

BAB I PENDAHULUAN

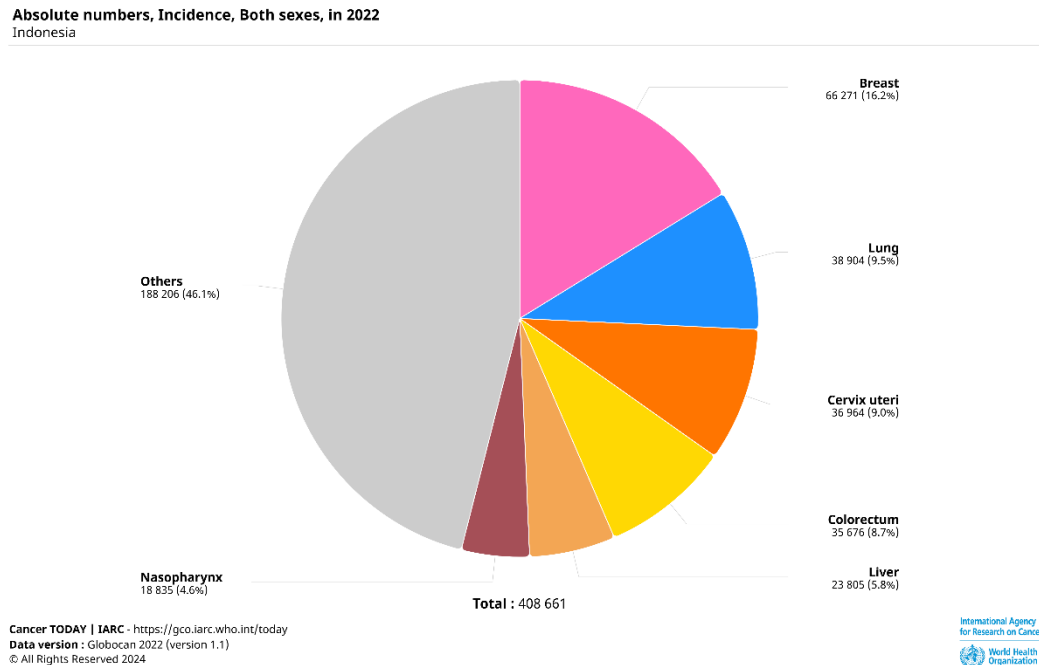
Pada bab ini menjelaskan tentang pemanfaatan *machine learning* di bidang kesehatan. Terdapat uraian mengenai latar belakang, perumusan masalah, tujuan, manfaat, batasan penelitian, dan sistematika penulisan.

I.1 Latar Belakang

Berdasarkan Undang-undang Republik Indonesia Nomor 36 Tahun 2009, kesehatan didefinisikan sebagai keadaan yang baik secara fisik, mental, spiritual maupun sosial yang memungkinkan setiap individu untuk hidup produktif secara sosial dan ekonomis (Undang-Undang Republik Indonesia Nomor 36 Tahun 2009 Tentang Kesehatan, 2009). Konsep ini menekankan pentingnya mencapai keseimbangan menyeluruh dalam keadaan kesehatan, tidak hanya tubuh yang bebas dari penyakit, tetapi juga dalam aspek jiwa, rohaniah, dan interaksi sosial yang mendukung. Undang-undang menjadi landasan bagi pemerintah dan masyarakat untuk menjaga serta meningkatkan kesehatan, termasuk pola makan, aktivitas fisik, kesehatan mental, dan lingkungan sehari-hari. Pada konteks ini, kesehatan fisik yang baik mencakup ketiadaan penyakit dan kondisi yang mendukung produktivitas tubuh. Salah satu ancaman serius terhadap kesehatan fisik adalah kanker, khususnya kanker payudara.

