Classification of Pulmonary Tuberculosis Using the C4.5 Algorithm at Jember Lung Hospital

Angga Rahagiyanto, S.ST., M.T. (*Supervisor* I)

Tria Tita Putri Fatchiah

Medical Record Study Program
Health Departement

ABSTRACT

Tuberculosis is an infectious disease caused by the bacterium Mycobacterium tuberculosis. According to the 2022 report on the top 10 inpatient diseases at the Jember Lung Hospital, pulmonary tuberculosis with positive BTA ranked third with a total of 1,124 cases. This study aims to classify pulmonary tuberculosis using the C4.5 algorithm at the Jember Lung Hospital. The research is quantitative and processed using RapidMiner tools with the C4.5 algorithm method. Sampling was conducted using a simple random sampling technique, resulting in 358 records with 11 variables: cough $\geq$ 2 weeks, productive cough, blood-tinged sputum, chest pain, shortness of breath, malaise, weight loss, loss of appetite, chills, fever, and night sweats. The study results show that the variable cough $\geq$ 2 weeks has the highest gain ratio and is the most influential variable in determining pulmonary tuberculosis. The ratio comparison of training data to testing data of 90:10 resulted in an accuracy of 80.56%, precision of 73.91%, and recall of 94.44%.