Analisis Klasterisasi Pasien Kanker Payudara (C50) Menggunakan Algoritma K-Means di Rumah Sakit Tingkat III Baladhika Husada Jember
(*Clustering Analysis of Breast Cancer Patients (C50) Using the K-Means Algorithm at Baladhika Husada Level III Hospital, Jember*). Mudafiq Riyan Pratama, S.Kom., M.Kom. (Pembimbing 1).

Erwiyandiningsih Kusumawardani
Program Studi Manajemen Informasi Kesehatan
Jurusan Kesehatan

ABSTRACT

Breast cancer has the highest incidence and mortality rate among women in Indonesia, including at Baladhika Husada Level III Hospital, Jember. The large volume of heterogeneous patient data requires an advanced analytical approach to group patients by characteristic similarities. This study aims to analyze the clustering of breast cancer patients using the K-Means algorithm with the assistance of the Orange Data Mining tool. The data used comprised 501 inpatient medical records of breast cancer patients from 2024–2025, covering the variables of age, BMI, family history of breast cancer, history of breast disease, breast pain, breast lump, axillary lump, change in breast shape and size, nipple change, nipple discharge, skin abnormalities, cancer locality status, length of stay, inpatient treatment, occupation, and education. The best clustering result was obtained at k=2 after excluding five variables with single-value dominance, yielding a Silhouette Score of 0.150 and two clusters: C1 (248 patients), labeled "Cluster of Patients without Changes in Breast Shape and Size" and C2 (253 patients), labeled "Cluster of Patients with Changes in Breast Shape and Size". These findings are expected to serve as a reference for data management and breast cancer patient care in hospitals, as well as to support strengthening early detection education through Breast Self-Examination (BSE) by hospitals and primary healthcare facilities to improve public awareness of the early signs and symptoms of breast cancer.

Keywords: *breast cancer, clustering, K-Means, Silhouette Coefficient, Orange Data Mining*