Classification of Pneumonia Disease Using the C4.5 Algorithm Method at Kaliwates General Hospital Jember

Angga Rahagiyanto,.S.ST, M. T. (*chief councelor*)

Indah Dwi Lestari

Health Information Manaement Study Program
Majoring of Health

ABSTRACT

Pneumonia is an acute inflammation of the lung parenchyma caused by infections from pathogens (bacteria, viruses, fungi, and parasites), excluding Mycobacterium tuberculosis. Pneumonia ranks first among the top 10 diseases treated in the inpatient ward of Kaliwates General Hospital, Jember, with 1,483 cases (31.2%) reported from January to December 2023. In terms of mortality, it ranks sixth among the top 10 causes of death, with 17 cases. This study aims to classify pneumonia based on symptoms (cough, fever, sputum/purulent production, chest pain, infection, and lung consolidation signs) and risk factors (gender, age, and comorbidities). The research is quantitative and uses primary data. The sample consists of 315 cases and 315 controls selected through purposive sampling. Classification is done using the C4.5 algorithm with split validation. The variables identified include fever (99%), lung consolidation signs (94.7%), cough (87.9%), sputum/purulent production (82.5%), infection (64.1%), chest pain (61.9%), and shortness of breath (53.9%). The risk factors influencing pneumonia patients are male gender (59.4%), age under 5 years (41.9%), and over 60 years (20.3%), as well as comorbidities (50.5%). Four variables are used to detect pneumonia: lung consolidation signs, cough, fever, and infection. Based on the entropy, gain, split info, and gain ratio calculations, lung consolidation signs are the most influential variable with the highest gain ratio value of 0.7829, making it the root. The data split between training and testing is 80%:20%, yielding an accuracy of 99.20%, precision of 100%, and recall of 98.21%. Future researchers can use other classification methods such as K-NN, Naïve Bayes, and K-Means, increase the amount of training and testing data to enhance the dataset, and use this dataset for developing an early detection system.

***Keywords**: C4.5 Algorithm, Pneumonia, Split Validation*