

# BAB 1. PENDAHULUAN

## 1.1 Latar Belakang

Rumah sakit merupakan institusi pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan secara paripurna, mencakup layanan promotif, preventif, kuratif, dan rehabilitatif (Presiden RI, 2021). Rumah sakit menyelenggarakan tidak hanya pelayanan pengobatan, tetapi juga pendidikan, penelitian, dan pengembangan teknologi kesehatan untuk meningkatkan mutu layanan. (Presiden RI (2009); Sondakh *et al.*, (2023)). Rumah sakit memiliki sistem dokumentasi yang baik, yaitu rekam medis, untuk mendukung tugas-tugas tersebut.

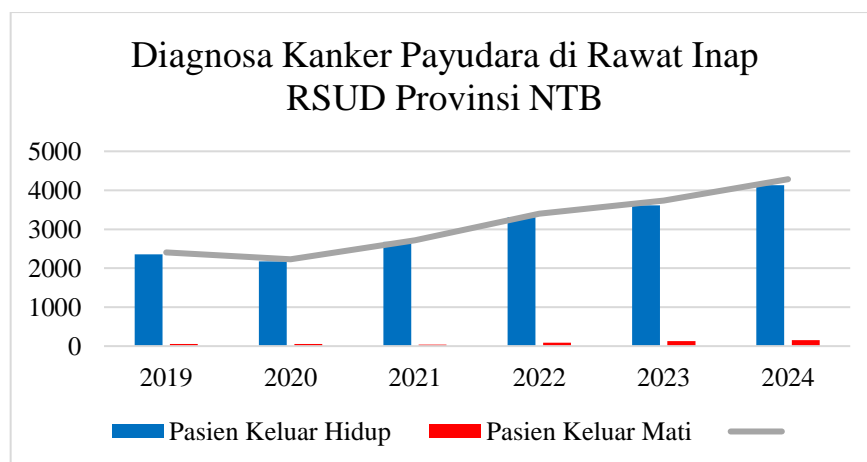
Rekam medis menyimpan seluruh informasi penting tentang pasien, termasuk identitas, hasil pemeriksaan, diagnosis, pengobatan, dan tindakan medis yang dilakukan (Kemenkes RI, 2022). Rekam medis tidak hanya berfungsi sebagai dokumen hukum, tetapi juga menjadi sumber data untuk analisis dalam mendukung perbaikan mutu pelayanan dan keputusan klinis maupun manajerial (Ramadani *et al.*, 2019). Perkam Medis dan Informasi Kesehatan (PMIK) memiliki peran penting dalam pengelolaan data, klasifikasi dan kodifikasi klinis, serta penerapan metode analitik dasar seperti statistik dan epidemiologi. (Kemenkes RI, 2020).

Data rekam medis kini dapat dianalisis lebih mendalam melalui teknik data *mining* berkat kemajuan teknologi. Data *mining* merupakan proses eksplorasi data dalam jumlah besar untuk menemukan pola yang valid dan berguna (Han *et al.*, 2011). Di bidang kesehatan, data *mining* banyak digunakan untuk mengidentifikasi tren penyakit, mengevaluasi efektivitas pengobatan, serta memprediksi kemungkinan pasien terkena penyakit tertentu berdasarkan riwayat klinis (Larose *et al.* (2014); Khomsah (2017)). Salah satu teknik dalam klasifikasi penyakit adalah algoritma C4.5, yang membentuk pohon keputusan berdasarkan atribut dalam *dataset*.

Kanker payudara merupakan salah satu penyakit tidak menular dengan tingkat kejadian tinggi di Indonesia maupun dunia. Penyakit ini menyerang jaringan epitel

payudara dan merupakan kanker paling umum pada wanita, serta penyebab utama kematian akibat kanker (Kemenkes RI (2015); Imaduddin *et al.* (2021)). Banyak kasus terlambat terdiagnosis akibat rendahnya kesadaran terhadap gejala awal (Dwika, 2021). Data WHO pada tahun 2018 mencatat 18,1 juta kasus kanker dengan 9,6 juta kematian, dan kanker payudara menjadi penyumbang kasus terbanyak kedua (Pangribowo, 2019). *Globacon* pada tahun 2022, kanker payudara menempati posisi pertama kasus kanker pada wanita sebesar 2.296.840 kasus dan posisi kedua untuk kanker secara keseluruhan. Di Indonesia pada tahun 2022, kanker payudara menjadi kasus terbanyak dengan total 66.271 kasus baru dan 22.598 kasus penderita yang meninggal dunia.

