

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit lambung merupakan salah satu gangguan pada pencernaan yang sering dijumpai dimasyarakat, dengan kondisi yang umum meliputi gastritis (radang lambung), tukak lambung (luka pada dinding lambung), dan GERD (*Gastroesophageal Reflux Disease*). Ketiga kondisi ini tidak hanya menimbulkan keluhan seperti nyeri ulu hati, mual, dan gangguan pencernaan, tetapi juga dapat memicu komplikasi serius jika tidak ditangani dengan tepat. Berbagai studi menunjukkan bahwa pola hidup tidak sehat menjadi pemicu utama timbulnya gangguan ini. Konsumsi makanan pedas, asam, berlemak, serta minuman berkafein secara berlebihan dapat mengiritasi lapisan lambung dan meningkatkan produksi asam lambung.

Selain itu, kebiasaan merokok, stres kronis, kurangnya aktivitas fisik, dan pola makan tidak teratur (seperti melewatkan sarapan atau makan larut malam) juga berkontribusi terhadap memburuknya kondisi lambung. Penelitian oleh Zhang *et al.*, (2021) mengungkapkan bahwa faktor gaya hidup seperti makan terburu-buru, tidur segera setelah makan, serta tingginya konsumsi makanan berlemak dan pedas secara signifikan meningkatkan risiko terjadinya GERD. Studi tersebut menekankan bahwa perubahan gaya hidup dapat menjadi pendekatan utama dalam pencegahan gejala lambung kronis. Selain itu, penelitian terbaru oleh Alatawi *et al.*, (2025) yang melibatkan lebih dari 900 responden dewasa di Arab Saudi menunjukkan hubungan yang signifikan antara kebiasaan merokok, konsumsi kafein, kurang olahraga, dan konsumsi makanan pedas lebih dari lima kali per minggu dengan kejadian GERD. Hasil ini memperkuat bukti bahwa faktor gaya hidup memiliki kontribusi nyata dalam memicu gangguan lambung dan menurunkan kualitas hidup pasien.

Sementara itu, secara fisiologis, sistem pencernaan memiliki peran vital dalam menjaga keseimbangan metabolisme tubuh, dan lambung merupakan organ utama yang berfungsi menyimpan serta mencerna makanan secara mekanik dan kimiawi sebelum nutrisi diserap di usus halus. Kerusakan atau gangguan pada

lambung dapat menghambat proses penyerapan nutrisi dan mengakibatkan masalah sistemik lain seperti anemia, penurunan berat badan, hingga gangguan imunitas. Oleh karena itu, penting untuk mengenali faktor risiko sejak dini dan menerapkan kebiasaan hidup sehat guna mencegah terjadinya gangguan lambung yang berkelanjutan.

Untuk membantu proses tersebut secara lebih sistematis dan efisien, teknologi informasi dapat berperan dalam mendukung proses identifikasi dan deteksi dini. Penerapan teknologi informasi, khususnya dalam bidang *data mining*, menjadi penting untuk mendukung deteksi dini dan pencegahan penyakit lambung. *Data mining* memungkinkan analisis terhadap pola hidup pasien untuk mengidentifikasi faktor-faktor risiko yang berkontribusi terhadap perkembangan penyakit lambung. Salah satu metode yang efektif dalam klasifikasi data adalah algoritma *decision tree*, dengan varian C5.0 yang dikenal memiliki keunggulan dalam hal kecepatan, akurasi, dan efisien memori.

Studi oleh Purwanti *et al.*, (2023) menunjukkan bahwa algoritma C5.0 dapat digunakan untuk mendeteksi risiko stroke dengan akurasi sebesar 84,42%, menggunakan 10 faktor risiko yang dianalisis melalui aplikasi web. Hal ini menunjukkan potensi algoritma C5.0 dalam mengklasifikasikan risiko penyakit berdasarkan data pola hidup pasien.