Menurut World Health Organization (WHO), kanker payudara adalah kondisi sel-sel payudara tumbuh tidak terkendali dan membentuk tumor yang jika tidak diatasi dapat berakibat fatal. Kanker ini yang paling umum terjadi pada wanita di seluruh dunia, dengan 2,3 juta kasus baru dan 670.000 kematian pada tahun 2022. Kanker ini terjadi pada wanita dari segala usia, terutama setelah pubertas, dan dapat juga terjadi pada pria meskipun sangat jarang (0,5-1%) (World Health Organization, 2024). Sebagai contoh, studi yang dilakukan oleh Pereira dkk., (2022) menunjukkan bahwa kanker payudara adalah salah satu penyakit yang dapat mengganggu kesehatan secara menyeluruh, karena tidak hanya berdampak pada aspek fisik melainkan menimbulkan beban psikologi dan sosial yang besar bagi penderita.

Kanker payudara umumnya terjadi pada saluran susu atau lobulus penghasil susu. Tahap awal dari penyakit ini tidak berbahaya, namun dapat menyebar ke jaringan sekitar dan menciptakan benjolan atau penebalan (World Health Organization, 2024). Diagram di bawah ini memberikan gambaran tentang distribusi jenis kanker di Indonesia, dengan jumlah kasus yang tercatat adalah 408.661 kasus (World Health Organization, 2022).



Gambar I.1 Jumlah Absolut dan Insiden Kanker Payudara Pada Kedua Jenis Kelamin di Indonesia Tahun 2022

Sumber : (<https://gco.iarc.who.int/today>)

Gambar I.1 di atas merujuk data dari World Health Organization (WHO) yang berasal dari Global Cancer Population (2022), menunjukkan persebaran penyakit kanker di Indonesia, yaitu kanker payudara (breast) sebesar 16,2%, *cancer lung* (kanker paru-paru) sebesar 9,5%, *cancer cervix uteri* (kanker leher rahim) sebesar 9,0%, *cancer colorectum* (kanker kolorektal) sebesar 8,7%, *cancer liver* (kanker hati) sebesar 5,8%, *cancer nasopharynx* (kanker nasofaring) sebesar 4,6%, dan jenis kanker lainnya sebesar 46,1%. Dari spesifik jenis kanker yang teridentifikasi, kanker payudara memiliki persentase yang lebih tinggi dari kanker paru-paru, serviks, kolorektal, dan hati.

Kanker payudara menjadi perhatian utama karena sifatnya multifaktorial yang dipengaruhi oleh struktur populasi, gaya hidup, faktor genetik, dan lingkungan. Beberapa faktor yang dapat meningkatkan risiko seseorang terkena kanker payudara adalah usia, riwayat keluarga, faktor hormon atau genetik, dan gaya hidup (Momenimovahed & Salehiniya, 2019). Perawatan untuk penyakit ini meliputi berbagai pendekatan seperti radioterapi, intervensi bedah, dan pengobatan yang disesuaikan dengan kondisi penderita serta jenis kankernya. Selain perawatan, deteksi dini kanker payudara juga merupakan langkah penting untuk meningkatkan kelangsungan hidup penderita (World Health Organization, 2024).

Deteksi dini kanker payudara dapat dilakukan melalui pemeriksaan payudara sendiri atau pemeriksaan payudara klinis oleh tenaga kesehatan (Rohan dkk., 2019). Selain pemeriksaan secara klinis, *machine learning* dapat dimanfaatkan dalam deteksi penyakit kanker payudara. Pada penelitian sebelumnya, *machine learning* dimanfaatkan untuk memprediksi kelangsungan hidup pasien kanker paru-paru dengan menggunakan regresi, klasifikasi, dan metode statistik dengan hasil yang signifikan (Bartholomai & Frieboes, 2018). Penggunaan *machine learning* bertujuan untuk meningkatkan akurasi prediksi, efisiensi pengolahan data, serta memungkinkan identifikasi dini faktor risiko dan pola penyakit. Pada penelitian ini *machine learning* digunakan untuk meningkatkan analisis pada data rekam medis dan memprediksi kemungkinan terkena penyakit kanker payudara serta mengembangkan strategi pencegahan dan penanganan yang lebih efektif sehingga memastikan deteksi dini dan pengobatan yang tepat.

