

BAB I

PENDAHULUAN

1.1. Latar Belakang

Penyakit Tidak Menular (PTM) menjadi masalah kesehatan utama di Indonesia, dengan proporsi kematian hingga mencapai 78%. Penyakit tidak menular ini terutama terjadi di negara-negara berpendapatan rendah dan menengah [1]. Diperkirakan bahwa Indonesia mengalami potensi kerugian total sebesar 4,47 triliun dolar AS dari tahun 2012 hingga 2030 akibat penyakit tidak menular. Tingginya prevalensi PTM dapat menyebabkan meningkatnya permintaan akan pelayanan kesehatan, pengobatan yang lebih mahal, dan peningkatan belanja kesehatan, yang pada gilirannya dapat mengurangi anggaran yang tersedia untuk investasi dalam kegiatan yang lebih produktif [2].

PTM tidak dapat ditularkan secara langsung dari orang ke orang, faktor-faktor seperti gaya hidup, lingkungan, dan genetik menjadi pemicu utama. Di antara jenis utama PTM, penyakit kardiovaskular (*Cardiovascular Disease/CVD*) menjadi penyebab kematian utama di Indonesia dengan proporsi mencapai 43%, diikuti kanker 11%, penyakit pernapasan kronis 6%, diabetes 3%, dan PTM lainnya 15%. Persentase kematian akibat CVD di bawah usia 70 tahun mencapai 46%, ini menunjukkan dampak yang signifikan terhadap populasi usia produktif. Penyakit kardiovaskular menjadi penyebab utama kematian di seluruh dunia, dengan angka kematian mencapai 17,9 juta jiwa per tahun. Di tingkat global, penyakit kardiovaskular menyumbang sekitar 32% dari total kematian, dari jumlah tersebut disebabkan oleh serangan jantung dan *stroke* [3]. Penyakit kardiovaskular mencakup berbagai gangguan pada jantung dan pembuluh darah, seperti penyakit jantung koroner, serebrovaskular, hipertensi, penyakit arteri perifer, penyakit jantung rematik, penyakit jantung bawaan, trombosis vena dalam dan emboli paru [4]. Penyakit jantung memiliki variasi yang luas dan dapat

menyebabkan komplikasi yang beragam, termasuk penurunan kualitas hidup hingga kematian [5].

Faktor-faktor utama penyebab penyakit kardiovaskular seperti konsumsi makanan tidak sehat, kurangnya aktivitas fisik atau olahraga yang membakar kalori, mengonsumsi alkohol, merokok, dan tingkat stres. Faktor risiko perilaku, seperti pola makan tidak sehat dan kurangnya aktivitas fisik, bersama dengan kebiasaan merokok dan konsumsi alkohol, menjadi perhatian utama dalam upaya pencegahan penyakit kardiovaskular [6]. Faktor utama terjadinya penyakit kardiovaskular dapat dibagi menjadi dua kelompok besar yaitu faktor risiko yang tidak dapat diubah (usia, jenis kelamin, dan riwayat keluarga) dan faktor risiko yang dapat diubah (obesitas, merokok, pola makan yang kurang sehat, kurangnya aktivitas fisik, konsumsi alkohol, dan kadar lipid yang tidak normal dalam tubuh). Semakin banyak faktor risiko yang terdapat pada seseorang, semakin tinggi kemungkinannya untuk mengalami penyakit kardiovaskular [7]. Oleh karena itu, analisis menyeluruh terhadap faktor risiko ini menjadi krusial untuk mengidentifikasi pengaruh maupun korelasi yang terkait dengan penyakit kardiovaskular. Beberapa wilayah di Indonesia, seperti Provinsi Jawa Tengah, mencatat prevalensi penyakit kardiovaskular yang tinggi, menunjukkan pentingnya pemahaman yang lebih mendalam tentang faktor-faktor risiko yang berkontribusi pada penyakit tersebut.

Provinsi Jawa Tengah termasuk dalam lima besar wilayah dengan prevalensi kardiovaskular (hipertensi) tertinggi [8]. Pada tahun 2023, jumlah estimasi penderita hipertensi di Jawa Tengah yang berusia di atas 15 tahun mencapai 8.554.672 orang atau sekitar 38,2% dari populasi usia tersebut. Angka ini mengalami peningkatan dibandingkan tahun sebelumnya, yang mengindikasikan bahwa hipertensi merupakan masalah kesehatan serius di wilayah ini [9].

Berdasarkan data lima tahun terakhir, menunjukkan tren prevalensi hipertensi yang cukup bervariasi, namun jumlah penderitanya tetap tinggi. Pada tahun 2019, tercatat 8.070.378 penderita, diikuti oleh 8.525.593 pada tahun 2020, 8.700.512 pada tahun 2021, dan 8.494.296 pada tahun 2022. Dibandingkan

dengan penyakit tidak menular lainnya, seperti diabetes, penyakit pernapasan, dan kanker, hipertensi memiliki prevalensi tertinggi di Jawa Tengah [9]. Hal ini menjadikan hipertensi sebagai masalah kesehatan utama yang perlu mendapatkan perhatian khusus. Oleh karena itu, pemilihan Jawa Tengah sebagai lokasi penelitian ini sangat tepat untuk memahami lebih dalam mengenai karakteristik geografis dan faktor-faktor risiko yang berpengaruh terhadap penyakit kardiovaskular dengan memanfaatkan teknologi *machine learning* atau pembelajaran mesin.

