

CHAPTER 3

RESEARCH METHODOLOGY

In this project, there are two algorithms used to examine the level of accuracy in predicting heart attacks. The first thing to do is download the dataset from Kaggle website that available online at: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>, this dataset was uploaded to the Kaggle website then created by combining various datasets that are already available independently but have not been combined before. In this dataset, 5 heart datasets from hospitals in five regions are combined with more than 11 common features of heart dataset making it the largest heart disease data set available so far for research purposes. This dataset amounted to an initial 1190 data, with data cleaning found 272 duplicate data, so that the final data available was 918 data.

Next is to process the dataset that has been collected using software called Orange, software used for processing data mining. This software is used to process downloaded heart attack datasets. The first thing to do is import the dataset that has been searched. The dataset here has various formats, you can use csv format, excel documents, SQL and even images. In this project the data used is csv data taken from the Kaggle site. Then after the dataset is successfully imported, the data will enter the transform stage.

At the transform stage, select the column select tool and data sampler. The select column is in charge of choosing which column will be used as the result of comparison with the predicted results of the algorithm that has been selected. The selected column in the Data sampler is used to take a random sample of data to be used in the data training process for the selected algorithm. In the data sampler, you can set how much data will be used for training the algorithm, and how much will be used for the comparison of the original data and the predicted results from the algorithm. After finishing setting the sampler data, it will enter the next stage, namely Modeling.

Entering the Modeling stage, here data is processed for training on predetermined algorithms, namely Neural Networks and Random Forests. Both algorithms will work to find the last output results and the output results will be compared with the original results from the dataset. After the results of the output training data from the two algorithms come out, we will compare these results with the actual results. We will enter the Evaluation stage.

Towards the final stage is Evaluation, where at this stage the results of the algorithm training in the Modeling process will be matched with the actual results in the remaining existing datasets. Here the results of the algorithm training in the Modeling process will predict the output value of the data provided and then the results will be compared with the results that already exist in the dataset. In this process, the error value of each algorithm will be shown for the dataset being tested. Not only error values and prediction results are displayed, but there are other assessments such as AUC, CA, F1, Precision, and Recall.

