

ABSTRAK

KLASIFIKASI *MICROARRAY* PADA SEL KANKER PAYUDARA MENGUNAKAN METODE *EXTREME GRADIENT BOOSTING*

Oleh

FRISKA DAESY ELVINA SIMBOLON

Kanker payudara adalah suatu penyakit yang ditandai dengan pembelahan sel yang tidak terkendali, yang disebabkan oleh perubahan gen (mutasi) sehingga membuat sel secara mendadak bisa berubah. Kanker disebut juga dengan tumor, yang dibagi menjadi dua golongan, yaitu tumor jinak, dan tumor ganas. Kanker payudara menjadi kanker yang paling mematikan di seluruh dunia, terutama yang diderita oleh kaum wanita. Klasifikasi penyakit kanker payudara sangat penting untuk menentukan pengobatan yang tepat dan memberikan perkiraan prognosis yang akurat untuk pasien. Metode yang digunakan dalam penelitian ini adalah *XGBoost*, yang merupakan teknik *ensemble learning* yang cukup populer dalam *machine learning*. Penelitian ini bertujuan untuk mengetahui hasil klasifikasi algoritma *XGBoost* dengan menggunakan data DNA microarray pada sel kanker payudara yang berasal dari *National Center for Biotechnology Information (NCBI)*. Pada tahap awal, model *XGBoost* dilatih dengan menggunakan parameter *default*, kemudian melakukan *hyperparameter* dengan menggunakan metode *GridSearch CV* untuk mencari parameter terbaik yang dapat meningkatkan performa model. Hasil penelitian menunjukkan bahwa pembagian data *10-fold cross validation* dan telah dilakukan *hyperparameter*, performa model *XGBoost* meningkat signifikan. Akurasi model meningkat dari 50% menjadi 76%, presisi 50% menjadi 86%, *recall* dari 50% meningkat dari 60%, dan *f1-score* menjadi 71%. Hal ini menunjukkan bahwa *hyperparameter* dapat meningkatkan performa model *XGBoost* dalam klasifikasi *microarray* pada sel kanker payudara.

Kata kunci : Klasifikasi, Kanker Payudara, *XGBoost*, *Hyperparameter*

ABSTRACT

CLASSIFICATION OF BREAST CANCER CELL MICROARRAY USING EXTREME GRADIENT BOOSTING METHOD

By

FRISKA DAESY ELVINA SIMBOLON

Breast cancer is a disease characterized by uncontrolled cell division, caused by genetic mutations that can lead to sudden changes in the cells. Cancer, also known as a tumor, is divided into two categories: benign and malignant tumors. Breast cancer is the most fatal cancer worldwide, especially among women. The classification of breast cancer is crucial in determining appropriate treatment and providing accurate prognosis for patients. The method used in this study is XGBoost, which is a popular ensemble learning technique in machine learning. The study aims to investigate the classification results of the XGBoost algorithm using DNA microarray data on breast cancer cells obtained from the National Center for Biotechnology Information (NCBI). Initially, the XGBoost model was trained using default parameters, then hyperparameters were tuned using the GridSearch CV method to find the best parameters that could improve the model's performance. The results showed that after 10-fold cross-validation and hyperparameter tuning, the XGBoost model's performance significantly improved. The accuracy of the model increased from 50% to 76%, precision increased from 50% to 86%, recall increased from 50% to 60%, and the f1-score increased to 71%. This indicates that hyperparameter tuning can improve the XGBoost model's performance in the microarray classification of breast cancer cells.

Keywords : *Classification, Breast Cancer, XGBoost, Hyperparameter*