

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Penyakit jantung merupakan salah satu masalah kesehatan yang hingga kini masih menjadi perhatian serius, baik di tingkat global maupun di Indonesia. Penyakit jantung juga merupakan suatu kondisi dimana ketika jantung mengalami gangguan fungsi dalam pemompaan cairan darah tubuh[1]. Yang dimana penyakit jantung menjadi penyebab utama kematian di seluruh dunia termasuk di Indonesia. Dimulai dari perubahan gaya hidup yang buruk seperti pola makan yang menjurus ke makanan *fast food* yang tidak sehat dan tidak seimbang karena mengandung kalori, lemak, protein, dan garam tinggi tapi rendah serat pangan disinyalir menjadi faktor risiko meningkatnya prevalensi penyakit jantung.[2] Penyakit jantung biasanya ditandai dengan rasa nyeri atau tidak nyaman di dada, juga disertai sesak napas. Serta rasa sakitnya bisa menyebar ke lengan kiri, rahang, atau punggung. Adapun gejala lain yang muncul misalnya tubuh jadi mudah membiru, bengkak di sekitar mata atau tungkai, berat badan sulit untuk naik, sampai detak jantung terasa tidak teratur. Sehingga sering juga disertai demam, perut bengkak, atau batuk yang tidak kunjung reda.[3]

Oleh karna itu penelitian ini berfokus pada penerapan *Gain ratio* untuk seleksi fitur pada klasifikasi penyakit jantung guna untuk membedakan pasien berisiko dan tidak berisiko penyakit jantung. Dalam klasifikasi ini terdapat keuntungan yaitu untuk menyederhanakan data dengan hanya memilih fitur yang relevan dan mampu meningkatkan hasil akurasi dari hasil klasifikasi.

Berdasarkan pencatatan dari *World Health Organization* penyakit kardiovaskular menjadi penyebab kematian terbesar di dunia, dengan angka sekitar 17,9 juta jiwa per tahun. Jumlah ini diperkirakan terus meningkat hingga lebih dari 23,3 juta pada 2030.[4] Sedangkan di Indonesia, hasil Survei Kesehatan Indonesia (SKI) 2023 menunjukkan kasus penyakit jantung paling banyak terjadi pada kelompok usia 25–34 tahun dengan 140.206 pasien, sedikit lebih tinggi dibanding kelompok usia 15–24 tahun yang berjumlah 139.891 pasien. Data ini menegaskan bahwa penyakit jantung tidak hanya menyerang usia lanjut, tetapi juga kelompok usia muda yang produktif.[5] Dalam hal ini peran machine learning sangat dibutuhkan dalam mengklasifikasi resiko penyakit jantung. Dikarenakan algoritma tersebut dapat mengolah data medis yang kompleks dan besar untuk menghasilkan prediksi yang akurat dan efisien.[8]

Untuk itu, penelitian ini menggunakan metode *K-Nearest Neighbors* karena algoritma ini termasuk dasar dalam machine learning dan sering dipakai pada klasifikasi. *K-Nearest Neighbors* sendiri masuk dalam kategori *supervised learning* dan banyak dimanfaatkan untuk pengenalan pola, data mining, hingga deteksi intrusi.[9] Adapun perbandingan hasil dari beberapa metode pada kasus serupa yang dimana pada setiap metode pasti memiliki hasil yang berbeda-beda.

Untuk itu pada penelitian dengan metode *Naïve Bayes* menghasilkan akurasi 86.84%, yang masih tergolong baik meski sedikit lebih rendah dari *K-Nearest Neighbors*. [6] Sementara itu, pada penelitian dengan metode *K-Nearest Neighbors*, didapatkan akurasi sebesar 91.00%, menunjukkan bahwa metode ini cukup efektif untuk klasifikasi penyakit jantung.[7] Sedangkan metode *Random Forest* hanya

mencapai akurasi 75,00%, sehingga performanya tidak sebaik dua metode sebelumnya.[8]

Dari penelitian – penelitian sebelumnya yang dibandingkan oleh penulis diatas, bahwa metode dengan nilai akurasi yang sangat tinggi yaitu *K-Nearest Neighbors* yang dimana hasil tersebut sebesar 91,00% artinya metode tersebut sangat baik dalam klasifikasi penyakit jantung.

Seleksi fitur adalah proses pemilihan atribut yang sesuai dari kumpulan atribut yang besar.[9] Selain itu, seleksi fitur juga digunakan untuk mengurangi atribut data yang asing atau tidak relevan yang dihasilkan dari ekstraksi yang dapat memengaruhi hasil kinerja pembelajaran mesin.[10] Salah satu metode yang efektif dalam proses seleksi fitur adalah *Gain ratio*, karena dapat membantu menemukan atribut yang paling berpengaruh dan membuang atribut yang tidak relevan. *Gain ratio* merupakan modifikasi perkembangan dari seleksi fitur *information gain* dengan mengurangi daya biasnya. Penguatan memperhitungkan jumlah dan ukuran cabang pada waktu pemilihan parameter. Peningkatan akses ke informasi dengan memasukkan informasi internal (distribusi entropi) dalam perhitungan.[11] Selain itu, *Gain ratio* dapat memperbaiki data yang tidak stabil, oleh karena itu cenderung cocok pada data numerik dua kelas, sederhana sehingga komputasi lebih cepat.[12] Seleksi fitur *Gain ratio* juga dapat memperbaiki dengan mengambil informasi intrinsik dari atribut.[9]

Untuk itu menurut penulis, *Gain Ratio* bisa menunjukkan fitur mana yang paling berpengaruh terhadap hasil klasifikasi, yang dapat memudahkan dalam

analisis data. Dan metode yang dipilih dapat mengurangi bias terhadap fitur yang punya banyak nilai, sehingga hasil pemilihan fitur jadi lebih akurat.

