

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan penyakit kronis yang menjadi salah satu penyebab kematian tertinggi di dunia dan berdampak besar pada kesehatan, sosial, dan ekonomi masyarakat. Salah satu jenis kanker yang paling umum menyerang perempuan adalah kanker payudara, dengan 2,3 juta kasus baru di seluruh dunia pada tahun 2022, atau sekitar 11,6% dari total kasus kanker (World Health Organization, 2024). Di Indonesia, sebagian besar pasien kanker payudara terdiagnosis dalam stadium lanjut, dengan lebih dari 66% pasien datang ke rumah sakit dan sudah dalam kondisi stadium III atau IV (Iskandar dkk., 2024). Kondisi ini menunjukkan masih lemahnya deteksi dini dan pentingnya metode yang akurat dan cepat dalam menentukan stadium kanker, agar penanganan bisa dilakukan lebih optimal sejak tahap awal.

Studi menunjukkan bahwa pasien yang tidak segera mengetahui stadium kanker yang diderita cenderung mengalami tekanan psikologis berat, seperti stres dan gangguan kecemasan. Ketidakpastian akan kondisi kesehatan dan proses pengobatan seringkali memperburuk keadaan fisik pasien. Keadaan emosional ini memengaruhi keputusan medis yang diambil serta efektivitas terapi, sehingga proses diagnosis awal memegang peran penting dalam keberhasilan pengobatan (Okaidat dkk., 2020). Penentuan stadium yang akurat tidak hanya memberikan arahan yang tepat pada pemilihan metode terapi, tetapi juga meningkatkan kepercayaan pasien terhadap proses perawatan yang dijalani.

Tingkat kesadaran masyarakat terhadap pentingnya pemeriksaan kanker payudara masih tergolong rendah. Banyak pasien baru memeriksakan diri ketika gejala sudah menunjukkan tanda-tanda yang serius. Hal ini sering disebabkan oleh kurangnya edukasi, keterbatasan akses layanan kesehatan, dan stigma terhadap pemeriksaan kanker. Akibatnya, sebagian besar kasus ditemukan dalam kondisi stadium lanjut yang secara signifikan memperkecil peluang kesembuhan (Das dkk., 2021). Oleh karena itu, diperlukan pendekatan teknologi yang dapat membantu mempercepat proses klasifikasi stadium kanker berdasarkan data klinis pasien.

Kemajuan teknologi informasi, khususnya dalam bidang kecerdasan buatan, membuka peluang besar untuk mendukung sektor kesehatan melalui pengembangan sistem prediktif berbasis data. Teknik pembelajaran mesin memungkinkan pengolahan data pada skala besar secara efisien dan akurat untuk mendeteksi pola yang tidak mudah diidentifikasi secara manual oleh tenaga medis. Penerapan algoritma pembelajaran mesin dalam diagnosis medis telah menunjukkan hasil yang menjanjikan dalam meningkatkan kecepatan dan akurasi prediksi, serta meminimalisir kesalahan diagnosis (B. Kaur & Popli, 2024).

Dari studi lain diketahui bahwa penggunaan algoritma *Random Forest* telah terbukti mampu memberikan tingkat akurasi tinggi dalam memprediksi kanker payudara (Kabiraj dkk., 2020). Pada penelitian tersebut Kabiraj dkk. (2020) mengungkapkan bahwa model ini bekerja dengan membentuk beberapa *Decision Tree* dan mengambil hasil terbanyak dari setiap pohon keputusan tersebut. Algoritma *Random Forest* memberikan akurasi sebesar 74.73% dalam studi dengan 275 data instansi dan 12 atribut klinis. Kombinasi *Random Forest* dengan teknik lain seperti AdaBoost menghasilkan peningkatan performa klasifikasi tumor, dengan akurasi 98.57% dan sensitivitas hingga 100% (Rohan dkk., 2019). Algoritma ini juga menunjukkan korelasi kuat antara fitur-fitur penting dengan jenis tumor, sehingga modelnya dapat diandalkan dalam klasifikasi kanker berdasarkan karakteristik klinis.

