

ABSTRAK

Penyakit jantung merupakan penyebab kematian nomor satu di dunia. Telah banyak dilakukan penelitian terkait pemanfaatan *machine learning* untuk memprediksi penyakit jantung. *Machine Learning* adalah bagian dari penelitian tentang kecerdasan buatan, mencari untuk memberikan pengetahuan kepada komputer melalui data, pengamatan dan berinteraksi dengan dunia. Terdapat berbagai macam metode dalam pemanfaatan *machine learning* diantaranya adalah Naïve Bayes dan Decision Tree yang akan digunakan pada penelitian ini. Penelitian ini bertujuan untuk meningkatkan akurasi dalam studi kasus prediksi penyakit jantung dari penelitian sebelumnya dengan menambahkan jumlah data dan membuktikan metode mana yang memberikan akurasi lebih baik antara Naïve Bayes atau Decision Tree. Data yang digunakan merupakan data sekunder yang didapat dari website Kaggle, berisi 14 subset atribut, yang terdiri dari 13 atribut sebagai *predictor* dan satu atribut sebagai target dengan dua kelas dengan jumlah 1328 data. Data kemudian akan diklasifikasikan untuk mencari nilai akurasi, sensitivitas, dan spesifisitas yang tertinggi dari kedua metode. Evaluasi sistem dilakukan dengan menggunakan *confusion matrix*, *K Fold Cross Validation*, dan Kurva ROC dengan menghitung nilai *Area Under Curve* (AUC). Dari hasil evaluasi sistem diperoleh nilai tertinggi pada metode Decision Tree pada pengujian kurva ROC dan *confusion matrix* dengan nilai *fold* 10, dengan rata-rata nilai akurasi sebesar 97.81%, sensitivitas sebesar 97.12%, spesifisitas sebesar 98.58%, dan AUC sebesar 0.979. Sedangkan metode Naïve Bayes memperoleh hasil paling tinggi pada pengujian kurva ROC dan *confusion matrix* dengan nilai *fold* 5 dengan rata-rata nilai akurasi sebesar 83.03%, sensitivitas sebesar 84.70%, spesifisitas sebesar 81.68%, dan AUC sebesar 0.832.

Kata Kunci : Naïve Bayes, Decision Tree, Heart Disease Classification, confusion matrix, Kurva ROC.

ABSTRACT

Heart disease is the number one cause of death in the world. There have been many studies related to the use of machine learning to predict heart disease. Machine Learning is part of research on artificial intelligence, looking to provide knowledge to computers through data, observation and interacting with the world. There are various methods in utilizing machine learning including Naïve Bayes and Decision Tree which will be used in this study. The purpose of this research is to improve accuracy in case studies of prediction of heart disease from previous studies by adding data and knowing which method provides better accuracy between Naïve Bayes or Decision Tree. The data used are secondary data obtained from the Kaggle website, containing 14 subsets of attributes, consisting of 13 attributes as predictors and one attribute as a target with two classes with total amount of 1328 datas. Data will then be classified to find the highest accuracy, sensitivity and specificity of the two methods. System evaluation is done using confusion matrix, K Fold Cross Validation, and ROC Curve by calculating the value of Area Under Curve (AUC). From the system evaluation results obtained the highest value in the Decision Tree method in testing the ROC curve and confusion matrix with a fold value of 10, with an average accuracy of 97.81%, an average sensitivity of 97.12%, an average specificity of 98.58%, and an average average AUC of 0.979. While the Naïve Bayes method obtained the highest results in testing the ROC curve and confusion matrix with a fold value of 5 with an average accuracy value of 83.03%, a sensitivity of 84.70%, a specificity of 81.68%, and an average AUC of 0.832.

Keywords : Naïve Bayes, Decision Tree, Heart Disease Classification, confusion matrix, ROC Curve.