RSUD Provinsi NTB merupakan rumah sakit tipe A, yaitu rumah sakit dengan fasilitas layanan kesehatan tingkat tertinggi yang diwajibkan menyelenggarakan pelayanan promotif, preventif, kuratif, dan rehabilitatif secara menyeluruh (Permenkes No. 30 Tahun 2019). Sebagai pusat rujukan tertinggi di Nusa Tenggara Barat, rumah sakit ini memiliki peran strategis dalam pelayanan kanker, khususnya kanker payudara yang menunjukkan tren peningkatan signifikan dalam lima tahun terakhir. Berdasarkan data rekam medis rawat inap, jumlah pasien dengan diagnosis kanker payudara (ICD-10 C50.9) meningkat dari sekitar 2.300 kasus pada tahun 2020 menjadi 4.284 kasus pada tahun 2024. Gambar 1 di bawah ini memperlihatkan tren grafik peningkatan jumlah pasien rawat inap kanker payudara di RSUD Provinsi NTB sejak tahun 2019.



Gambar 1. 1 Grafik Diagnosa Kanker Payudara di Rawat Inap RSUD Provinsi NTB

Grafik diatas menunjukkan tren peningkatan jumlah pasien kanker payudara selama lima tahun terakhir. Pasien yang keluar hidup mendominasi, namun angka kematian tetap ada dan menunjukkan bahwa meskipun deteksi dan penanganan dilakukan, tantangan penanganan kanker tetap signifikan. Kenaikan jumlah kasus ini mencerminkan urgensi dalam melakukan intervensi yang lebih efektif, khususnya pada tahap deteksi dini. Angka tersebut menempatkan kanker payudara sebagai diagnosis terbanyak, melampaui penyakit kronis lainnya seperti anemia, diabetes, dan pneumonia. Detail jumlah pasien yang keluar hidup maupun meninggal dunia dapat dilihat pada Tabel 1.1.

Tabel 1. 1 Data 10 Besar Penyakit Rawat Inap di RSUD Provinsi NTB Tahun 2024

No	Diagnosa		Pasien Keluar Hidup		Pasien Keluar Mati		Total
	Kode	Nama Penyakit	L	P	L	P	
1	C50.9	<i>Malignant neoplasm of breast, unspecified</i>	2	4132	0	150	4284
2	D64.9	<i>Anemia, unspecified</i>	1085	1136	216	236	2673
3	E11.9	<i>Non-insulin-dependent diabetes mellitus without complications</i>	819	989	149	169	2126
4	J18.9	<i>Pneumonia, unspecified</i>	826	605	349	276	2056
5	E87.1	<i>Hypo-osmolality and hyponatraemia</i>	640	523	223	209	1595
6	E88.0	<i>Disorders of plasma-protein metabolism, not elsewhere classified</i>	537	482	300	251	1570
7	E87.6	<i>Hypokalaemia</i>	497	641	148	214	1500
8	N18.5	<i>Chronic kidney disease, stage 5</i>	717	508	147	111	1483
9	I15.9	<i>Secondary hypertension, unspecified</i>	571	690	71	75	1407
10	D63.0	<i>Anemia in neoplastic disease (C00-D48+)</i>	381	753	60	76	1270

Sumber : Data Sekunder (2024)

Meskipun sebagian besar pasien keluar hidup, jumlah kematian tetap tercatat, yang mengindikasikan bahwa sebagian pasien mungkin datang dalam kondisi lanjut atau terlambat terdiagnosis. Hal ini menjadi perhatian serius karena keterlambatan diagnosis sangat berkaitan erat dengan *Grade* atau stadium kanker. Setiap stadium memiliki karakteristik gejala yang berbeda. Pada stadium awal (*Grade* rendah), gejala bisa berupa benjolan kecil yang tidak nyeri, perubahan

bentuk atau kulit payudara, yang sering tidak disadari atau dianggap tidak berbahaya. Sebaliknya, stadium lanjut (*Grade* tinggi) ditandai dengan gejala yang lebih berat seperti retraksi puting, keluarnya cairan berdarah, pembengkakan kelenjar getah bening, atau ulserasi kulit (*American Cancer Society*, 2024).

