

PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR (KNN)* DAN *GAUSSIAN NAIVE BAYES (GNB)* DALAM KLASIFIKASI *BREAST CANCER COIMBRA*

Johan Taruna Wijaya¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

johantaruna007@gmail.com¹, hardian@unmuhjember.ac.id²,

habibatulazizah@unmuhjember.ac.id³.

ABSTRAK

Kanker payudara didefinisikan sebagai suatu penyakit *neoplasma* ganas yang berasal dari *parenchyma* dan menghasilkan frekuensi kematian yang menjadi penyebab utama kekhawatiran di dunia. Kanker payudara merupakan kanker kedua yang paling banyak diderita dan penyebab kelima kematian kanker di seluruh dunia dengan presentase 6,4% dari semua penyebab kematian. Pada penelitian ini dilakukan klasifikasi terhadap data kanker payudara, dimana data tersebut terdapat adalah data darah pengidap kanker payudara. Metode yang digunakan pada klasifikasi ini adalah *K-Nearest Neighbor (KNN)* dan *Gaussian Naive Bayes (GNB)*. Pengujian akurasi pada penelitian ini menggunakan *Cross Validation* dan evaluasi data ujia dengan *Confusion Matrix*. Dari penelitian ini didapatkan hasil pada 116 data darah kanker payudara, metode KNN menghasilkan akurasi 86,9% lebih baik dari pada GNB, dan untuk presisi dan recall, metode KNN menghasilkan presisi sebesar 87,3%, dan *recall* sebesar 86,7%, pengujian pada metode KNN menggunakan nilai $K=4$.

Kata Kunci : Klasifikasi data, Kanker Payudara, *K-Nearest Neighbor*, *Gaussian Naive Bayes*

**COMPARISON BETWEEN K-NEAREST NEIGHBOR (KNN) AND
GAUSSIAN NAIVE BAYES (GNB) ALGORITHM IN THE COIMBRA
BREAST CANCER CLASSIFICATION**

Johan Taruna Wijaya¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³

Informatics Engineering Study Program, Faculty of Engineering,

University of Muhammadiyah Jember

johantaruna007@gmail.com¹, hardian@unmuhjember.ac.id², habibatulazizah@unmuhjember.ac.id³

ABSTRACT

Breast cancer is defined as a malignant neoplasm that originates from the parenchyma and produces a death frequency which is a major cause of concern in the world. Breast cancer is the second most common cancer and the fifth leading cause of cancer deaths worldwide, accounting for 6.4% of all causes of death. In this study, the classification of breast cancer data was carried out, where the data contained was blood data for people with breast cancer. The methods used in this classification are *K-Nearest Neighbor* (KNN) and *Gaussian Naive Bayes* (GNB). Testing the accuracy in this study using Cross Validation and evaluation of the test data with the Confusion Matrix. From this study, it was found that on 116 breast cancer blood data, the KNN method produced an accuracy of 86.9% better than GNB, and for precision and recall, the KNN method produced a precision of 87.3%, and a recall of 86.7%, testing on the KNN method uses the value of $K = 4$.

Keywords : Data classification, Breast Cancer, K-Nearest Neighbor, Gaussian Naive Bayes