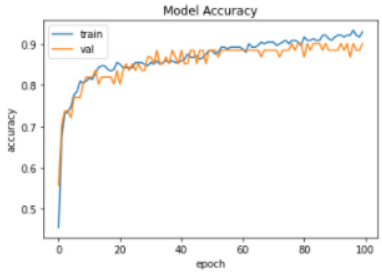


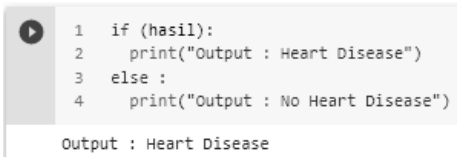

Figure 13 Accuracy Model

Based on the visualization image of the accuracy model, the accuracy calculation of the training set is basically in the same line (match) with validation accuracy.

3. Testing on the Model: This stage tests the independent variable to produce predictions on the dependent variable, namely y = df['target']. From the test results, the accuracy value in the prediction stage is obtained. The Artificial Neural Networks model was tested by entering the value of the independent variable ann.predict (sc.transform([[63, 1, 3, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1]]) and the resulting a value of 1 which means the presence of a heart attack disease.

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

TABLE VI
TEST MODEL

| Feature | Value |
|---|---|
| age | 63 |
| sex | 1 |
| cp | 3 |
| trestbps | 145 |
| chol | 223 |
| fbs | 1 |
| restecg | 0 |
| thalach | 150 |
| exang | 0 |
| oldpeak | 2.3 |
| slope | 0 |
| ca | 0 |
| thal | 1 |
| ann.predict(sc.transform([[63, 1, 3, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1]])   | |
| target | 1 (heart disease) |

D. Model Result

In the testing phase of the ANN model, it is compared with the machine learning technique of the Random Forest algorithm, the aim is to see the results of the accuracy values of each model.

The Matthews correlation coefficient (MCC) [16], is a more reliable statistical level that results in a high score only if the prediction is good in all of the four categories of the confusion matrix (true positives, false negatives, true negatives, and false positives). proportional both to the size of the positive elements and the sizes of the negative elements in the

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

data set (dataset). In using the MCC function, you can use the library from scikit-learn. Below is Figure 14 Result Model for Neural Network and Figure 15 Result Model for Random Forest.

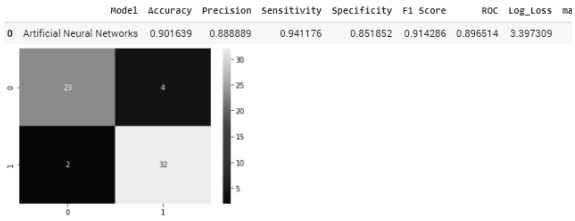



Figure 14. Result Model for Neural Network

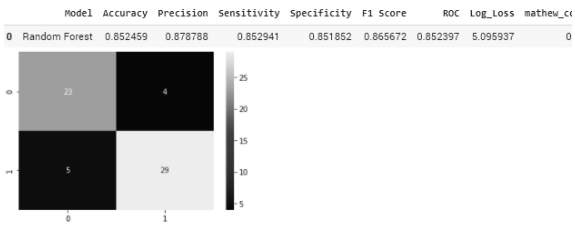
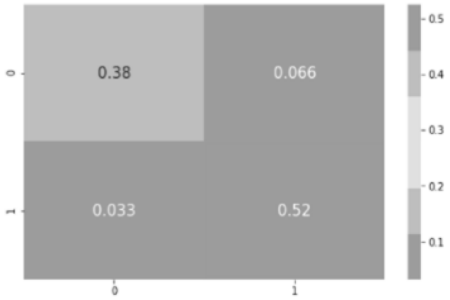



Figure 15. Result Model for Random Forest

After making the model results from each of these algorithms, below is a visualization to measure the performance of the model, Figure 10 Confusion Matrix Model Neural Network and Figure 11 Confusion Matrix Model Random Forest. The confusion matrix is used to measure the performance of the formed model.

Artificial Neural Network Model Accuracy of the model: 0.9016393442622951.




Gambar 16. Confusion Matrix Model Neural Network

Random Forest Model

Random Forest accuracy: 0.8524590163934426

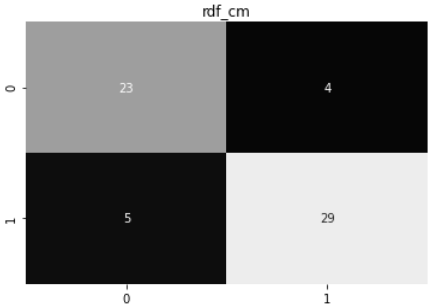



Figure 17. Confusion Matrix Model Random Forest

The confusion matrix in this case presents data in the form of numbers. As for the 2-class classification problem (binary classification), if you want to show classification algorithm performance data in the form of a graph, you can use Receiver Operating

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

Characteristics (ROC) or Precision-Recall Curve.