Pada penelitian ini, *machine learning* digunakan untuk menganalisis karakteristik geografis dan faktor risiko penyakit tidak menular, khususnya penyakit kardiovaskular di Jawa Tengah. *Machine learning* dapat mengidentifikasi pola yang kompleks dan hubungan tersembunyi dalam data yang mungkin tidak terlihat dengan metode statistik tradisional atau analisis manual [10]. *Machine learning* dan teknik *data mining* banyak digunakan dalam analisis data medis dan mengekstrak informasi yang berharga. Berbagai pendekatan *machine learning*, seperti prediksi, klasifikasi, dan pengenalan pola dapat memprediksi maupun menganalisis masalah kardiovaskular [11]. Keunggulan utama dari *machine learning* adalah kemampuannya untuk memodelkan hubungan antara berbagai variabel dalam *dataset*, sehingga memungkinkan untuk mengeksplorasi dan memahami pengaruh maupun hubungan kompleks antara faktor-faktor risiko yang berpotensi memengaruhi prevalensi penyakit kardiovaskular. Teknik pembelajaran mesin dapat digunakan untuk mengekstrak data yang berharga dari kumpulan data yang besar. Algoritma pembelajaran mesin juga dapat digunakan untuk mengatasi kumpulan data yang tidak seimbang dan pemilihan fitur yang dapat meningkatkan akurasi [12].

Decision tree merupakan salah satu metode atau algoritma dalam *machine learning* yang populer untuk analisis data dan pengambilan keputusan. Struktur *decision tree* terdiri dari simpul akar (*root node*), cabang (*branches*), simpul internal (*internal nodes*), dan simpul daun (*leaf nodes*). Simpul akar (*root node*) merepresentasikan pertanyaan atau masalah yang ingin diselesaikan, sedangkan cabang-cabangnya merupakan jalur keputusan yang mengarah ke berbagai

alternatif jawaban atau keputusan (*internal node*). Setiap *decision tree* memiliki beberapa *internal node* yang mewakili alternatif jawaban atau keputusan dan *internal node* tersebut juga dapat memiliki cabang node lain yaitu *leaf node*, yang merupakan keputusan akhir. *Decision tree* termasuk salah satu algoritma yang mampu memilih atribut terbaik untuk memaksimalkan informasi yang diperoleh dalam pengambilan keputusan. Selain membantu dalam pengambilan keputusan dan memberikan gambaran yang jelas mengenai alasan di balik keputusan tersebut, *decision tree* juga sangat berguna untuk keperluan data *mining*. Kelebihan *decision tree* termasuk kemampuannya yang fleksibel dalam menangani tugas klasifikasi dan regresi, serta cocok untuk berbagai jenis data seperti data diskrit, data kontinu, dan data kategorik [13], [14].

Penelitian ini menggunakan *decision tree* jenis CART (*Classification And Regression Tree*), yang digunakan untuk pengambilan keputusan berkaitan dengan klasifikasi dan regresi. Pada klasifikasi, CART menghasilkan pohon keputusan yang memetakan pengamatan ke dalam kelas atau kategori. Sedangkan pada regresi, CART menghasilkan pohon keputusan yang memprediksi nilai numerik atau kontinu dari variabel target [15]. Penelitian ini akan berfokus pada regresi atau *regression tree*, pohon regresi yang dihasilkan digunakan untuk memprediksi nilai numerik atau kontinu dari variabel target berdasarkan hubungan dengan variabel independen. Melalui algoritma CART, pohon regresi ini dapat digunakan untuk memahami hubungan antara variabel independen (*predictors*) dengan variabel dependen (*target*).

Adanya penerapan teknologi *machine learning* dengan menggunakan metode CART, diharapkan dapat diperoleh pemahaman yang lebih baik tentang bagaimana berbagai faktor risiko berkontribusi terhadap prevalensi penyakit kardiovaskular. Faktor risiko yang memiliki pengaruh terbesar akan diidentifikasi melalui analisis *feature importance*. Faktor risiko dengan *feature importance* yang lebih tinggi menunjukkan bahwa faktor tersebut memiliki kontribusi yang lebih besar dalam menyebabkan penyakit kardiovaskular. Sebaliknya, faktor dengan *feature importance* lebih rendah memberikan kontribusi yang lebih kecil. Selain itu, penelitian ini juga akan disertai visualisasi pemetaan hasil analisis secara

geografis, sehingga dapat mengetahui distribusi geografis dari faktor-faktor risiko penyakit kardiovaskular. Hasil dari penelitian ini diharapkan tidak hanya bermanfaat bagi para praktisi kesehatan dan pembuat kebijakan di Jawa Tengah, tetapi juga bagi penelitian serupa di wilayah lain yang menghadapi tantangan kesehatan yang serupa. Dengan demikian, penelitian ini dapat menjadi langkah penting dalam upaya pencegahan dan pengendalian PTM, serta meningkatkan kualitas hidup masyarakat.