Tingkat kesintasan juga sangat dipengaruhi oleh stadium saat terdiagnosis. Menurut *American Cancer Society* (2024), tingkat kelangsungan hidup lima tahun untuk pasien stadium I bisa mencapai 99%, namun angka ini menurun drastis hingga di bawah 30% untuk stadium lanjut. Rendahnya kesadaran gejala awal membuat deteksi dini menjadi sangat penting.

Penentuan *Grade* kanker payudara memiliki nilai penting dalam proses identifikasi tingkat keganasan dan perilaku biologis kanker, serta dapat digunakan sebagai dasar dalam klasifikasi berbasis data. Setiap *Grade* mencerminkan perbedaan tingkat kedewasaan dan kecepatan pembelahan sel kanker, yang berhubungan erat dengan gejala klinis yang muncul. Misalnya, kanker *Grade* rendah (*Grade* 1) biasanya menunjukkan pertumbuhan lambat dan gejala minimal, sedangkan kanker *Grade* tinggi (*Grade* 3) bersifat lebih agresif dan sering disertai dengan gejala yang lebih nyata seperti benjolan keras, perubahan bentuk payudara, atau keluarnya cairan abnormal dari puting (Rakha *et al.*, 2010; Adesiyun *et al.*, 2022). Hubungan antara gejala dan *Grade* ini memungkinkan penggunaan data klinis sebagai fitur penting dalam model klasifikasi, di mana algoritma seperti C4.5 dapat mengidentifikasi pola-pola gejala yang berkorelasi dengan tingkat keganasan. Penelitian terbaru menunjukkan bahwa pendekatan data *mining* berbasis gejala dan karakteristik histopatologi dapat meningkatkan ketepatan identifikasi *Grade* dan membantu dalam perencanaan terapi yang lebih individual (Kourou *et al.*, 2015; Jiang *et al.*, 2020). Dengan demikian, integrasi data gejala dan *Grade* dalam sistem klasifikasi berbasis *machine learning* menjadi langkah strategis untuk mendukung pengambilan keputusan klinis secara lebih akurat dan efisien.

RSUD Provinsi NTB memiliki tanggung jawab sebagai pusat layanan rujukan tertinggi yang menangani kasus-kasus berat dan kompleks, termasuk kanker payudara yang merupakan salah satu penyakit tidak menular dengan angka kejadian tinggi di provinsi tersebut (Dinkes NTB, 2023). Rumah sakit tipe A

dituntut untuk menyediakan layanan yang cepat, akurat, dan berbasis teknologi, terutama dalam proses diagnosis dan penentuan rencana terapi. Dalam konteks ini, klasifikasi kanker payudara menjadi langkah penting karena mampu mengelompokkan pasien berdasarkan tingkat keganasan (*Grade*) yang berkaitan erat dengan strategi pengobatan dan prediksi klinis (Rakha *et al.*, 2010; Adesiyun *et al.*, 2022). Penerapan metode klasifikasi berbasis algoritma C4.5 dapat membantu tenaga medis dalam memproses data klinis dan menentukan kategori kanker secara lebih efisien. Model klasifikasi ini dapat mempercepat proses deteksi dini, mendukung pengambilan keputusan medis, dan diarahkan untuk memperkuat program promotif dan preventif. Selain itu, implementasi sistem ini dapat dimanfaatkan dalam perencanaan sumber daya, program edukasi dan skrining terarah, serta pengelolaan epidemiologi kanker secara digital.

Metode algoritma C4.5 merupakan klasifikasi yang banyak digunakan dalam sistem rekomendasi dengan beberapa keunggulan (Tentua *et al.*, 2019). Algoritma C4.5 adalah yang digunakan dalam *decision tree*. Algoritma C4.5 memiliki konsep mengubah data menjadi keputusan dan aturan-aturan keputusan (*rule*) (Yunus *et al.*, 2023). Proses klasifikasi ini akan menghasilkan aturan-aturan yang dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru (Gustami *et al.*, 2022).