Penelitian-penelitian sebelumnya juga menunjukkan bahwa *Random Forest* memiliki performa kompetitif dibanding algoritma pembelajaran mesin lainnya seperti XGBoost dan LightGBM (Ono & Mitani, 2021). Dengan teknik *recursive feature elimination*, *Random Forest* mampu meningkatkan akurasi prediksi dengan mengurangi noise dari fitur yang tidak relevan. Evaluasi model melalui pengujian silang (*cross-validation*) serta penggunaan *dataset* dari *UCI Machine Learning Repository* meningkatkan validitas eksternal dari model prediksi tersebut (Murugan dkk., 2017). Dapat diketahui untuk keunggulan model ini terletak pada kemampuannya menyeimbangkan antara interpretabilitas, akurasi, dan kecepatan proses klasifikasi.

Preferensi terhadap metode diagnosis berbasis *Machine Learning* semakin meningkat karena kemampuannya dalam memberikan hasil secara instan dan objektif. Model prediktif seperti *Random Forest* tidak hanya mampu melakukan klasifikasi secara otomatis berdasarkan data medis, tetapi juga dapat memberikan informasi fitur yang paling

berpengaruh terhadap hasil diagnosis. Hal ini sangat membantu para dokter dalam mengidentifikasi karakteristik pasien yang berisiko tinggi, sehingga strategi pengobatan dapat dirancang lebih personal dan efektif (Wang dkk., 2025). Efisiensi ini menjadikan *Random Forest* sebagai salah satu algoritma unggulan untuk aplikasi di bidang kesehatan, khususnya dalam prediksi kanker payudara.

Dari uraian latar belakang diatas, maka penelitian mengenai “klasifikasi stadium kanker berdasarkan karakteristik klinis menggunakan algoritma *Random Forest*”, bertujuan untuk membangun model klasifikasi berbasis *Machine Learning*. Algoritma *Random Forest* digunakan karena memiliki kapabilitas untuk menangani data kompleks, mengurangi *overfitting*, dan menghasilkan interpretasi fitur yang akurat. Melalui penelitian ini, diharapkan hasil yang diperoleh dapat mendukung proses pengambilan keputusan medis secara lebih cepat dan tepat, serta membantu pasien mendapatkan penanganan sejak dini dengan pendekatan yang berbasis data.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini adalah bagaimana algoritma Random Forest dalam mengklasifikasikan stadium kanker payudara.

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun dan mengevaluasi model klasifikasi stadium kanker payudara berbasis algoritma Random Forest dengan fokus pada pengukuran performa model dari segi akurasi, sensitivitas, dan presisi. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi fitur-fitur penting yang paling berpengaruh dalam proses klasifikasi, sehingga hasil yang diperoleh dapat mendukung proses diagnosis stadium kanker secara lebih cepat, tepat, dan berbasis data, guna membantu pengambilan keputusan medis serta meningkatkan kualitas perawatan pasien sejak tahap awal.

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat bagi beberapa pihak sebagai berikut:

1. Menghasilkan model klasifikasi yang dapat digunakan untuk memprediksi stadium kanker payudara.
2. Memberikan referensi bagi peneliti dan pengembang sistem dalam mengimplementasikan algoritma *Random Forest* untuk analisis medis.

1.5 Sistematika Penulisan

Untuk memberikan gambaran yang terstruktur, penulisan laporan ini disusun secara sistematis ke dalam beberapa bab berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan laporan skripsi mengenai klasifikasi stadium kanker berdasarkan karakteristik klinis menggunakan algoritma *Random Forest*.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori-teori yang relevan, seperti konsep klasifikasi dalam *Machine Learning*, algoritma *Random Forest*, karakteristik klinis pada penyakit kanker, serta penelitian terdahulu yang mendukung.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, jenis dan sumber data, teknik pengumpulan data, tahapan pemrosesan data, serta proses penerapan algoritma *Random Forest*.

BAB IV ANALISIS DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan data, evaluasi model klasifikasi stadium kanker, serta pembahasan mengenai temuan penelitian berdasarkan hasil prediksi.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian dan rekomendasi atau saran untuk pengembangan sistem selanjutnya serta implikasi penelitian di masa mendatang.