Mengacu pada penelitian yang berjudul “*Machine Learning* Diagnosis Kanker Payudara Menggunakan Algoritma *Support Vector Machine*” oleh Chazar & Widhiaputra, (2020), algoritma *Support Vector Machine* (SVM) terbukti dapat meningkatkan hasil diagnosis terhadap penyakit kanker payudara. Proses klasifikasi menggunakan SVM dapat digunakan pada *dataset* yang memiliki jumlah fitur tinggi. Namun, SVM kurang cocok untuk *dataset* yang memiliki jumlah data besar karena membutuhkan waktu eksekusi yang lama dan juga tidak ideal untuk data yang mengandung banyak fitur atau dimensi tinggi (Li, 2015). Selain itu, penelitian yang dilakukan oleh Kabiraj dkk., (2020) menggunakan

dataset dari *UCI Machine Learning Repository* dengan 275 *instance* dan 12 fitur digunakan untuk memprediksi risiko kanker payudara menggunakan algoritma *Random Forest* dan *XGBoost*. Hasil penelitian tersebut menunjukkan bahwa *Random Forest* mencapai akurasi 74,73% dan *XGBoost* mencapai akurasi 73,63%.

Di dalam penelitiannya, Chazar & Widhiaputra (2020) dan Kabiraj dkk., (2020) menyarankan penggunaan algoritma lain untuk meningkatkan keakuratan hasil prediksi dan penerapan pada *dataset* yang lengkap serta tahapan *preprocessing* yang lebih efisien. Oleh karena itu, penelitian ini mengembangkan penelitian sebelumnya dengan menggunakan data rekam medis pasien RSUD Al-Ihsan untuk kasus kanker payudara namun menggunakan algoritma *Random Forest*. Algoritma *Random Forest* dikenal memiliki keunggulan dalam menangani data yang tidak seimbang serta memiliki kemampuan untuk memberikan hasil yang lebih akurat dibandingkan metode lain. Pada penelitian oleh Hairani dkk., (2023), penggunaan algoritma *Random Forest* pada data diabetes menunjukkan peningkatan signifikan dalam akurasi, *precision*, *recall*, dan *f1-score*. Sementara itu, penelitian oleh Macaulay dkk., (2021) menunjukkan bahwa *Random Forest Classifier* memiliki potensi yang baik dalam memprediksi risiko kanker payudara pada wanita Afrika, dengan hasil yang lebih baik dibandingkan teknik *machine learning* lainnya seperti *Support Vector Machine* dan *Neural Network*. Penelitian lain oleh Iman & Ahmad (2020) juga menunjukkan bahwa parameter yang diestimasi dari *Random Forest* dapat meningkatkan kinerja.

Dengan demikian, berdasarkan saran yang diberikan oleh Chazar & Widhiaputra (2020) dan Kabiraj dkk., (2020) untuk menggunakan algoritma lain dan *dataset* lebih lengkap, penelitian ini memberikan pendekatan berbeda dengan menggunakan *dataset* rekam medis pasien RSUD Al-Ihsan dan mengadopsi metode *Knowledge Discovery in Database* (KDD). Pendekatan ini diharapkan dapat meningkatkan akurasi dalam prediksi penyakit kanker payudara dan mendukung pengambilan keputusan yang lebih tepat berdasarkan analisis data medis yang komprehensif. Manfaat utama dari penelitian ini adalah memberikan peringatan dini (*early warning*) bagi masyarakat, membantu dalam upaya preventif, serta memperkuat diagnosis dokter untuk pengobatan yang lebih efektif.