Asmara (2020) melakukan penelitian dengan membandingkan beberapa metode terhadap penyakit diabetes. Penelitian dilakukan dengan tujuan mengetahui metode yang paling akurat adalah algoritma C4.5 dengan nilai *accuracy* 67% dibandingkan dengan metode *Naives Bayes* dan *K-Nearest Neighbor* (Asmara *et al.*, 2020).

Penelitian ini menerapkan algoritma C4.5 untuk memfokuskan proses klasifikasi kanker payudara dalam mengelompokkan data gejala pasien ke dalam *Grade* 1, 2, atau 3 berdasarkan tingkat keganasannya. Hasil klasifikasi ini memungkinkan identifikasi gejala-gejala klinis yang paling relevan terhadap masing-masing *Grade*, sehingga dapat memberikan gambaran mengenai pola perkembangan kanker payudara. Informasi ini tidak hanya berguna untuk menunjang proses diagnosis, tetapi juga sebagai dasar pengambilan keputusan

klinis yang lebih cepat dan terarah. Penelitian ini menjadi penting dalam mendukung penguatan sistem klasifikasi berbasis data di rumah sakit tipe A seperti RSUD Provinsi NTB, yang menangani beban kasus kanker dalam jumlah besar sebagaimana tercermin dalam grafik dan tabel laporan kasus tahunan. Selain meningkatkan efisiensi identifikasi *Grade*, penerapan C4.5 melalui sistem *Orange* juga diharapkan membantu rumah sakit dalam mengurangi keterlambatan klasifikasi klinis yang dapat berdampak pada keterlambatan terapi dan penurunan prognosis pasien.

Berdasarkan uraian tersebut, peneliti bermaksud melakukan penelitian berjudul “Klasifikasi Penyakit Kanker Payudara dengan Algoritma C4.5 (Studi Kasus di RSUD Provinsi NTB)”.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah bagaimana klasifikasi dengan metode algoritma C4.5 dalam klasifikasi penyakit kanker payudara di RSUD Provinsi NTB?

## **1.3 Tujuan Penelitian**

### **1.3.1 Tujuan Umum**

Mengklasifikasikan penyakit kanker payudara menggunakan algoritma C4.5 di RSUD Provinsi NTB.

### **1.3.2 Tujuan Khusus**

- a. Melakukan *Data Acquisition* untuk mengidentifikasi karakteristik penyakit kanker payudara yang digunakan sebagai variabel untuk mengklasifikasikan penyakit kanker payudara di RSUD Provinsi NTB.
- b. Melakukan *Preprocessing Data* untuk menyiapkan *Dataset* untuk mengklasifikasikan penyakit kanker payudara di RSUD Provinsi NTB.
- c. Melakukan *Data Mining process* untuk mengklasifikasikan penyakit kanker payudara di RSUD Provinsi NTB menggunakan algoritma C4.5.
- d. Melakukan *Pattern Evaluation* untuk menganalisis hasil *Confusion matrix* dari klasifikasi penyakit kanker payudara menggunakan algoritma C4.5.

## **1.4 Manfaat**

### **1.4.1 Bagi Rumah Sakit**

Bahan masukan bagi rumah sakit sebagai informasi pengembangan mengenai klasifikasi penyakit kanker payudara di RSUD Provinsi NTB.

### **1.4.2 Bagi Institusi Pendidikan**

Hasil penelitian diharapkan dapat menambah wawasan dan masukan untuk institusi pendidikan khususnya bagi mahasiswa D-IV Manajemen Informasi Kesehatan dalam mengembangkan dan meningkatkan ilmu pengetahuan serta sebagai literatur bagi peneliti selanjutnya.

### **1.4.3 Bagi Peneliti**

Penelitian ini menjadi syarat dalam menyelesaikan pendidikan serta menjadi tambahan wawasan, pengetahuan, dan keterampilan dalam dunia kesehatan khususnya dibidang *data mining* dan klasifikasi.