

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Kanker payudara atau *Carcinoma Mammarum* (*Ca Mammarum*) yang bisa disebut juga dengan neoplasma ganas payudara merupakan keganasan pada jaringan payudara, biasanya berasal dari saluran susu (*epitel ductus*) yang mengalirkan ASI ke puting maupun kelenjar susu (*lobulus*) tempat ASI diproduksi (Menkes RI, 2018). Penyakit ini dapat berkembang dan menyebar ke jaringan di sekitarnya maupun ke organ lain apabila tidak terdeteksi dan ditangani sejak dini (Ketut & Kartika, 2022).

Kanker payudara merupakan salah satu masalah kesehatan utama di dunia karena angka kejadian dan kematiannya yang sangat tinggi dengan sekitar 2,3 juta kasus baru dan 670.000 kematian di seluruh dunia pada tahun 2022, menjadikannya kanker yang paling sering didiagnosis pada perempuan secara global (WHO, 2022). Data epidemiologi menunjukkan bahwa insiden dan angka kematian kanker payudara terus meningkat di banyak negara, terutama di wilayah berpenghasilan rendah dan menengah, yang mencerminkan tantangan besar dalam upaya pencegahan dan penanggulangan penyakit ini (WHO, 2022). Analisis global terbaru menunjukkan adanya peningkatan signifikan jumlah kasus kanker payudara yang dapat mencapai lebih dari enam juta kasus per tahun pada tahun 2050 (Freihat *et al.*, 2025).

Kanker payudara di Indonesia merupakan jenis kanker dengan insidensi dan mortalitas tertinggi pada perempuan dan menjadi penyebab kematian ketiga tertinggi setelah penyakit jantung dan stroke. Berdasarkan data *Global Cancer Observatory* (Globocan), jumlah kasus baru kanker di Indonesia pada tahun 2022 tercatat 408.661 kasus dan 242.099 kematian, di mana kanker payudara menyumbang proporsi terbesar pada perempuan (Menkes RI, 2024). Adapun berdasarkan data Globocan tahun 2022, kanker payudara di Indonesia menempati peringkat pertama dengan jumlah kasus baru sebanyak 66.271 kasus atau 16,2% dari seluruh kasus

kanker, serta menempati peringkat ketiga sebagai penyebab kematian akibat kanker payudara sebanyak 22.598 kasus (Globocan, 2022).

Tingginya angka kejadian dan kematian tersebut menunjukkan bahwa kanker payudara masih menjadi permasalahan kesehatan yang signifikan dan memerlukan perhatian khusus dalam upaya penanganan dan pengelolaan data pasien. Dalam konteks pelayanan kesehatan, khususnya rumah sakit rujukan, memiliki peran penting dalam penatalaksanaan pasien kanker, seperti kanker payudara, sekaligus mengelola data pasien yang terus meningkat (Fajar *et al.*, 2020). Data tersebut memiliki karakteristik klinis yang beragam, seperti perbedaan stadium penyakit, adanya komorbiditas, serta variasi tindakan medis yang diberikan (Fajar *et al.*, 2020). Penelitian sebelumnya juga menunjukkan bahwa data pasien kanker payudara di rumah sakit memiliki heterogenitas yang tinggi, sehingga memerlukan pengelolaan dan analisis data yang sistematis agar dapat dimanfaatkan secara optimal (Latamaosandhi & Suryadarma, 2025).

Kondisi tersebut juga tercermin pada tingkat pelayanan kesehatan di rumah sakit. Sebagai salah satu fasilitas pelayanan kesehatan rujukan, Rumah Sakit Tingkat III Baladhika Husada Jember menangani berbagai jenis kasus penyakit dengan jumlah pasien yang cukup tinggi. Berdasarkan data diagnosis pasien yang tercatat dalam rekam medis di Rumah Sakit Tingkat III Baladhika Husada Jember tahun 2025, diperoleh 10 besar penyakit dengan jumlah kasus terbanyak sebagaimana disajikan pada tabel berikut.

Tabel 1.1 Data 10 Besar Penyakit Rawat Inap Tahun 2025

No	Jenis Penyakit	Jumlah Kasus
1.	Neoplasma ganas payudara	1.105
2.	Demam tifoid dan paratifoid	1.064
3.	Penyakit sistem kemih lainnya	432
4.	Pneumonia	394
5.	Anemia lainnya	373
6.	Janin dan bayi baru lahir yang dipengaruhi faktor ibu dan kehamilan	326
7.	Neoplasma jinak lainnya	312
8.	Diare & gastroenteritis oleh penyebab infeksi	282
9.	Kondisi lain yang bermula pada masa perinatal	273
10.	Neoplasma ganas kolon	232

Berdasarkan Tabel 1.1, neoplasma ganas payudara atau kanker payudara mendominasi sebagai penyakit rawat inap terbanyak di Rumah Sakit Tingkat III Baladhika Husada Jember pada tahun 2025. Data ini menunjukkan bahwa kanker payudara merupakan jenis neoplasma tertinggi dibandingkan neoplasma lainnya, seperti neoplasma jinak lainnya dan neoplasma ganas kolon, sehingga memerlukan prioritas pengelolaan data dan sumber daya khusus.

