

## Abstrak

Jantung merupakan organ dalam manusia yang fungsinya sangatlah penting yaitu untuk mengedarkan darah yang berisi oksigen dan nutrisi ke seluruh tubuh dan mengangkut sisa hasil metabolisme tubuh. Akan sangat fatal apabila di dalam organ jantung terdapat gangguan seperti penyumbatan pembuluh darah, dan lain-lain sehingga dapat menyebabkan jantung tidak dapat bekerja dan dapat menyebabkan kematian. Penyakit jantung juga bisa disebut dengan istilah *sudden death* karena penyakit ini sering kali tidak menimbulkan gejala namun, tiba-tiba pembuluh darah di jantung tersumbat sehingga tidak dapat memompa darah dan menyalurkan ke seluruh tubuh, sehingga dapat menyebabkan kematian.

Proses pendeteksian apakah seseorang terkena penyakit jantung dapat dilakukan dengan konsultasi kepada dokter spesialis jantung yang nantinya akan dilakukan pemeriksaan laboratorium dan dikonsultasikan kembali kepada dokter spesialis jantung. Namun cara tersebut tidaklah efisien, selain memakan waktu yang lama, juga karena memakan biaya yang cukup tinggi. Oleh karena itu perlu dilakukan pendeteksian penyakit jantung secara digital supaya dapat meningkatkan efektifitas kerja. Dalam kasus prediksi penyakit jantung, penelitian sebelumnya sudah banyak dilakukan dengan menggunakan berbagai macam algoritma klasifikasi yang ada. Dari berbagai macam algoritma yang sudah digunakan dalam penelitian sebelumnya, tentunya masing-masing memiliki kelebihan dan kekurangan. Contohnya adalah terjadi *overfitting* dimana akurasi pada *data training* sangat tinggi namun akurasi pada *data testing* sangatlah rendah. Oleh karena itu dalam penelitian ini algoritma yang dipilih adalah algoritma XgBoost yang dikenal memiliki kelebihan dapat mencegah *overfitting* namun tetap mempertahankan akurasi yang tinggi. Untuk data yang digunakan adalah data *cleveland heart diseases* yang berjumlah 303 dan memiliki 13 parameter.

Model yang dibangun dalam penelitian ini dilakukan *tuning hyper parameter* menggunakan Randomized Search untuk menemukan *hyper parameter* yang terbaik. Hasil dari penelitian ini adalah model klasifikasi penyakit jantung. Pengukuran akurasi model dilakukan menggunakan *confusion matrix*, dari hasil pengukuran didapatkan akurasi model secara keseluruhan adalah 90% dengan rincian akurasi pada *data training* adalah 91% dan akurasi pada *data testing* adalah 83%. Pengukuran juga dilakukan dengan *dataset* lain yaitu menggunakan *statelog heart diseases* dan mendapatkan akurasi sebesar 78%.

**Kata kunci:** penyakit jantung, prediksi, *overfitting*, XgBoost

## Abstract

*Heart is an organ in humans whose function is very important, namely to circulate blood containing oxygen and nutrients throughout the body and transport the rest of the body's metabolic products. It will be very fatal if there are disturbances in the heart organ such as blockage of blood vessels, etc. so that it can cause the heart to not work and can cause death. Heart disease can also be referred to as sudden death because this disease often causes no symptoms, but suddenly the blood vessels in the heart become blocked so that they cannot pump blood and distribute it throughout the body, which can cause death.*

*The process of detecting whether a person has heart disease can be done by consulting a cardiologist who will later be subjected to laboratory tests and consulted again with a cardiologist. However, this method is not efficient, in addition to taking a long time, also because it is quite expensive. Therefore, it is necessary to digitally detect heart disease in order to increase work effectiveness. In the case of prediction of heart disease, many previous studies have been carried out using various existing classification algorithms. Of the various algorithms that have been used in previous studies, of course, each has advantages and disadvantages. An example is overfitting where the accuracy of the training data is very high but the accuracy of the testing data is very low. Therefore, in this study the algorithm chosen is the XgBoost algorithm which is known to have the advantage of being able to prevent overfitting but still maintain high accuracy. The data used is cleveland heart disease data, which is 303 and has 13 parameters.*

*The model built in this research is done by tuning the hyper parameters using Randomized Search to find the best hyper parameters. The result of this study is a classification model of heart disease. Measurement of model accuracy is carried out using a confusion matrix, from the measurement results obtained the overall model accuracy is 90% with details of accuracy in training data is 91% and accuracy in testing data is 83%. Measurements were also carried out with other datasets, namely using a heart disease statelog and obtained an accuracy of 78%.*

**Keywords:** *heart disease, prediction, overfitting, XgBoost*