1.2. Rumusan Masalah

Tingginya proporsi kematian akibat Penyakit Tidak Menular (PTM), terutama Penyakit Kardiovaskular di Indonesia, menunjukkan urgensi untuk memahami faktor-faktor risiko yang berpengaruh terhadap prevalensi penyakit ini. Jawa Tengah, sebagai salah satu dari lima provinsi dengan prevalensi hipertensi tertinggi, dipilih sebagai fokus wilayah penelitian ini. Berdasarkan data Profil Kesehatan Jawa Tengah 2023, menunjukkan bahwa hipertensi sebagai salah satu jenis penyakit kardiovaskular, memiliki jumlah penderita yang paling tinggi dibandingkan dengan jenis PTM lainnya. Dengan demikian, penelitian ini perlu dilakukan untuk mengidentifikasi dan memberikan informasi mengenai faktor risiko yang paling berpengaruh terhadap penyakit kardiovaskular di Jawa Tengah, dengan menggunakan metode CART, sehingga dapat ditemukan langkah-langkah preventif yang lebih efektif dalam pengelolaan dan pencegahan penyakit kardiovaskular. Selain itu, penelitian ini juga akan disertai visualisasi pemetaan hasil analisis secara geografis, sehingga dapat mengetahui distribusi geografis dari faktor-faktor risiko penyakit kardiovaskular. Pemetaan ini akan dibuat menggunakan bahasa pemrograman Python dan memanfaatkan *library* seperti GeoPandas untuk membaca dan mengolah data geospasial.

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan di atas, maka penelitian ini bertujuan untuk menganalisis faktor risiko yang paling berpengaruh terhadap penyakit kardiovaskular di Jawa Tengah menggunakan metode CART.

1.4. Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut :

1. Sesuai data dari World Health Organization (WHO), faktor-faktor risiko yang menjadi fokus adalah konsumsi tidak sehat, kurangnya aktivitas fisik, konsumsi alkohol, merokok, tingkat stres, dan akses transportasi publik.
2. Cakupan atau ruang lingkup dari *dataset* faktor-faktor risiko terbatas pada tempat atau lokasi yang terdeteksi saat proses *scraping* menggunakan ekstensi Instant Data Scraper. Jika ada tempat yang tidak terdeteksi selama proses tersebut, maka tempat tersebut tidak akan dimasukkan dalam *dataset*.
3. Informasi yang dikumpulkan melalui proses *scraping* hanya mencakup nama tempat dan lokasinya di kabupaten/kota Jawa Tengah. Penelitian ini tidak mencakup karakteristik geografis maupun demografis masing-masing wilayah dalam analisisnya.
4. Penelitian ini menggunakan metode CART untuk menganalisis dan mengidentifikasi faktor risiko yang paling berpengaruh terhadap penyakit kardiovaskular di Jawa Tengah.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini dapat dijelaskan sebagai berikut:

Mengetahui faktor risiko yang paling berpengaruh terhadap penyakit kardiovaskular, sehingga memungkinkan pihak terkait untuk merancang program-program intervensi yang lebih efektif dan efisien untuk mengurangi prevalensi penyakit kardiovaskular di Jawa Tengah.

1.6. Sistematika Penulisan

Penulisan penelitian ini disusun secara sistematis dalam beberapa bab untuk memberikan gambaran yang jelas mengenai proses, analisis, dan hasil penelitian. Berikut adalah sistematika penulisannya:

1. BAB I PENDAHULUAN

Bab ini memuat latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan dan asumsi penelitian, manfaat penelitian, serta sistematika penulisan. Bagian ini memberikan dasar dan alasan pentingnya penelitian dilakukan.

2. BAB II LANDASAN TEORI

Bab ini membahas teori-teori dan literatur yang relevan dengan penelitian, termasuk definisi penyakit kardiovaskular, faktor-faktor risiko yang memengaruhi, metode analisis CART, dan pentingnya analisis geografis dalam konteks kesehatan masyarakat.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk sumber data, diagram alir penelitian, teknik pengumpulan data (*scraping* Google Maps dan data dari Dinas Kesehatan Jawa Tengah), serta metode analisis data menggunakan CART.

4. BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini menjelaskan proses pengumpulan data dan penjelasan langkah-langkah pengolahan data untuk memastikan data siap dianalisis.

5. BAB V ANALISIS DAN PEMBAHASAN

Bab ini menyajikan hasil analisis faktor risiko penyakit kardiovaskular menggunakan metode CART. Selain itu, terdapat pemetaan distribusi geografis faktor risiko penyakit kardiovaskular di Jawa Tengah, serta pembahasan terkait hasil yang diperoleh.

6. BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan serta saran yang dapat diberikan berdasarkan hasil penelitian.