Selain dilihat dari jumlah kasus, tingkat keparahan suatu penyakit dapat ditinjau dari angka kematian yang ditimbulkannya. Berdasarkan hal tersebut, berikut disajikan data sepuluh besar penyebab kematian pasien di Rumah Sakit Tingkat III Baladhika Husada Jember tahun 2025.

Tabel 1.2 Data 10 Besar Kematian Tahun 2025

No	Jenis Penyakit	Jumlah Kematian
1.	Neoplasma ganas payudara	21
2.	Gangguan endokrin, nutrisi, dan metabolik	16
3.	Pneumonia	16
4.	Penyakit sistem napas lainnya	16
5.	Diabetes mellitus tidak bergantung insulin	14
6.	Infark serebral	14
7.	Tuberkulosis paru lainnya	12
8.	Gagal jantung	12
9.	Septisemia	11
10.	Diabetes mellitus YTT	9

Berdasarkan Tabel 1.2, neoplasma ganas payudara atau kanker payudara menempati urutan pertama sebagai penyebab kematian utama di Rumah Sakit Tingkat III Baladhika Husada Jember pada tahun 2025. Temuan ini menunjukkan bahwa kanker payudara tidak hanya memiliki angka kejadian yang tinggi, tetapi juga menjadi penyebab kematian utama dibandingkan penyakit metabolik dan kardiovaskular, sehingga perlu mendapatkan perhatian khusus dalam pengelolaan dan pemanfaatan data pasien kanker payudara.

Tingginya jumlah kasus kanker payudara di Rumah Sakit Tingkat III Baladhika Husada Jember menghasilkan data pasien dalam jumlah besar dan kompleks dengan karakteristik klinis yang beragam, seperti variasi usia, stadium penyakit, ukuran tumor, dan manifestasi patologis yang berbeda antar pasien. Kondisi serupa juga ditemukan pada penelitian di RSUD Arifin Achmad Pekanbaru dan RSUD Bali Mandara yang menunjukkan bahwa data klinis kanker payudara memiliki heterogenitas yang tinggi akibat perbedaan distribusi usia, stadium penyakit, serta karakteristik klinikopatologis pasien (Latamaosandhi & Suryadarma, 2025; Utami *et al.*, 2025).

Heterogenitas data dengan jumlah variabel yang beragam tersebut menyebabkan penyajian data secara deskriptif menjadi kurang memadai dalam menggambarkan kompleksitas hubungan antar karakteristik klinis pasien kanker payudara (Iryanti *et al.*, 2025). Namun, hingga saat ini data pasien kanker payudara di rumah sakit umumnya masih dianalisis secara deskriptif, sehingga belum mampu menggambarkan pola kemiripan antar pasien secara komprehensif. Oleh karena itu, diperlukan pengolahan dan analisis data lanjutan menggunakan pendekatan yang mampu mengelompokkan data pasien berdasarkan kemiripan karakteristik, seperti melalui pendekatan statistik multivariat atau metode *data mining* yang mampu mengintegrasikan dan mengeksplorasi pola dalam dataset besar (Nurmayeni & Windarti, 2022). Sejalan dengan kebutuhan tersebut, penelitian ini memfokuskan pada analisis data kanker payudara melalui *data mining* dengan teknik *clustering* sebagai tahap awal untuk mengidentifikasi pola kelompok pasien yang memiliki karakteristik serupa.

Teknik *clustering* dipilih karena mampu mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label kelas sebelumnya, sehingga sesuai dengan karakteristik data klinis kanker payudara yang bersifat heterogen dan kompleks. Beberapa penelitian telah membuktikan efektivitas teknik ini dalam konteks medis, seperti penelitian yang dilakukan oleh Susilowati *et al.* (2023) menunjukkan bahwa penerapan *clustering* dapat mengelompokkan pasien kanker payudara berdasarkan kesamaan gejala klinis tanpa label kelas. Selain itu, penelitian internasional yang dilakukan oleh Ferro *et al.* (2021) membuktikan bahwa pendekatan *unsupervised learning* seperti teknik *clustering* mampu menemukan struktur alami dalam data pasien kanker payudara yang kompleks.

Salah satu teknik *clustering* dalam analisis data kesehatan adalah algoritma K-Means. Algoritma ini memiliki keunggulan dalam kesederhanaan, efisiensi komputasi, serta kemampuannya dalam

menangani data berukuran besar, seperti pada dataset pasien kanker payudara. Beberapa penelitian menunjukkan bahwa algoritma K-Means lebih unggul dari sisi efisiensi dan kualitas pengelompokan pada data berukuran besar yang telah melalui tahap *preprocessing*. Penelitian yang dilakukan oleh Yohannes (2016) menunjukkan bahwa K-Means memiliki waktu komputasi yang lebih cepat dibanding Fuzzy C-Means, sehingga lebih sesuai untuk analisis data kesehatan skala besar. Selain itu, dibandingkan K-Medoids dan DBSCAN, K-Means terbukti mampu menghasilkan kluster yang lebih kompak dan terpisah dengan baik, khususnya pada data yang relatif bersih dan tidak didominasi oleh *noise* (Adha *et al.*, 2021; Putra *et al.*, 2025).

