

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Penyakit jantung koroner merupakan salah satu penyakit penyebab utama kematian pada semua kelompok umur setelah stroke. Penyakit jantung koroner adalah penyakit yang disebabkan oleh plak yang menumpuk di arteri koroner yang memasok oksigen ke otot jantung. Data mining dapat diterapkan di bidang kesehatan. Keputusan dan prediksi mengenai penyakit yang menjadi sasaran akan dihasilkan dari pengambilan keputusan berdasarkan data dan informasi yang akurat[1].

Menurut Organisasi Kesehatan Dunia (WHO), tidak kurang dari 17,9 juta orang meninggal akibat penyakit kardiovaskular pada tahun 2019, atau 32% dari seluruh kematian global. 85 % dari kematian ini disebabkan oleh serangan jantung dan stroke. Kemudian, pada tahun 2019, 17 juta kematian dini (dibawah usia 70 tahun) disebabkan oleh penyakit tidak menular (PTM), 38% di antaranya disebabkan oleh penyakit kardiovaskular[2].

Penyakit jantung koroner merupakan masalah kesehatan dapat memberikan dampak pada faktor sosial-ekonomi, karena perawatan dan pengobatan yang membutuhkan biaya yang mahal sehingga diperlukan upaya pencegahan dan pengendalian sejak dini sangat diperlukan. Beberapa faktor yang menyebabkan penyakit jantung koroner yaitu: usia, jenis kelamin, hipertensi, obesitas, diabetes

mellitus, stroke, merokok, penyakit arteri perifer, letak geografis, makanan tinggi lemak dan alkohol, hipertensi dan dislipidemia.

Penyakit jantung koroner terjadi akibat kerusakan pada arteri koroner yang menyuplai darah kaya oksigen ke otot jantung. Akibatnya aliran darah ke jantung berkurang mengakibatkan kerusakan otot jantung dan kehilangan kemampuan untuk memompa darah dengan efisien. Komplikasi serius terjadi seperti nyeri dada, serangan jantung, dan gagal jantung bisa terjadi. Deteksi dini dan pengobatan tepat sangat penting untuk menjaga fungsi jantung tetap optimal.

Hasil Riset Kesehatan Dasar Tahun 2019 menunjukkan bahwa Prevalensi Penyakit jantung di Indonesia sebesar 1,5% atau 15 dari 1.000 penduduk yang menderita penyakit jantung koroner, dengan jumlah kasus terbanyak ada di daerah Kalimantan Utara sebesar 2,2%. Sumatera Barat menempati pada urutan ke-10 dengan jumlah kasus penyakit jantung koroner sebesar 1,6%. Angka kematian yang disebabkan oleh Penyakit Jantung Koroner di Indonesia cukup tinggi mencapai 1,25 juta jiwa jika populasi penduduk Indonesia 250 juta jiwa[3].

Penyakit jantung koroner memiliki angka kematian yang tinggi dan merupakan penyakit yang sangat ditakuti oleh masyarakat. Oleh karena itu, klasifikasi data mining sangat diperlukan untuk membantu dokter dalam pengambilan keputusan yang cepat dan akurat dalam menangani penyakit ini.

Klasifikasi termasuk ke dalam *supervised learning* karena dalam proses klasifikasi terdapat proses pembelajaran lampau. Proses ini digunakan algoritma untuk mengenali pola dari data yang nantinya akan diterapkan kepada data baru yang belum diketahui kelompoknya. Beberapa algoritma yang digunakan untuk

proses klasifikasi, diantaranya adalah K-Nearest Neighbor (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), Random Forest (RF), serta Ada Boost Classifier).

Algoritma RF (Random Forest) algoritma dalam machine learning yang digunakan untuk pengklasifikasian data set dalam jumlah besar.

Adapun beberapa penelitian yang melakukan penelitian dalam klasifikasi suatu penyakit seperti halnya penelitian [4], yang menerapkan algoritma Random Forest untuk klasifikasi penyakit Store. Hasil penelitian ini memperlihatkan hasil akurasi sebesar 95,2%. Selain itu, penelitian yang dilakukan [5], dengan menggunakan algoritma Random Forest untuk Klasifikasi Penyakit Diabetes. Hasil penelitian ini memperlihatkan hasil akurasi sebesar 97,88%. Selain itu, penelitian yang dilakukan [6], dengan menggunakan algoritma Random Forest untuk Klasifikasi Penyakit Kanker Kulit. Hasil penelitian ini memperlihatkan hasil akurasi sebesar 85%.

