BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Kanker payudara merupakan urutan pertama kanker pada perempuan di dunia (*incidence rate* 40 per 100.000 perempuan), kasus baru yang ditemukan 30,5% dengan jumlah kematian 21,5% per tahun dari seluruh kasus kanker pada perempuan di dunia. Insiden kanker payudara di Indonesia adalah sebesar 26 per 100.000 penduduk. Angka kematian kanker payudara lebih tinggi pada negara berkembang dibandingkan negara maju. Penyebab utama meningkatnya mortalitas kanker di negara berkembang adalah kurangnya program skrining efektif yang dapat mendeteksi keadaan sebelum kanker, maupun mendeteksi kanker pada stadium dini sehingga penanganannya dilakukan sebelum kanker pada stadium lanjut [1].

Machine Learning yaitu bidang kesehatan dengan memprediksi atau mendiagnosa penyakit pasien. Prediksi penyakit menggunakan teknologi *Machine Learning* akan memudahkan tenaga medis untuk dapat mendiagnosa penyakit secara dini sehingga pasien dapat langsung ditangani dengan tepat [2].

Ensemble Learning merupakan teknik kombinasi algoritma *Machine Learning* yang bertujuan untuk memperoleh hasil akurasi prediksi yang lebih tinggi dibandingkan dengan algoritma *Machine Learning* yang digunakan secara tunggal.Penelitian perbandingan algoritma yang diimplementasikan pada prediksi penyakit kanker ini, akan menggunakan algoritma *Random Forest*, *AdaBoost*, *XGBoost, LightGBM* dan *ExtraTrees* [3].

Selain itu, dalam penelitian ini juga ditambahkan model *Decision Tree* sebagai salah satu algoritma pembelajaran mesin dasar dan sebagai model pembanding yang memiliki interpretasi yang sangat baik dan struktur pohon yang intuitif. Ville *et al* [4], menurutnya *decision tree* merupakan model yang kuat, fleksibel, dan dapat digunakan secara luas dalam klasifikasi dan prediksi karena kemampuannya untuk menghasilkan pemodelan berbasis aturan yang mudah dipahami serta mendukung berbagai jenis data termasuk yang memiliki nilai hilang [4].

1. 2 RUMUSAN MASALAH

1. Bagaimana prediksi hasil prediksi kanker payudara dengan metode *ensemble learning*?
2. Bagaimana kinerja metode *ensemble learning* dalam memprediksi jenis kanker payudara menggunakan dataset *Wisconsin Breast Cancer*?
3. Model *ensemble learning* manakah yang memberikan akurasi terbaik dalam klasifikasi kanker payudara dibandingkan dengan model pembelajaran mesin individu?
4. Bagaimana kinerja algoritma *Decision Tree* dalam memprediksi jenis kanker payudara dibandingkan dengan metode *ensemble learning*?

1. 3 BATASAN MASALAH

1. Penelitian ini hanya menggunakan dataset *Breast Cancer Data Winconsin (Original)* yang tersedia secara publik di *Kaggle*.
2. Model *ensemble learning* yang digunakan *Random Forest, AdaBoost, XGBoost, LightGBM ,ExtraTrees* dan model dasar *Decision Tree.*

3. Evaluasi model dilakukan menggunakan metrik klasifikasi seperti akurasi, *precision, recall, F1-score, dan ROC-AUC*, tanpa mempertimbangkan aspek medis lanjutan seperti stadium kanker atau riwayat pasien.

1. 4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

1. Menganalisis dan mengevaluasi kinerja metode *Ensemble Learning Random Forest*, *AdaBoost*, *XGBoost, LightGBM* dan *ExtraTrees* dalam prediksi kanker payudara.
2. Menentukan model ensemble learning yang paling optimal berdasarkan metrik evaluasi seperti akurasi, *precision, recall*, dan *F1-score* dalam klasifikasi kanker payudara.
3. Menganalisis performa algoritma *Decision Tree* dalam klasifikasi kanker payudara

1.4.2 Manfaat Penelitian

1. Memberikan referensi ilmiah tentang efektivitas metode *ensemble learning* dalam klasifikasi penyakit kanker payudara, khususnya dalam konteks pengolahan data medis.
2. Membantu tenaga medis atau peneliti dalam mengembangkan sistem pendukung keputusan berbasis machine learning yang lebih akurat dan andal.
3. Mendorong penerapan teknologi kecerdasan buatan dalam bidang kesehatan sebagai alat bantu diagnosis awal yang cepat dan efisien.

1. 5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I Pendahuluan: berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II Tinjauan Pustaka: membahas teori-teori dasar yang digunakan dalam penelitian, serta kajian literatur terkait.

BAB III Metodologi Penelitian: menjelaskan langkah-langkah penelitian, data yang digunakan, metode ensemble learning, dan teknik evaluasi.

BAB IV Hasil dan Pembahasan: menyajikan hasil implementasi model, evaluasi performa, serta analisis perbandingan model.

Bab V Kesimpulan dan Saran: menyimpulkan hasil penelitian dan memberikan saran untuk pengembangan lebih lanjut.