The ROC curve is made based on the value already obtained in the calculation with the confusion matrix, which is between the False Positive Rate and the True Positive Rate, is:

a) False Positive Rate (FPR) = False Positive / (False Positive + True Negative)

b) True Positive Rate (TPR) = True Positive / (True Positive + False Negative)

The Precision-Recall Curve is made based on the value obtained in the calculation with the confusion matrix, namely between Precision and Recall, is:

a) precision = True Positive / (True Positive + False Positive)

b) recall = True Positive / (True Positive + False Negative)

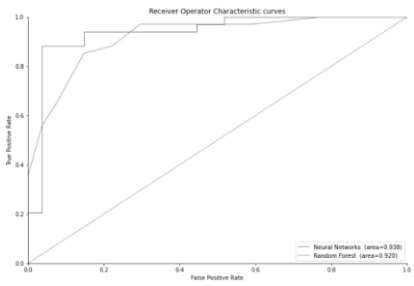



Figure 18 Receiver Operating Characteristics (ROC)

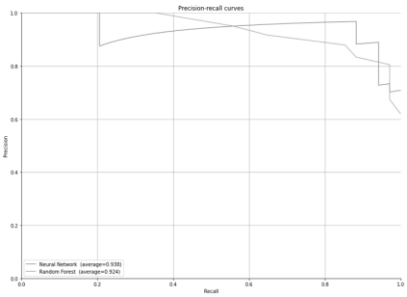



Figure 19 Precision-Recall Curve

The results of this study illustrate that the Artificial Neural Network model in predicting heart disease using a free dataset from the UCI Machine Learning Repository has 90% accuracy compared to the Random Forest model which produces 85% accuracy which can be seen in Table VI Comparison of Models. In the compilation stage of this neural network model, epoch=100 is given and data validation parameter attributes are added. The results of the validation data information produce loss: 0.2091 - accuracy: 0.9298 - val_loss: 0.3413 - val_accuracy: 0.9016. The features of the heart disease dataset that contribute to the results of the data analysis are st_slope_upsloping, st_slope_flat, exercise_induced_angina, sex and cholesterol have an influence on the target variable.

TABLE VII

MODEL COMPARISON

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

| Algoritma | | Precision | Recall | f1-score |
|---|---|---|---|---|
| Artificial Neural Network | 0 | 0.92 | 0.85 | 0.88 |
| | 1 | 0.89 | 0.94 | 0.91 |
| Accuracy | 0.90 | | | |
| Random Forest | 0 | 0.82 | 0.85 | 0.84 |
| | 1 | 0.88 | 0.85 | 0.87 |
| Accuracy | 0.85 | | | |

## CONCLUSION

Based on the results of research that has been carried out to produce outcome predictions in predicting heart disease using the Artificial Neural Network algorithm, it produces accuracy = 90% and as a comparison material for accuracy results compared to the Random Forest algorithm, namely accuracy = 85%. There are 5 (five) features that contribute to influencing the target variable, namely st_slope_upsloping, st_slope_flat, exercise_induced_angina, sex and cholesterol. Testing the model created by filling in the feature values and producing predictions 1 = heart disease 0 = No heart disease. This neural network technique uses 3 (three) hidden layers with a value of epochs = 100.

Subsequent research will predict more than one dataset regarding heart disease other than the UCI Machine Learning Repository in order to have a comparison of the accuracy values through the selection of certain algorithms and exploration activities for data analysis (Exploration Data Analyst) that need to be improved to knowing the correlation of the features in the dataset and further handling of pre-processing data as well as the research activities that have been carried out can make a good contribution in the field of health data science and as material for further research.

## BIBLIOGRAPHY

[1] United Nations Department of Economic and Social Affairs, "World's population increasingly urban with more than half living in urban areas," *Report on World Urbanization Prospects, Jul. 2014*, 2014.

[2] G. Perkasa, "Penyakit Jantung Penyebab Kematian Utama di Dunia," *Kompas.com*, 2020.

[3] A. Riani, Y. Susianto, N. Rahman, dan U. D. Ali, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Mengunakan

**Jurnal Informatika dan Komputer (INFOKOM)**
Volume 9 Nomor 1 Tahun 2021
E ISSN : 2722-4147 P ISSN : 2339-188X
OJS : http://journal.piksi.ac.id/index.php/INFOKOM

Metode Naive Bayes Data Mining Implementation to Predict Heart Disease using Naive Bayes Method," vol. 1, no. 01, hal. 25–34, 2019, doi: 10.35970/jinita.v1i01.64.

[4] S. Komputer *dkk.*, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," vol. 4, no. 1, hal. 84–88, 2020.

[5] A. Alhamad, A. I. S. Azis, B. Santoso, dan S. Taliki, "Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote," vol. 5, no. 3, hal. 352–360, 2019.

6] F. Shamout, T. Zhu, dan D. A. Clifton, "Machine Learning for Clinical Outcome Prediction," *IEEE Rev. Biomed. Eng.*, vol. 14, hal. 116–126, 2021, doi: 10.1109/RBME.2020.3007816.

[7] A. D. Kelleher, John D, Brian Mac Namee, *FUNDAMENTALS OF MACHINE LEARNING FOR PREDICTIVE DATA ANALYTICS*. London: The MIT Press Cambrid, 2015.

[8] L. W. Scikit-learn dan M. L. W. Scikit-, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media, 2017.

[9] M. B. Nugroho, "Machine Learning For Beginners Algorithms, Decision Tree & Random Forest Introduction," *J. Chem. Inf. Model.*, vol. 53, no. 9, hal. 1689–1699, 2013.

[10] D. Radečić, "Master Machine Learning: Random Forest From Scratch With Python," 2014. https://towardsdatascience.com/master-machine-learning-random-forest-from-scratch-with-python-3efdd51b6d7a.