Berdasarkan kajian penelitian sebelumnya Algoritma Random Forest dalam klasifikasi suatu penyakit menghasilkan nilai akurasi yang sangat baik dan tinggi. Oleh karena itu, pada penelitian ini dalam mengklasifikasi penyakit Jantung Koroner penulis menggunakan Algoritma Random Forest karena algoritma tersebut merupakan algoritma yang baik dalam mengklasifikasi penyakit dan menghasilkan nilai akurasi yang tinggi.

Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna memberi solusi terhadap masalah yang terjadi dengan mengakat judul
”PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT

JANTUNG KORONER MENGGUNAKAN ALGORITMA RANDOM FOREST”

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana Menerapkan *Data Mining* Untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Random Forest?
2. Seberapa besar tingkat akurasi Algoritma Random Forest apabila digunakan untuk mengklasifikasi penyakit jantung koroner?

1.3 BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu:

1. Pada penelitian ini peneliti hanya menggunakan metode klasifikasi dengan algoritma Random Forest
2. Data yang diperlukan adalah dataset Penyakit Jantung Koroner (data penyakit jantung koroner yang telah didapat berjumlah 4.239 data dan tahun 2019), dataset diambil dari website kaggle.com yang dapat didownload dengan link: (<https://www.kaggle.com/code/adithyabshetty100/coronary-heart-disease-prediction/notebook>).
3. Variabel yang digunakan dalam penelitian ini adalah gender, age, education, current smoke, cigs per day, bp meds, prevalent stroke, prevalent hypertension, diabetes, total cholesterol, sys, bp, dia bp, bmi, heart rate,

glucose dan ten year chd yang memberikan informasi mengenai penyakit jantung koroner.

4. Alat evaluasi yang digunakan meliputi Python, Scikit-learn, dan Pandas
5. Evaluasi dilakukan dengan mengukur kinerja model berdasarkan metrik akurasi, *precision*, *recall*.
6. Evaluasi performa menggunakan *confusion matrix*
7. Penelitian ini akan menggunakan pustaka scikit-learn dalam pemrograman Python
8. Data *preprocessing* terbatas pada Cleaning Data, Transformasi Data, Selection Data, Data *Reduction* ,Data Split
9. Alat bantu (tools) yang digunakan penelitian ini adalah google colab, rapidminer

1.4 TUJUAN PENELITIAN

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Menerapkan Algoritma Random Forest untuk mengklasifikasi penyakit jantung koroner maupun bukan penyakit jantung koroner.
2. Mengetahui berapa besar tingkat akurasi Algoritma Random Forest dalam klasifikasi Penyakit jantung koroner
3. Mengevaluasi kinerja model yang dibangun berdasarkan metrik akurasi, *precision*, *recall*.

1.5 MANFAAT PENELITIAN

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Dapat mengetahui status penderita Jantung koroner dengan melihat tingkat akurasi yang tinggi
2. Dapat memberikan kemudahan dalam mengenali penyakit jantung koroner .
3. Mendapatkan akurasi yang terbaik dalam klasifikasi penyakit jantung koroner dengan menggunakan tools Google Colab, Rapidminer sebagai alat bantu klasifikasi.
4. Dapat digunakan sebagai acuan untuk penelitian selanjutnya.

1.6 SISTEMATIKA PENULISAN

Sistematika dibuat untuk memberi suatu gambaran yang jelas terhadap penulisan ilmiah agar menghindari terjadinya pembahasan diluar tema dan judul penelitian. Secara garis besar penulisan ilmiah terdiri dari 5 bab, dapat dilihat dari sistematikan penulisan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

Dalam landasan teori ini, bab memuat kosep-konsep teori yang relevan sebagai kerangka atau dasar yang digunakan untuk mendukung pemahaman yang penulis lakukan.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini berisi metode pengumpulan data, prosedur penelitian, dan metode analisis berupa pendekatan penyelesaian masalah yang dilakukan untuk mendukung penelitian.

BAB IV : ANALISIS DAN HASIL

Pada bab ini berisi tentang gambaran umum tentang data yang akan di analisis, serta hasil analisis dari pengujian data yang dilakukan dan hasil yang dicapai dari pengujian data yang telah ditentukan.

BAB V : PENUTUP

Bab ini disajikan simpulan dari hasil penelitian dan juga saran-saran yang ditujukan kepada semua pihak yang bersangkutan.