ABSTRAK

Kanker payudara merupakan salah satu jenis kanker dengan angka kejadian dan kematian tertinggi di dunia. Deteksi dini menjadi faktor penting dalam meningkatkan peluang kesembuhan pasien, sehingga diperlukan metode diagnosis yang akurat dan efisien. Perkembangan teknologi machine learning memungkinkan pengembangan model prediksi yang dapat membantu proses diagnosis kanker payudara berdasarkan data medis. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest dalam prediksi kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic). Dataset yang digunakan terdiri dari 569 data pasien dengan 30 fitur numerik yang merepresentasikan karakteristik sel payudara serta dua kelas diagnosis, yaitu benign (jinak) dan malignant (ganas). Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, normalisasi menggunakan StandardScaler, pembagian data menjadi data latih (80%) dan data uji (20%), pelatihan model, serta evaluasi menggunakan metrik akurasi, sensitivitas (recall), spesifisitas, dan Area Under Curve (AUC). Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine memiliki performa yang lebih baik dibandingkan Random Forest. SVM memperoleh nilai akurasi sebesar 97,37%, sensitivitas sebesar 97,62%, spesifisitas sebesar 97,22%, dan AUC sebesar 0,9921. Sementara itu, Random Forest memperoleh nilai akurasi sebesar 92,98%, sensitivitas sebesar 92,86%, spesifisitas sebesar 93,06%, dan AUC sebesar 0,9878. Berdasarkan hasil tersebut, dapat disimpulkan bahwa Support Vector Machine merupakan metode yang lebih optimal untuk prediksi kanker payudara pada dataset Breast Cancer Wisconsin (Diagnostic) dibandingkan Random Forest.

Kata Kunci: Kanker Payudara, Support Vector Machine, Random Forest, Machine Learning, Prediksi Klasifikasi.

ABSTRACT

Breast cancer is one of the types of cancer with the highest incidence and mortality rates worldwide. Early detection is a critical factor in improving patients' chances of recovery, making accurate and efficient diagnostic methods essential. Advances in machine learning technology have enabled the development of predictive models that can assist in the diagnosis of breast cancer based on medical data. This study aims to analyze and compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms in predicting breast cancer using the Wisconsin Breast Cancer (Diagnostic) dataset. The dataset consists of 569 patient records with 30 numerical features representing breast cell characteristics, as well as two diagnostic classes: benign and malignant. The research stages included data collection, data preprocessing, normalization using StandardScaler, splitting the data into training data (80%) and test data (20%), model training, and evaluation using accuracy, sensitivity (recall), specificity, and Area Below the Curve (AUC) metrics. The results of the study show that the Support Vector Machine algorithm performs better than Random Forest. SVM achieved an accuracy of 97.37%, a sensitivity of 97.62%, a specificity of 97.22%, and an AUC of 0.9921. Meanwhile, Random Forest achieved an accuracy of 92.98%, a sensitivity of 92.86%, a specificity of 93.06%, and an AUC of 0.9878. Based on these results, it can be concluded that Support Vector Machine is a more optimal method for breast cancer prediction on the Wisconsin Breast Cancer (Diagnostic) dataset compared to Random Forest.

Keywords: Breast Cancer, Support Vector Machine, Random Forest, Machine Learning, Classification Prediction.