The Classification of Breast Cancer Disease Using The C4.5 Algorithm
(*A Case Study at the West* Nusa Tenggara *Provincial Hospital*).

Muhammad Yunus, S.Kom., M.Kom (*Supervisor*)

Adhisty Maulia
Health Information Management Study Program
Department of Health

ABSTRACT

Breast cancer is one of the most common cancers and a leading cause of cancer-related deaths among women. Determining the grade of breast cancer is important for assessing disease severity and supporting appropriate treatment decisions. Medical records contain various clinical information that can be utilized to assist in identifying breast cancer grades. This study aims to classify breast cancer grades using the C4.5 algorithm based on patient medical record data from RSUD Provinsi NTB. The dataset consisted of 276 patient records, including 66 Grade 1 cases, 100 Grade 2 cases, and 110 Grade 3 cases. The research variables included clinical characteristics of breast cancer and anatomical pathology examination results obtained from patients' medical records. The results showed that the classification model achieved an accuracy of 74.07%, with precision values of 73.46%, 66.30%, and 80.33%, and recall values of 85.00%, 58.10%, and 81.67% for Grade 1, Grade 2, and Grade 3, respectively. Based on the generated decision tree, the variable "red, dry, scaly, or thickened breast skin or nipple" was identified as the root node, indicating that it was the most influential attribute in classifying breast cancer grades. The results demonstrated that the C4.5 algorithm is capable of classifying breast cancer into Grade 1, Grade 2, and Grade 3 based on clinical features available in medical record data.

***Keyword* :** *Breast Cancer, C4.5 Algorithm, Classification, Confusion matrix*