

BAB I PENDAHULUAN

1.1 Latar Belakang

Hipertensi merupakan kondisi peningkatan tekanan darah sistolik ≥ 140 mmHg dan/atau diastolik ≥ 90 mmHg yang terjadi secara menetap. Kondisi ini mencerminkan adanya gangguan keseimbangan hemodinamik pada sistem kardiovaskular yang disebabkan oleh berbagai faktor sehingga tidak dapat dijelaskan oleh satu penyebab tunggal. Hipertensi tidak hanya meningkatkan risiko penyakit jantung, tetapi juga berpotensi menimbulkan gangguan pada organ lain seperti ginjal, sistem saraf, dan pembuluh darah, di mana risiko komplikasi akan semakin besar seiring meningkatnya tekanan darah [1].

Penyebab pasti terjadinya hipertensi hingga saat ini masih belum diketahui secara pasti. Namun, terdapat sejumlah faktor risiko yang berperan dalam terjadinya hipertensi, antara lain usia, jenis kelamin, obesitas, kebiasaan merokok, serta kurangnya aktivitas fisik. Selain itu, hipertensi juga dipengaruhi oleh perilaku hidup tidak sehat seperti kurang berolahraga, konsumsi alkohol berlebihan, stres berkepanjangan, serta rendahnya konsumsi buah dan sayur yang dapat memicu peningkatan tekanan darah. Apabila kondisi ini tidak ditangani secara tepat, hipertensi dapat menimbulkan komplikasi serius seperti stroke, dan kerusakan ginjal [2].

Secara global, hipertensi masih menjadi masalah kesehatan utama. Data World Health Organization (WHO) tahun 2015 menunjukkan bahwa sekitar 1,13 miliar penduduk dunia menderita hipertensi, atau sekitar satu dari tiga orang dewasa terdiagnosis hipertensi [1]. Di Indonesia, prevalensi hipertensi tergolong tinggi, yaitu mencapai 34%, dengan estimasi jumlah penderita sebanyak 63.309.620 orang. Selain itu, pada tahun 2018 tercatat sebanyak 427.218 kematian di Indonesia yang berkaitan dengan hipertensi, menunjukkan bahwa penyakit ini masih menjadi penyebab utama morbiditas dan mortalitas [2].

Perkembangan teknologi yang semakin pesat telah memberikan dampak positif dalam bidang kesehatan, terutama dalam membantu proses analisis dan deteksi penyakit. Pemanfaatan kecerdasan buatan melalui pendekatan *machine learning* memungkinkan sistem komputer untuk mempelajari pola dari data kesehatan secara otomatis tanpa perlu pemrograman berulang. Teknologi ini mampu mengolah data dalam jumlah besar secara efektif sehingga dapat menghasilkan informasi dan prediksi yang lebih akurat.

Klasifikasi adalah salah satu metode dalam *machine learning* yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola atau atribut yang dimiliki. Dalam bidang kesehatan, klasifikasi membantu memprediksi status kesehatan pasien, termasuk

identifikasi risiko penyakit seperti hipertensi, sehingga memungkinkan tenaga medis memberikan penanganan yang lebih cepat dan tepat [3].

Dalam konteks klasifikasi penyakit, berbagai algoritma *machine learning* dapat diterapkan. Salah satu algoritma yang banyak digunakan adalah *Decision Tree C4.5*. Algoritma ini merupakan pengembangan dari ID3 dan digunakan untuk membangun pohon keputusan yang fleksibel dan mudah dipahami. Pohon keputusan sendiri merupakan model prediksi yang mengubah data menjadi aturan keputusan (if-then rules). C4.5 memiliki beberapa keunggulan, antara lain mampu menangani atribut numerik maupun kategorikal, mengelola data dengan nilai hilang (missing value), dan mendukung pruning untuk mengurangi kompleksitas pohon sehingga mengurangi risiko *overfitting*. Selain itu, aturan keputusan yang dihasilkan mudah dipahami tenaga medis, sehingga algoritma ini sangat sesuai untuk klasifikasi penyakit hipertensi [4].

Pada penelitian sebelumnya, telah dilakukan perbandingan antara algoritma data mining, seperti *Decision Tree C4.5* dan CART, untuk menentukan metode dengan akurasi terbaik menggunakan dataset penyakit stroke. Hasil penelitian menunjukkan bahwa C4.5 menghasilkan akurasi sebesar 95,76%, sedangkan CART memperoleh akurasi sebesar 95,11% [5]. Penelitian lain yang membandingkan C4.5 dengan Naive Bayes pada dataset hipertensi memperoleh akurasi 74% untuk C4.5 dan 67% untuk Naive Bayes [6]. Meskipun demikian, penelitian yang menggunakan C4.5 pada dataset hipertensi lokal masih terbatas, sehingga performa algoritma pada kondisi dan fitur pasien lokal perlu dievaluasi lebih lanjut. Penelitian ini menggunakan dataset pasien lokal dari UPT Puskesmas Modopuro, Kecamatan Mojosari, Kabupaten Mojokerto, Jawa Timur, untuk mengevaluasi algoritma C4.5 dalam klasifikasi penyakit hipertensi.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree C4.5* dalam klasifikasi penyakit hipertensi menggunakan dataset lokal, serta mengevaluasi performa model agar dapat mendukung deteksi dini dan pengambilan keputusan medis secara efektif.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Penerapan algoritma klasifikasi seperti *Decision Tree C4.5* sering kali dihadapkan pada masalah *overfitting* jika kedalaman pohonnya tidak dibatasi, serta ketidakstabilan akurasi jika proporsi pembagian data tidak tepat. Selain itu, model komputasi tidak hanya dituntut memiliki akurasi tinggi, tetapi juga harus menghasilkan aturan keputusan yang logis dan transparan saat diujikan pada kasus nyata. Berdasarkan hal tersebut, permasalahan yang diangkat dalam