Terkait konteks data kanker payudara, salah satu penelitian di Indonesia menunjukkan bahwa metode K-Means dapat digunakan untuk mengelompokkan pasien kanker payudara berdasarkan karakteristik tertentu, sehingga berpotensi membantu proses analisis awal data medis tanpa memerlukan label kelas sebelumnya (Susilowati *et al.*, 2023). Temuan ini sejalan dengan penelitian internasional yang membuktikan bahwa K-Means mampu menghasilkan pengelompokan yang bermakna pada data kanker payudara, baik data klinis maupun data kesehatan lainnya sehingga mendukung penggunaannya sebagai pendekatan *unsupervised learning* dalam eksplorasi pola data medis yang kompleks (Salloum, 2025). Dengan keunggulan tersebut, algoritma K-Means dinilai relevan dan sesuai digunakan dalam penelitian *data mining* kesehatan, khususnya untuk analisis data pasien kanker payudara.

Meski demikian, penelitian yang secara khusus mengelompokkan pasien kanker payudara berdasarkan karakteristik klinis maupun nonklinis menggunakan data rekam medis di Rumah Sakit Tingkat III Baladhika Husada Jember dengan pendekatan algoritma K-Means masih terbatas. Oleh karena itu, penelitian ini difokuskan pada penerapan algoritma K-Means untuk melakukan pengelompokan data pasien kanker payudara berdasarkan karakteristik yang tersedia. Tujuan dari proses klusterisasi ini

adalah untuk mengelompokkan pasien kanker payudara berdasarkan tingkat kemiripan karakteristik klinis maupun nonklinis yang dimiliki, sehingga diperoleh klaster pasien dengan karakteristik pasien serupa. Selanjutnya, hasil klaster tersebut diharapkan dapat digunakan sebagai dasar analisis data lebih lanjut serta sebagai referensi pendukung dalam pengambilan keputusan di bidang kesehatan, khususnya dalam upaya peningkatan kualitas pelayanan kesehatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini yakni “Bagaimana analisis klasterisasi pasien kanker payudara menggunakan algoritma K-Means di Rumah Sakit Tingkat III Baladhika Husada Jember?”.

1.3 Tujuan Penelitian

1.3.1 Tujuan Umum

Tujuan umum dari penelitian ini adalah untuk mengklasterisasi pasien kanker payudara menggunakan algoritma K-Means di Rumah Sakit Tingkat III Baladhika Husada Jember.

1.3.2 Tujuan Khusus

Adapun tujuan khusus dari penelitian ini, di antaranya:

- a. Melakukan *preprocessing* data dan menentukan *centroid* terbaik terhadap pasien kanker payudara di Rumah Sakit Tingkat III Baladhika Husada Jember.
- b. Mengidentifikasi karakteristik data pasien kanker payudara di Rumah Sakit Tingkat III Baladhika Husada Jember.
- c. Melakukan klasterisasi data pasien kanker payudara menggunakan algoritma K-Means di Rumah Sakit Tingkat III Baladhika Husada Jember.
- d. Melakukan evaluasi hasil klasterisasi data pasien kanker payudara menggunakan metode *Silhouette Coefficient*.

- e. Melakukan analisis dan pelabelan dari karakteristik masing-masing klaster pasien kanker payudara yang terbentuk sebagai hasil penerapan algoritma K-Means.

1.4 Manfaat Penelitian

1.4.1 Bagi Rumah Sakit

Memberikan gambaran pengelompokan pasien kanker payudara berdasarkan karakteristik klinis yang serupa sebagai bahan pertimbangan dalam pengelolaan data pasien, perencanaan pelayanan kesehatan, serta sebagai dasar awal pengambilan keputusan klinis dan kebijakan manajemen rumah sakit, khususnya terkait penanganan pasien kanker payudara.

1.4.2 Bagi Politeknik Negeri Jember

Dapat menambah referensi ilmiah dan bahan pembelajaran di bidang manajemen informasi kesehatan dan data mining kesehatan, khususnya terkait penerapan algoritma K-Means dalam analisis data klinis. Selain itu, penelitian ini dapat menjadi rujukan bagi penelitian selanjutnya dengan topik serupa.

1.4.3 Bagi Peneliti

Menambah pengetahuan dan pengalaman peneliti dalam pengolahan, analisis, dan pemanfaatan data rekam medis menggunakan metode data mining, khususnya teknik *clustering*. Penelitian ini juga menjadi sarana penerapan konsep teoritis yang diperoleh selama perkuliahan ke dalam permasalahan nyata di bidang kesehatan, serta mendukung pemenuhan kompetensi peneliti sebagai Perekam Medis dan Informasi Kesehatan (PMIK) dalam pengelolaan dan analisis data kesehatan sesuai dengan standar kompetensi yang berlaku.