

ABSTRACT

PERFORMANCE EVALUATION OF ADABOOST AND XGBOOST ALGORITHMS FOR BREAST CANCER CLASSIFICATION USING THE WISCONSIN BREAST CANCER DATASET

By

Ardila Manda Putri

Breast cancer is one of the most common types of cancer among women worldwide and remains a leading cause of cancer-related deaths. Early detection plays a crucial role in improving patient survival rates. The advancement of machine learning techniques enables the development of classification models that can assist in the diagnosis of breast cancer more quickly and accurately. This study aims to compare the performance of the AdaBoost and XGBoost algorithms in classifying breast cancer using the Wisconsin Breast Cancer Dataset (WBCD). The research process includes data preprocessing, normalization, and handling class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE). The models were trained using the AdaBoost and XGBoost algorithms and evaluated using the 5-Fold Cross Validation method. Model performance was assessed using several evaluation metrics, including accuracy, precision, recall, and F1-score. The results show that the AdaBoost model achieved an accuracy of 96.2%, precision of 94.7%, recall of 94.4%, and an F1-score of 94.5%. Meanwhile, the XGBoost model achieved an accuracy of 96.1%, precision of 93.3%, recall of 95.5%, and an F1-score of 94.3%. These findings indicate that both algorithms demonstrate excellent classification performance. AdaBoost performs better in terms of accuracy, precision, and F1-score, while XGBoost shows a higher recall in detecting breast cancer cases.

Keywords: machine learning, breast cancer classification, AdaBoost, XGBoost.

ABSTRAK

EVALUASI KINERJA ALGORITMA ADABOOST DAN XGBOOST UNTUK KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN WISCONSIN BREAST CANCER DATASET

Oleh

Ardila Manda Putri

Kanker payudara merupakan salah satu jenis kanker dengan tingkat kejadian yang tinggi pada perempuan di dunia dan menjadi penyebab utama kematian akibat kanker. Deteksi dini sangat penting untuk meningkatkan peluang kesembuhan pasien. Perkembangan metode machine learning memungkinkan pengembangan model klasifikasi yang dapat membantu proses diagnosis kanker payudara secara lebih cepat dan akurat. Penelitian ini bertujuan untuk membandingkan kinerja algoritma AdaBoost dan XGBoost dalam melakukan klasifikasi kanker payudara menggunakan Wisconsin Breast Cancer Dataset (WBCD). Tahapan penelitian meliputi proses preprocessing data, normalisasi, serta penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Model kemudian dilatih menggunakan algoritma AdaBoost dan XGBoost dengan metode evaluasi 5-Fold Cross Validation. Kinerja model dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model AdaBoost memperoleh accuracy sebesar 96,2%, precision 94,7%, recall 94,4%, dan F1-score 94,5%. Sementara itu, model XGBoost memperoleh accuracy sebesar 96,1%, precision 93,3%, recall 95,5%, dan F1-score 94,3%. Hasil tersebut menunjukkan bahwa kedua algoritma memiliki performa klasifikasi yang sangat baik, dengan AdaBoost unggul pada accuracy, precision, dan F1-score, sedangkan XGBoost memiliki nilai recall yang lebih tinggi dalam mendeteksi kasus kanker payudara.

Kata-kata kunci: *machine learning*, klasifikasi kanker payudara, AdaBoost, XGBoost.