penelitian ini adalah bagaimana performa dan penerapan algoritma *Decision Tree C4.5* melalui variasi rasio pembagian data serta tingkat kedalaman pohon dalam mendeteksi penyakit hipertensi guna menentukan model optimal beserta struktur aturan keputusannya.

1.2.2 Metode Usulan

Metode usulan dalam penelitian ini adalah penerapan algoritma *Decision Tree C4.5* yang diuji secara sistematis melalui variasi rasio pembagian data dan batas kedalaman pohon (*max_depth*) guna menghasilkan model dengan akurasi tertinggi yang terhindar dari *overfitting*, sekaligus mengekstraksi struktur pohonnya untuk mengetahui hierarki atribut klinis yang paling dominan.

1.2.3 Pertanyaan Penelitian

Bagaimana performa dan pengaruh variasi rasio pembagian data serta tingkat kedalaman pohon terhadap akurasi algoritma *Decision Tree C4.5* dalam menentukan model optimal beserta struktur aturan keputusan untuk deteksi penyakit hipertensi?

1.3 Tujuan

Penelitian ini bertujuan untuk mengevaluasi kinerja dan efektivitas penerapan algoritma *Decision Tree C4.5* dengan kriteria *entropy* dalam mendeteksi penyakit hipertensi secara akurat berdasarkan variasi rasio pembagian data (70:30, 80:20, dan 90:10) serta tingkat kedalaman maksimal pohon (*max_depth*). Melalui serangkaian pengujian dan evaluasi matriks performa, penelitian ini dimaksudkan untuk mengidentifikasi konfigurasi model yang paling optimal, menganalisis pola pengaruh volume data latih terhadap kestabilan sistem, serta menyajikan representasi struktur aturan keputusan (*decision rules*) secara transparan guna mendukung keandalan proses diagnosis klinis.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat secara teoritis sebagai referensi dan kontribusi pengembangan ilmu pengetahuan di bidang *machine learning*, khususnya terkait penerapan algoritma *Decision Tree C4.5* dalam klasifikasi data medis. Secara praktis, hasil penelitian ini diharapkan mampu memberikan kontribusi nyata sebagai acuan atau sarana pendukung bagi tenaga kesehatan maupun masyarakat dalam melakukan deteksi dini penyakit hipertensi secara lebih cepat, akurat, dan transparan melalui struktur aturan keputusan yang mudah diinterpretasikan.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut :

1. Penelitian ini hanya membahas klasifikasi penyakit hipertensi menggunakan algoritma *Decision Tree C4.5*

2. Data yang digunakan berasal dari UPT Puskesmas Modopuro, Kecamatan Mojosari, Kabupaten Mojokerto, Provinsi Jawa Timur, dengan rentang waktu mulai dari bulan Mei hingga September 2022
3. Jumlah data keseluruhan adalah sebanyak 1750 data.
4. Data yang digunakan memiliki 2 kelas, Hipertensi dan bukan Hipertensi
5. Data Hipertensi berjumlah 1000 data, dan kelas bukan Hipertensi berjumlah 750 data
6. Parameter yang digunakan dalam penelitian mencakup jenis kelamin, usia (dalam tahun), indeks massa tubuh (IMT), tinggi badan (cm), berat badan (Kg), tekanan sistolik (dalam mm Hg), tekanan diastolic (dalam mm Hg), tingkat nafas (per menit), dan denyut nadi (per menit).

1.5 Sistematika Penelitian

Dalam penulisan skripsi di bagi menjadi 5 bagian sebagai berikut :

1. BAB I PENDAHULUAN

Bab ini akan menguraikan konteks sejarah penelitian, perumusan pertanyaan penelitian, pembatasan lingkup penelitian, tujuan dan kegunaan penelitian, serta metodologi penyusunan skripsi.

2. BAB II KAJIAN PUSTAKA

Pada bab ini akan dibahas kajian pustaka yang berisi penelitian-penelitian sebelumnya terkait klasifikasi penyakit menggunakan algoritma *Decision Tree C4.5*.

3. BAB III METODE USULAN

Pada bab ini akan menjelaskan secara rinci tentang alur pada penelitian ini, meliputi dataset yang akan digunakan, tahapan-tahapan dalam klasifikasi dengan algoritma *Decision Tree C4.5*.

4. BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan secara rinci mengenai penerapan dan hasil klasifikasi hipertensi yang didapatkan menggunakan algoritma *Decision Tree C4.5*

5. BAB V PENUTUP

Bab ini berisi kesimpulan berdasarkan penelitian yang telah dilakukan serta saran yang diharapkan dapat mengembangkan penelitian ini lebih lanjut.