Berdasarkan latar belakang yang sudah dijelaskan, maka peneliti tertarik untuk mengambil judul “PENERAPAN GAIN RATIO UNTUK SELEKSI FITUR PADA KLASIFIKASI PENYAKIT JANTUNG DENGAN MENGGUNAKAN K-NEAREST NEIGBORS”.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah di atas maka penulis dapat merumuskan masalah dari pembahasan tersebut, yaitu:

1. Bagaimana menerapkan KNN dengan seleksi fitur *Gain ratio* pada klasifikasi penyakit jantung?
2. Bagaimana menganalisis algoritma KNN pada klasifikasi penyakit jantung?

1.3 BATASAN MASALAH

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi:

1. Penelitian ini hanya akan fokus pada penggunaan metode yang digunakan pada klasifikasi penyakit jantung.
2. Penelitian ini akan menggunakan teknik seleksi fitur untuk meningkatkan akurasi klasifikasi.
3. Penelitian tidak membahas aspek diagnosis klinis secara langsung, melainkan fokus pada pemrosesan data dan klasifikasi menggunakan machine learning.

1.4 TUJUAN PENELITIAN

Berdasarkan pada permasalahan yang telah disampaikan sebelumnya yang terdapat di dalam penelitian ini. Memiliki tujuan-tujuan sebagai berikut:

1. Menerapkan metode *K-Nearest Neighbors* dengan seleksi fitur menggunakan *Gain Ratio* untuk klasifikasi penyakit jantung, sehingga hanya fitur yang relevan yang digunakan agar hasil prediksi lebih akurat.
2. Menganalisis kinerja algoritma *K-Nearest Neighbors* dalam klasifikasi penyakit jantung, terutama dari segi akurasi, kecepatan, dan keandalan prediksi.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang diperoleh dari penelitian ini adalah:

1. Membuat model menjadi lebih akurat dalam memilih fitur yang relevan, seperti tekanan darah atau kolesterol, serta proses klasifikasi menjadi lebih cepat dan hasil prediksi lebih tepat.
2. Memberikan ketepatan *K-Nearest Neighbors* dalam membantu memprediksi risiko penyakit jantung dengan membandingkan data pasien baru dengan data sebelumnya secara akurat.
3. Meningkatkan akurasi dan efisiensi prediksi pada fitur yang tidak relevan, sehingga *K-Nearest Neighbors* bekerja lebih tepat dan cepat dalam memprediksi risiko penyakit jantung. Serta mengurangi risiko kesalahan dalam diagnosis dengan hanya menggunakan data penting, sehingga prediksi

menjadi lebih andal dan membantu pengambilan keputusan medis yang lebih tepat.

4. Meningkatkan kesadaran masyarakat tentang pentingnya deteksi dini dan pencegahan penyakit jantung, terutama di kalangan kelompok usia muda.
5. Membantu peneliti dan praktisi di bidang data mining dan kesehatan dalam memahami dampak seleksi fitur terhadap algoritma klasifikasi, khususnya penggunaan *Gain Ratio* dan *K-Nearest Neighbors*, sejalan dengan tujuan menganalisis pengaruh metode ini terhadap kinerja klasifikasi.

1.6 SISTEMATIKA PENULISAN

Laporan penelitian ini dibuat dalam sistematika yang sesuai dengan kaidah penulisan yang benar dan dibagi dalam bab-bab sebagai berikut:

BAB I: PENDAHULUAN

Bab ini berisi pendahuluan yang membahas tentang latar belakang masalah, perumusan masalah, pembatasan dan ruang lingkup masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II: LANDASAN TEORI

Bab ini menjelaskan secara singkat mengenai teori-teori yang mendasari penelitian ini, termasuk penyakit jantung, seleksi fitur, *Gain ratio* metode *K-NEAREST NEIGHBORS*, serta studi-studi sebelumnya yang relevan dengan penerapan gain ratio untuk seleksi fitur pada klasifikasi penyakit jantung dengan menggunakan *K-NEAREST NEIGHBORS*.

BAB III: METODOLOGI PENELITIAN

Pada bab ini dijelaskan bagaimana cara-cara pelaksanaan penelitian yang mencakup metode pengumpulan data yang dilakukan selama penelitian, metode pengembangan model penerapan *Gain Ratio* untuk seleksi fitur pada klasifikasi penyakit jantung dengan menggunakan *K-NEAREST NEIGHBORS*, serta perangkat lunak yang digunakan dalam penelitian.

BAB IV: HASIL DAN PEMBAHASAN

Dalam bab ini memuat proses penelitian dari seleksi fitur untuk menentukan hasil dari penerapan *Gain Ratio* untuk seleksi fitur pada klasifikasi penyakit jantung dengan menggunakan *K-NEAREST NEIGHBORS*. Pembahasan meliputi pengolahan dataset, implementasi algoritma *K-NEAREST NEIGHBORS* menggunakan bahasa pemrograman dan alat bantu yang sesuai.

BAB V: PENUTUP

Bab ini merupakan penutup dari penelitian ilmiah ini yang berisi kesimpulan dari pembahasan bab-bab sebelumnya dan saran-saran yang berguna bagi pihak-pihak yang berkepentingan dalam penelitian ini serta kemungkinan pengembangan lebih lanjut dari penerapan *Gain Ratio* untuk seleksi fitur pada klasifikasi penyakit jantung dengan menggunakan *K-NEAREST NEIGHBORS*.