

PERBANDINGAN METODE *RANDOM FOREST* DAN ALGORITMA C4.5 PADA DIAGNOSIS PENYAKIT *LIVER*

Antisa Kurnia Hayatri

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

ABSTRAK

The liver is the largest and important organ for our body. The human cannot live unless it works properly. A liver disease may be caused by some factors, as unhealthy lifestyles, virus infections, immune system abnormalities, and genetics factors. Liver disease is considered to kill its patient quietly because it doesn't show its symptoms, especially in the first phase. Treatments in the first phase of the disease can prolong the age of its patients. To detect liver disease, patient needs take a blood test. A mistake in diagnosing the disease may endanger the patient's health. So, classification is essential in diagnosing the disease. Classification method which is used in this research is the Random Forest method and Algorithm C4.5 using Indian Liver Patient Dataset that is taken from the kaggle website. This data consists of clinical data of patients who are detected to positively or negatively suffer liver disease. The research's outcome that used Random Forest method is in the optimum model on $Mtry = 2$ and $Ntree = 100$ with an accuracy of 79%. The Algorithm C4.5 method resulted in a judgment tree with 4 node terminals and accuracy of 72%, even though the model cannot predict the liver negative category.

Keyword: Liver, Diagnose, Liver Disease, Random Forest, Algorithm C4.5