

ABSTRAK

Kanker payudara terjadi karena adanya sel-sel abnormal pada payudara yang dapat berasal dari saluran susu (*epitel ductus*) dan kelenjar penghasil susu (*lobulus*). Umumnya kanker payudara terjadi pada wanita dengan rentang usia 20 sampai 50 tahun, serta menjadi jenis kanker yang menyebabkan kematian bagi wanita di dunia. Oleh karena itu, penelitian ini bertujuan untuk membuat sistem klasifikasi jenis kanker payudara yang dapat membantu tenaga kesehatan dalam mendignosis dan pengobatan pasien. Pada penelitian ini, enam tahapan *Cross Industry Standard Process for Data Mining* (CRIPS-DM) digunakan dalam proses data mining. Adapun implementasi sistem dilakukan menggunakan *framework* Flask dengan menggunakan pengembangan model RAD. Dataset kanker payudara yang berjumlah 569 dan 32 fitur diproses pada tahap persiapan data dengan pembersihan, normalisasi, transformasi dan seleksi fitur. Terdapat ketidakseimbangan pada jumlah data, maka teknik SMOTE digunakan untuk mengatasi permasalahan tersebut. Model klsifikasi dibangun menggunakan Algoritma Logistic Regression. Pada evaluasi model *Confusion Matrix* digunakan dengan menerapkan *K-Fold Cross Validation* untuk membagi data pada set pengujian dan pelatihan. Berdasarkan hasil evaluasi model, hasil terbaik didapatkan dengan menerapkan teknik SMOTE. Hasil didapatkan *accuracy* sebesar 90%, *precision* 89%, dan *recall* 92%. Adapun hasil evaluasi sistem dilakukan menggunakan *blackbox testing*, menunjukan bahwa fungsi pada sistem bekerja dengan baik. Namun, pada uji validasi perlu dipastikan dengan hasil patologi anatomi.

Kata Kunci: algoritma logistic regression, kanker payudara, smote



ABSTRACT

Breast cancer occurs due to abnormal cells in the breast that can originate from the milk ducts (duct epithelium) and milk-producing glands (lobules). Generally, breast cancer occurs in women with an age range of 20 to 50 years, and is the type of cancer that causes death for women in the world. Therefore, this study aims to create a breast cancer type classification system that can help health workers in diagnosing and treating patients. In this research, six stages of the Cross Industry Standard Process for Data Mining (CRIPS-DM) are used in the data mining process. The system implementation was carried out using the Flask framework using the RAD development model. The 569 breast cancer datasets and 32 features were processed at the data preparation stage with cleaning, normalization, transformation and feature selection. There is an imbalance in the amount of data, so the SMOTE technique is used to overcome this problem. The classification model was built using Logistic Regression Algorithm. In the model evaluation Confusion Matrix is used by applying K-Fold Cross Validation to divide the data in the testing and training sets. Based on the model evaluation results, the best results were obtained by applying the SMOTE technique. The results obtained accuracy of 90%, precision 89%, and recall 92%. The system evaluation results were carried out using blackbox testing, showing that the functions on the system work well. However, the validation test needs to be confirmed with the results of anatomical pathology.

Keyword: algorithm logistic regression, breast cancer, smote

