

BAB I PENDAHULUAN

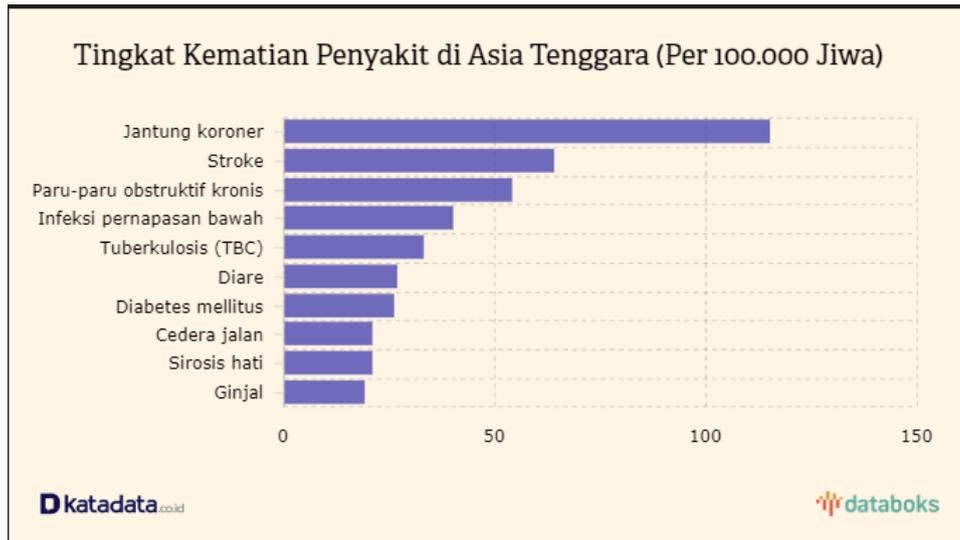
I.1 Latar Belakang

Penyakit Penyakit jantung merupakan penyakit yang paling mematikan dan menjangkit di dunia dengan jumlah penderita yang sangat besar. Setiap tahun, lebih dari 17,8 juta orang meninggal dunia akibat penyakit jantung (RI, 2021) (Lazulfa & Agustina, 2017) (Rahmat et al., 2023). Pada tahun 2022, menurut Kementerian Kesehatan Republik Indonesia (Kemenkes RI) jumlah kasus penyakit jantung telah mencapai 13,96 juta pasien dan menghabiskan dana JKN sekitar Rp10,9 triliun (Firdaus, n.d.). Menurut data dari Institute for Health Metrics and Evaluation (IHME), pada tahun 2019, terdapat 251,09 kematian akibat penyakit jantung per 100.000 penduduk di Indonesia. Angka tersebut menunjukkan peningkatan sebesar 1,25% dibandingkan dengan tahun sebelumnya, di mana jumlah kematian mencapai 247,99 per 100.000 orang (Ridhwan, n.d.). Dapat dilihat pada mengenai data kasus kematian penyakit di asia tenggara.



Gambar I.1 Kasus Penyakit Jantung Di Indonesia

Sumber (PT. Alpha, 2015)



Gambar I.2 Kasus Kematian Penyakit Di Asia Tenggara

Sumber Data : databoks.katadata.co.id

Pada dasarnya penyebab penyakit jantung sangat bervariasi, termasuk pola hidup yang tidak sehat, mengonsumsi makanan tinggi kolesterol, faktor lingkungan serta faktor keturunan atau genetik (p2ptm kemkes, 2024) (Atthohiroh et al., 2023) (Khaerunnisa, n.d.). Masalah ini disebabkan oleh sulitnya mendeteksi dini penyakit jantung, karena banyak orang sering mengabaikan gejala awal dan seringkali seseorang tidak menyadari potensi penyakit jantung yang ada pada dirinya (Maulana, n.d.). Kebanyakan orang enggan melakukan pemeriksaan kesehatan jantung karena terkendalanya waktu tunggu atau antrian untuk mendapatkan pelayanan kesehatan yang sesuai (Handayani et al., 2022), seperti di poli jantung RSUD Haji Provinsi Jawa Timur, yang mengalami kekurangannya jumlah dokter. Hal ini dapat menyebabkan meningkatnya risiko pasien terpapar virus lain (Meilinda & Gustini, 2021). Idealnya waktu tunggu pelayanan rawat jalan di rumah sakit adalah <60 menit berdasarkan Keputusan Menteri Kesehatan Republik Indonesia Nomor 129/Menkes/SK/II/2008 (Rahmat et al., 2023).

