

ABSTRAK

Kanker payudara adalah salah satu kanker paling umum dan penyebab utama kematian wanita di dunia. Banyak pasien menyadari kondisinya saat sudah stadium lanjut, sehingga deteksi dini sangat penting sebagai peringatan awal dan untuk penguatan diagnosis serta efektivitas pengobatan. Penelitian ini bertujuan untuk menerapkan *machine learning* dengan algoritma *Random Forest* untuk prediksi kanker payudara serta untuk mengidentifikasi faktor yang berpengaruh dalam prediksi tersebut. Data yang digunakan dalam penelitian ini berasal dari rekam medis pada RSUD Al-Ihsan yang terdiri dari dua kelas yaitu Ca Mammae dan Ca Mammae + Penyerta. Terdapat ketidakseimbangan data antara kedua kelas tersebut yaitu 85,62% (941 data) adalah Ca Mammae dan 14,38% (158 data) adalah Ca Mammae + Penyerta. Ketidakseimbangan data ini diatasi dengan teknik *random undersampling*, *random oversampling*, dan SMOTE untuk memastikan model yang dibangun lebih optimal. Proses pembangunan model *Random Forest* diterapkan pada tiga rasio *data splitting* yaitu 70:30, 80:20, dan 90:10 dan *grid search* untuk menentukan model optimal. Evaluasi model dilakukan menggunakan *confusion matrix*, menunjukkan bahwa model optimal dengan pembagian data 90:10 serta penanganan ketidakseimbangan menggunakan SMOTE dengan parameter *criterion="entropy"*, *max_depth=20*, *min_samples_leaf=1*, *min_samples_split=2*, *n_estimators=200*. Hasil penelitian menunjukkan bahwa *Random Forest* yang diatur dengan parameter optimal dapat memberikan performa yang baik dalam prediksi kanker payudara, memberikan potensi untuk digunakan sebagai alat bantu dalam diagnosis klinis dan pengambilan keputusan medis. Model ini mencapai akurasi sebesar 87,27% dengan *precision* 64,71%, *recall* 57,89%, *f1-score* 61,02%, dan jumlah duplikasi data sebanyak 145. Berdasarkan hasil model terbaik menunjukkan bahwa faktor yang mempengaruhi meningkatnya kanker payudara meliputi Trombosit, Hemoglobin, AST (SGOT), Ureum, dan ALT (SGPT). Penelitian ini diharapkan bermanfaat sebagai peringatan dini untuk meningkatkan kesadaran deteksi dini kanker payudara.

Kata kunci—algoritma *Random Forest*, kanker payudara, *machine learning*, penyeimbangan data, rekam medis