Algoritma C5.0 merupakan peningkatan dari metode *decision tree* C4.5 yang dikenal lebih unggul dalam hal kecepatan pemrosesan, akurasi hasil, serta efisiensi penggunaan memori, terutama saat menangani data dengan jumlah atribut yang besar. C5.0 menggunakan pendekatan *boosting* dan *pruning* yang lebih efisien untuk menghasilkan pohon keputusan yang tidak hanya akurat tetapi juga sederhana dan mudah diinterpretasikan. Keunggulan ini menjadikannya sangat cocok untuk aplikasi dalam bidang kesehatan, di mana interpretabilitas model sangat penting untuk mendukung pengambilan keputusan klinis.

Salah satu implementasi C5.0 yang relevan ditunjukkan dalam penelitian Lestari *et al.*, (2024) yang membandingkan algoritma ini dengan CART dalam klasifikasi risiko stroke berdasarkan data klinis dan gaya hidup pasien. Hasilnya menunjukkan bahwa C5.0 mampu memberikan akurasi yang lebih tinggi dan

struktur pohon yang lebih sederhana, menunjukkan kemampuannya dalam mengidentifikasi faktor risiko utama secara efisien. Dengan mempertimbangkan hal tersebut, penelitian ini bertujuan membangun model klasifikasi untuk mendeteksi risiko penyakit lambung berdasarkan pola hidup pasien secara lebih tepat, cepat, dan mudah diterapkan dalam praktik medis.

Algoritma C5.0 dipilih karena kemampuannya dalam mengolah data numerik dan kategorikal, kecepatan eksekusi yang tinggi, serta tingkat akurasi yang baik. Metode ini akan digunakan untuk mengklasifikasikan risiko penyakit lambung pasien ke dalam dua kategori, yaitu parah dan tidak parah, berdasarkan data real dari kebiasaan hidup sehari-hari pasien. Penelitian ini menggunakan pendekatan kuantitatif untuk mengukur tingkat akurasi model klasifikasi yang dibangun. Melalui pendekatan ini, diharapkan dapat dihasilkan sistem prediksi risiko yang tidak hanya tepat sasaran, tetapi juga dapat dijadikan alat bantu dalam pengambilan keputusan oleh tenaga medis atau pihak-pihak terkait dalam upaya pencegahan dan pengelolaan penyakit lambung.

Berdasarkan hal tersebut, penulis tertarik untuk melakukan penelitian dengan judul “Implementasi Metode C5.0 untuk Klasifikasi Risiko Penyakit Lambung Berdasarkan Pola Hidup Pasien”.

Selain kontribusi pada bidang akademik, penelitian ini juga diharapkan dapat memberi dampak praktis bagi sektor kesehatan. Dengan adanya sistem klasifikasi berbasis algoritma C5.0, proses *skrining* awal terhadap risiko penyakit lambung dapat dilakukan secara lebih efisien dan objektif. Ini akan sangat membantu dalam meningkatkan kesadaran masyarakat akan pentingnya gaya hidup sehat dan mendukung program promotif dan preventif dalam bidang kesehatan.

1.2 Rumusan Masalah

Berikut ini adalah rumusan masalah yang dirumuskan untuk membimbing penelitian dalam menganalisis risiko penyakit lambung menggunakan algoritma C5.0:

1. Apa yang dimaksud dengan model klasifikasi risiko penyakit lambung berbasis pola hidup pasien menggunakan algoritma C5.0?

2. Mengapa algoritma C5.0 dipilih untuk mengklasifikasikan risiko penyakit lambung ke dalam kategori parah dan tidak parah?
3. Bagaimana proses penerapan data mining dilakukan untuk mengolah data pola hidup pasien dalam mendeteksi dini risiko penyakit lambung menggunakan algoritma C5.0?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menerapkan algoritma C5.0 dalam menganalisis data pola hidup pasien guna:

