

KLASIFIKASI DATA PENDERITA *BREAST CANCER* MENGGUNAKAN METODE *GAUSSIAN NAIVE BAYES*

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
Rahmat Yulianto, Deni Arifianto

ABSTRAK

Kanker payudara merupakan tumor ganas yang menyerang jaringan sel-sel payudara. Kanker payudara merupakan masalah paling besar bagi wanita di seluruh dunia dan menyebabkan kematian utama bagi penderita kanker payudara. Penyakit kanker payudara di negara berkembang menunjukkan bahwa penyakit kanker dengan persentase kasus tertinggi, kurang lebih 43% kasus dan persentase kematian yaitu 12,9%. Pesatnya perkembangan teknologi sekarang ini memungkinkan untuk mendeteksi suatu penyakit kanker payudara dengan menggunakan teknik penalaran *Soft computing*. Salah satu teknik *data mining* yang dipakai dalam penelitian terhadap dunia kesehatan adalah klasifikasi. Penelitian ini berisi tentang pengukuran metode *Gaussian Naive Bayes* terhadap klasifikasi penyakit *breast cancer* terhadap pasien. Data yang digunakan berasal dari *UCI Machine Learning* dengan total data 116 pasien dengan partisi 80 data digunakan pada skenario uji dan 36 data digunakan pada uji validasi. Penelitian ini menggunakan skenario uji *K Fold Cross Validation* dengan nilai $k = 2, 4, 5, 8$ dan 10. Pada penelitian ini diperoleh hasil akurasi tertinggi yaitu 62,5% yang dihasilkan pada pengujian 10 *Fold* skenario 5 dan 9 dan presisi yang sama diperoleh yaitu 62,5%. Model pada 10 *Fold* skenario 5 dan 9 merupakan skenario terbaik maka keduanya akan diuji validasi. Hasil uji validasi menunjukkan pada 10 *Fold* skenario 5 dan 9 memperoleh hasil akurasi sebesar 61,11% dan presisi sebesar 56,66%.

Kata kunci : *Data Mining*, Klasifikasi, *Breast Cancer*, *Gaussian Naive Bayes*.

CLASSIFICATION OF DATA BREAST CANCER PATIENTS USING GAUSSIAN NAIVE BAYES METHOD

*Department of Informatics Engineering Faculty of Engineering, University of
Muhammadiyah Jember
Rahmat Yulianto, Deni Arifianto*

ABSTRACT

Breast cancer is a malignant tumor that attacks the breast tissue cells. Breast cancer is the biggest problem for women around the world and the main cause of death for breast cancer sufferers. Breast cancer in developing countries shows that cancer has the highest percentage of cases, approximately 43% of cases and the percentage of deaths is 12.9%. The rapid development of technology today makes it possible to detect a breast cancer by using soft computing reasoning techniques. One of the data mining techniques used in research on the world of health is classification. This study contains the measurement of the Gaussian Naive Bayes method for the classification of breast cancer in patients. The data used comes from UCI Machine Learning with a total data of 116 patients with 80 data partitions used in the test scenario and 36 data used in the validation test. This study uses the K Fold Cross Validation test scenario with k values = 2,4,5, 8 and 10. In this study, the highest accuracy results were 62.5% which was produced in the 10 Fold test scenario 5 and 9 and the same precision was obtained. ie 62.5%. The model in 10 Fold scenarios 5 and 9 is the best scenario, so both will be validated. The results of the validation test show that in 10 Fold scenarios 5 and 9, the accuracy is 61.11% and the precision is 56.66%.

Keywords: Data Mining, Classification, Breast Cancer, Gaussian Naive Bayes