Untuk mengatasi permasalahan ini, diperlukannya sistem diagnosa yang cepat, akurat, dan efektif untuk mengurangi waktu tunggu atau antrian. Oleh karena itu, penelitian ini bertujuan membuat sistem diagnosa untuk melakukan prediksi kategori penyakit jantung guna membantu para dokter maupun perawat dalam mendiagnosa pasien. Salah satu caranya yaitu menggunakan data mining. Data

mining adalah pendekatan yang dapat digunakan untuk meningkatkan akurasi diagnosis dan membantu dalam mengategorikan penyakit jantung (Dimsyiar M Al Hafiz et al., 2021). Salah satu teknik data mining yang digunakan dalam penelitian ini adalah klasifikasi, dengan menggunakan algoritma Naive Bayes. Naive Bayes merupakan algoritma klasifikasi yang memiliki kemampuan cukup dalam membangun sebuah model pengklasifikasian (Atthohiroh et al., 2023). Metode klasifikasi yang digunakan dalam penelitian yaitu Naive Bayes yang diimplementasikan untuk mengetahui sebuah diagnosa penyakit jantung berdasarkan perhitungan probabilitas data historis penyakit jantung sebelumnya.

Sudah banyak penelitian lainnya yang serupa untuk memprediksi penyakit jantung dengan menggunakan metode yang sama, salah satunya penelitian yang dilakukan oleh Dimsyiar M Al Hafiz, Khoirul Amaly, Javen Jonathan, M. Teranggono Rachmatullah, Rosidi pada tahun 2023 dengan menggunakan metode yang sama memberikan akurasi sebesar 85,25%. Adapun juga penelitian lainnya yaitu oleh Tri Retnasari, Eva Rahmawati pada tahun 2023 dengan membandingkan 2 metode yaitu C4.5 dan Naive Bayes dengan hasil akurasi tertinggi dengan nilai accuracy sebesar 85.57%, nilai precision sebesar 86.54%, nilai recall sebesar 87.57% dan nilai f1-score sebesar 87.05%.

Sistem diagnosa yang dikembangkan ini, akan memanfaatkan data yang berasal dari rekam medik RSUD Haji Surabaya yaitu data pasien penyakit jantung tahun 2022. Terdapat 15.846 data pasien yang menderita penyakit jantung, yang telah dikumpulkan dari RSUD Haji Surabaya selama bulan Januari hingga Desember 2022. Data yang telah diperoleh berbagai atribut seperti, nomor rekam medis (no_cm), nama, usia, tanggal lahir, alamat, jenis kelamin, tanggal periksa, nama dokter, jenis pasien, keluhan, riwayat penyakit keluarga, tekanan darah, hdl kolesterol, ldl kolesterol, gula darah 2 jam setelah makan (2 JPP), Gda stik (gula darah secara acak), laju pernapasan (RR), tinggi badan, berat badan, suhu, diagnosa, serta catatan medis (soap_plan).

Dari data yang diperoleh akan dilakukan preprocessing data dimana menghilangkan atau membuang data atau atribut yang tidak diperlukan atau tidak relevan serta atribut yang relevan akan digunakan untuk memprediksi kategori

penyakit jantung. Hasil dari analisis ini akan digunakan untuk mengklasifikasikan pasien ke dalam kelas kategori penyakit jantung.

Model diuji melalui perhitungan akurasi terhadap hasil prediksi menggunakan confusion matrix dengan satuan accuracy, precision, recall, dan FI-score. Evaluasi model dibangun berdasarkan 6 proporsi pembagian data testing dan data training yaitu, 10:90, 20:80, 30:70, 40:60, 50:50 dan 60:40. Dengan menggunakan metode naive bayes diharapkan hasil dari model ini mampu memberikan nilai akurasi yang tinggi dan hasil akhir dari model ini akan berupa sistem diagnosa yang diimplementasikan kedalam website untuk melakukan prediksi kategori penyakit jantung. Dengan menggunakan google collab untuk melakukan perhitungan rumus prediksi dan penyiapan data serta mengeksekusi kode python di cloud, VSCode sebagai editor kode program dalam berbagai bahasa pemrograman seperti python, html, dan css. Hasil yang diharapkan dari sistem diagnosa kategori penyakit jantung berbasis web dengan metode naive bayes adalah mampu mendapatkan akurasi yang terbaik dengan membandingkan jenis Naive bayes yaitu *Gaussian* dan *Bernoulli*, sehingga sistem dapat membantu dan memudahkan dokter dalam mendiagnosa kategori penyakit jantung dengan cepat, tepat dan akurat.

I.2 Perumusan Masalah

Dengan mengetahui latar belakang yang telah dijabarkan diatas terdapat rumusan masalah dari “Penerapan CRISP-DM Untuk Memprediksi Kategori Penyakit Jantung Menggunakan Metode Naive Bayes (Studi Kasus : Poli Jantung RSUD Haji Surabaya)”.