1. Untuk membangun model risiko penyakit lambung berdasarkan pola hidup pasien menggunakan algoritma C5.0, yang mampu mengelompokkan pasien ke dalam dua kategori, yaitu risiko parah dan tidak parah, dengan memanfaatkan pola kebiasaan hidup seperti konsumsi makanan, kebiasaan merokok, stres, dan aktivitas fisik sebagai variabel prediktor.
2. Untuk mengevaluasi alasan pemilihan algoritma C5.0 dalam proses klasifikasi, mengingat keunggulannya dalam hal akurasi yang tinggi, efisiensi komputasi, serta kemampuannya menghasilkan aturan klasifikasi yang jelas dan mudah diinterpretasikan, dibandingkan dengan algoritma pohon keputusan lainnya seperti ID3 atau C4.5.
3. Untuk menerapkan tahapan proses *data mining* dalam pengolahan data pola hidup pasien, mulai dari pengumpulan data, pra-pemrosesan, pembentukan model klasifikasi menggunakan algoritma C5.0, hingga evaluasi performa model, dengan tujuan mendukung deteksi risiko penyakit lambung secara lebih sistematis dan berbasis data.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Penelitian ini memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *data mining*, khususnya dalam penerapan algoritma C5.0 untuk klasifikasi risiko penyakit berdasarkan data pola hidup. Selain itu, penelitian ini juga dapat menjadi referensi bagi peneliti selanjutnya yang ingin mengkaji pemanfaatan algoritma pohon keputusan dalam konteks kesehatan.

2. Hasil penelitian ini diharapkan dapat digunakan oleh tenaga medis atau instansi kesehatan sebagai alat bantu dalam proses *skrining* awal terhadap pasien, sehingga deteksi dini risiko penyakit lambung dapat dilakukan secara lebih efisien, akurat, dan objektif.
3. Dengan adanya sistem klasifikasi risiko berbasis pola hidup, masyarakat dapat lebih sadar terhadap pentingnya menjaga pola hidup sehat untuk mencegah penyakit lambung. Sistem ini juga mendukung upaya promotif dan preventif di bidang kesehatan masyarakat.

1.5 Ruang Lingkup Dan Batasan Penelitian

Ruang lingkup dan Batasan dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berasal dari data pola hidup pasien, seperti kebiasaan makan, konsumsi makanan pedas dan asam, kebiasaan merokok, tingkat stres, pola tidur, serta konsumsi kafein.
2. Proses klasifikasi dilakukan terhadap tingkat risiko penyakit lambung, yang dikelompokkan ke dalam dua kategori, yaitu parah dan tidak parah.
3. Tujuan klasifikasi adalah untuk membantu proses deteksi dini dan *skrining* awal risiko penyakit lambung, bukan untuk diagnosis klinis atau pengobatan langsung.
4. Metode yang digunakan adalah algoritma C5.0, tanpa kombinasi atau perbandingan dengan metode lain seperti CART, *Random Forest*, atau SVM.
5. Penelitian ini menggunakan bantuan perangkat lunak *Python* sebagai alat bantu pemrosesan data dan pembuatan model klasifikasi.
6. Data yang digunakan dalam penelitian ini hanya mencakup informasi mengenai pola hidup pasien, seperti kebiasaan makan, merokok, stres, dan tidur. Penelitian ini tidak menggunakan data medis mendalam seperti faktor genetik, hasil pemeriksaan *endoskopi*, atau data laboratorium klinis, fokus penelitian adalah pada gaya hidup sebagai faktor risiko.
7. Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 200 sampai 300 data, yang dikumpulkan dalam periode Juli hingga Januari 2026.

8. Atribut yang digunakan dalam klasifikasi meliputi variabel gaya hidup dan demografis, yaitu jenis kelamin, usia, pola makan, kebiasaan merokok, kebiasaan minum kopi, tingkat stres, dan konsumsi alkohol. Atribut lain seperti hasil pemeriksaan medis, riwayat penyakit sebelumnya, atau data laboratorium tidak disertakan.
9. Klasifikasi risiko dilakukan dalam dua kelas parah dan tidak parah, berdasarkan bobot risiko dari kombinasi gaya hidup tertentu yang dianalisis oleh algoritma C5.0.
10. Hasil klasifikasi hanya digunakan sebagai bahan pertimbangan awal dalam promosi kesehatan dan tindakan pencegahan, bukan sebagai keputusan final medis atau simulasi intervensi.