I.2 Perumusan Masalah

Perumusan masalah yang mendasari penelitian ini adalah:

1. Bagaimana hasil penerapan *machine learning* dengan algoritma *Random Forest* untuk prediksi kanker payudara pada data rekam medis pasien RSUD Al-Ihsan?
2. Apa saja faktor yang mempengaruhi kanker payudara pada pasien RSUD Al-Ihsan berdasarkan algoritma *Random Forest*?

I.3 Tujuan Penelitian

Berdasarkan pada perumusan masalah, penelitian ini bertujuan untuk:

1. Mengetahui hasil penerapan *machine learning* dengan algoritma *Random Forest* untuk prediksi kanker payudara pada data rekam medis pasien RSUD Al-Ihsan.
2. Mengetahui faktor-faktor yang mempengaruhi kanker payudara pada pasien RSUD Al-Ihsan.

I.4 Manfaat Penelitian

Manfaat penelitian ini:

1. Bagi masyarakat luas, penelitian ini memberikan peringatan dini (*early warning*) yang membantu mencegah kanker payudara. Dengan adanya informasi ini, masyarakat dapat lebih waspada dan mengambil tindakan untuk mengurangi risiko terkena kanker payudara.
2. Bagi RSUD Al-Ihsan, penelitian ini bermanfaat untuk mengetahui hasil penerapan *machine learning* dan faktor yang mempengaruhi penyakit kanker payudara. Selain itu, penelitian ini membantu memahami lebih baik profil pasien yang beresiko terkena penyakit serta dapat meningkatkan layanan kesehatan dan perawatan pasien dengan lebih tepat sasaran.
3. Bagi Universitas Telkom, penelitian ini dapat menjadi kontribusi pada kemajuan pengetahuan di bidang kesehatan berupa diagnosis penyakit kanker payudara. Dengan menerapkan *machine learning*, penelitian ini juga akan menjadi awal mula kolaborasi pihak universitas dan instansi dalam meningkatkan reputasi.

4. Bagi peneliti lain yang bergerak dalam sistem informasi, penelitian ini bermanfaat untuk referensi melakukan penelitian serupa yang berhubungan dengan pendekatan *machine learning* dan algoritma *Random Forest*.

I.5 Batasan Penelitian

Berdasarkan perumusan masalah yang telah dipaparkan, penelitian ini dibatasi pada penggunaan *dataset* rekam medis pasien dari Instalasi Rawat Inap di RSUD Al-Ihsan. Data yang digunakan dari tahun 2020-2023 pada pasien kanker payudara.

I.6 Sistematika Penulisan

Laporan penelitian ini disusun dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang pemanfaatan *machine learning* di bidang kesehatan. Terdapat uraian mengenai latar belakang, perumusan masalah, tujuan, manfaat, batasan penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini menjelaskan landasan teori dan penelitian terdahulu sebagai acuan dalam penelitian ini. Landasan teori yang dituliskan diantaranya penggunaan rekam medis, metode *Knowledge Discovery in Database* (KDD), dan deteksi dini penyakit kanker payudara menggunakan algoritma *Random Forest*.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan metode dalam penelitian. Metode ini berupa kerangka konseptual yang mengintegrasikan ilmu perilaku, ilmu desain, *machine learning* dan sistematika penelitian yang dilakukan.

BAB IV ANALISIS DAN PERANCANGAN

Pada bab ini menjelaskan proses pengambilan data dan tahapan pengolahan. Tahapan ini meliputi penyusunan, pembersihan, membangun model dan beberapa tahapan lainnya untuk analisis lebih lanjut dengan algoritma *Random Forest*.

BAB V IMPLEMENTASI DAN PENGUJIAN

Pada bab ini menjelaskan pengujian model dengan perbandingan performa klasifikasi hingga evaluasi menggunakan *confusion matrix* dan *classification report*. Disajikan juga hasil pengujian model dan visualisasi fitur.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini menjelaskan kesimpulan dari serangkaian penelitian yang dilakukan. Selain itu, berisi saran untuk para peneliti selanjutnya.