1. Bagaimana memprediksi kategori penyakit jantung dengan menggunakan pendekatan data mining pada data pasien Poli Jantung RSUD Haji Surabaya?
2. Bagaimana perbedaan dan evaluasi hasil model yang diperoleh dari metode Gaussian Naive Bayes dan Bernoulli Naive Bayes pada data pasien Poli Jantung RSUD Haji Surabaya ?

3. Bagaimana membangun sistem diagnosis berbasis web untuk memprediksi kategori penyakit jantung menggunakan metode Naive Bayes pada Poli Jantung RSUD Haji Surabaya?

I.3 Tujuan Penelitian

Dengan rumusan masalah yang ada, tujuan yang ingin dicapai oleh peneliti dalam penelitian ini adalah:

1. Menerapkan metode CRISP-DM dengan algoritma Naive Bayes untuk memprediksi kategori penyakit jantung melalui pengembangan sistem diagnosis berbasis web.
2. Melakukan perhitungan dan membandingkan akurasi terbaik berdasarkan *test size* dan *random state* dari hasil sistem diagnosis kategori penyakit jantung dengan metode *Gaussian Naive Bayes* dan *Bernoulli Naive Bayes* menggunakan *confusion matrix*.
3. Membangun sistem diagnosis penyakit berbasis web untuk memprediksi penyakit jantung menggunakan algoritma Naive Bayes dengan menerapkan studi kasus pasien penyakit jantung di RSUD Haji Surabaya.

I.4 Batasan Penelitian

Pada penelitian yang akan dikerjakan terdapat beberapa batasan-batasan yang sudah ditentukan oleh peneliti dalam melakukan penelitian agar memiliki batas dalam mengerjakan penelitian, yaitu:

1. Penelitian ini dilakukan di RSUD Haji Surabaya, khususnya bagian Poli Jantung. Data rekam medik pasien, data training serta data testing yang digunakan adalah data pasien tahun 2022.
2. Atribut data rekam medik yang digunakan terdiri dari 12 atribut untuk modelling terdiri dari jenis kelamin, usia, keluhan, riwayat penyakit keluarga, *sistole*, *diastole*, laju pernapasan (RR), hdl kolesterol, ldl kolesterol, gula darah 2 jam setelah makan (2 JPP), Gda stik (gula darah secara acak), dan diagnosa.
3. Kategori yang digunakan dalam penelitian ini yaitu Penyakit Jantung Koroner (PJK), Hipertensi, dan Gagal Jantung

4. Pada sistem diagnosa berbasis web untuk memprediksi kategori penyakit jantung, tidak bisa dilakukan prediksi multi-label, sehingga hanya dapat memprediksi satu kategori penyakit jantung

I.5 Manfaat Penelitian

Manfaat yang akan didapatkan dari penelitian ini akan menjadi bantuan bagi pihak RSUD Haji Surabaya maupun masyarakat kedepannya, yaitu:

1. Manfaat bagi dokter
 - Membantu dokter dalam membuat keputusan diagnosa yang lebih cepat dan tepat berdasarkan data pasien.
 - Dokter dapat mengurangi risiko terpapar virus atau penyakit lain dengan mengefisien waktu pemeriksaan pasien.
2. Manfaat bagi pasien
 - Waktu tunggu pasien menjadi lebih cepat
 - Pasien dapat cepat tertangani
3. Manfaat bagi umum
 - Dapat menjadi bahan rujukan dalam pembuatan penelitian selanjutnya
 - Dapat menjadi bahan pembelajaran mengenai Data Mining dengan algoritma *Naive Bayes*.

I.6 Metodologi Penelitian

Metodologi penelitian ini terdiri dari serangkaian langkah sistematis yang dirancang untuk mencapai tujuan dalam penelitian ini. Langkah awal yaitu melakukan identifikasi masalah terkait prediksi kategori penyakit jantung pada pasien RSUD Haji Surabaya yang menjadi fokus penelitian. Selanjutnya, melakukan pencarian studi literatur dari jurnal dan sumber web sebagai sumber acuan atau landasan untuk mendapatkan pemahaman yang mendalam terkait penelitian yang akan dilakukan.

Langkah berikutnya melakukan observasi dan wawancara dengan bidang diklit dan rekam medik RSUD Haji Surabaya terkait data yang akan digunakan. Serta melakukan wawancara pada dokter spesialis jantung untuk mendapatkan wawasan dan informasi mengenai informasi dan data apa saja yang dapat

digunakan sebagai bahan penguat penelitian. Melakukan pengumpulan data yang diperoleh dari rekam medik sebagai bahan penelitian. Melakukan preprocessing data untuk membuang atau menghilangkan data yang tidak digunakan.

Implementasi algoritma naive bayes sebagai metode klasifikasi penelitian. Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur keakuratan dan kinerja model dan validasi model dengan melibatkan dokter. Langkah terakhir dari metodologi ini adalah membangun sistem diagnosis berbasis web untuk memprediksi kategori penyakit jantung.