

BAB I

PENDAHULUAN

1.1. Latar Belakang

Diabetes merupakan penyakit yang tidak menular namun termasuk cukup serius bagi manusia. Diabetes ditandai dengan adanya gangguan metabolisme yang diakibatkan pankreas tidak memproduksi insulin dengan cukup atau bisa juga terjadi akibat tubuh tidak efektif dalam menggunakan insulin yang tersedia. Insulin berperan penting dalam mengatur penyerapan glukosa ke dalam sel-sel tubuh yang akan dijadikan sumber energi. Pada penderita diabetes, mekanisme ini terganggu sehingga glukosa yang ada di dalam darah tidak terserap ke dalam sel yang kemudian glukosa akan menumpuk di dalam darah, sehingga berdampak langsung pada rusaknya pembuluh darah, saraf dan struktur internal lainnya[1]. Menurut organisasi *International Diabetes Federation* (IDF) memperkirakan sedikitnya terdapat 463 juta orang pada usia 20-79 tahun di dunia menderita diabetes pada tahun 2019 atau setara dengan angka prevalensi sebesar 9,3% dari total penduduk usia yang sama. Angka ini diprediksi terus meningkat hingga mencapai 578 juta di tahun 2030 dan 700 juta di tahun 2045[2]. Diabetes dapat dideteksi melalui gejala klinis dan pemeriksaan laboratorium, namun dikarenakan gejala awal yang tidak spesifik sehingga seringkali diagnosis terlambat dilakukan. Oleh karena itu, diperlukan model yang efektif untuk membantu mengklasifikasikan dan mendeteksi dini penderita diabetes guna mencegah komplikasi lebih lanjut.

Pada saat ini terdapat beberapa tantangan yang menghalangi proses diagnosis diabetes. Pertama yaitu adanya keterbatasan sumber daya medis, terutama di daerah terpencil sehingga mengakibatkan kurangnya akses terhadap fasilitas diagnostic yang memadai. Kedua yaitu gejala awal penyakit diabetes sering mirip dengan gejala penyakit lain, sehingga sulit bagi tenaga medis untuk melakukan diagnosis yang tepat tanpa bantuan atau Teknik tambahan. Yang ketiga yaitu analisis yang cermat sangat diperlukan untuk menemukan pola yang relevan dalam diagnosis diabetes yang besar dan kompleks[3]. Proses ini dapat memakan waktu serta rentan terhadap terjadinya kesalahan apabila tidak dibantu dengan alat bantu yang tepat.

Salah satu masalah krusial yang perlu diselesaikan yaitu pengembangan metode klasifikasi yang akurat dan efisien untuk mengidentifikasi penderita diabetes berdasarkan data medis yang tersedia. Misalnya, jika seorang pasien mengalami gejala seperti sering merasa haus, dengan penurunan berat badan yang tidak jelas penyebabnya. Tenaga medis mungkin memerlukan serangkaian tes laboratorium yang memakan waktu untuk memastikan diagnosis diabetes jika mereka tidak memiliki alat bantu yang tepat[4]. Dengan adanya sistem klasifikasi yang handal, tenaga medis dapat dengan cepat mengidentifikasi kemungkinan diabetes pada pasien tersebut dan mengambil tindakan yang diperlukan dengan lebih awal.

Beberapa pendekatan telah dikembangkan untuk mengatasi masalah klasifikasi diabetes. Metode statistik tradisional seperti regresi logistik telah digunakan untuk memprediksi kemungkinan seseorang menderita diabetes berdasarkan variabel-variabel tertentu. Selain itu, teknik *machine learning* seperti *Neural Networks* dan *Support Vector Machines* (SVM) juga telah diterapkan dalam upaya meningkatkan akurasi prediksi. Namun, kompleksitas model dan kebutuhan akan data yang besar seringkali menjadi hambatan dalam implementasi praktis di lapangan[5].

Metode regresi logistik memiliki keunggulan dalam interpretabilitas yang mudah, namun seringkali kurang mampu menangani hubungan non-linear antara variabel. *Neural Networks* dan SVM menawarkan akurasi yang lebih tinggi, tetapi memerlukan waktu pelatihan yang lama dan sumber daya komputasi yang besar. Selain itu, model-model ini sering dianggap sebagai "*black box*" karena sulitnya interpretasi terhadap proses pengambilan keputusan yang dilakukan oleh model, sehingga menyulitkan tenaga medis dalam memahami dasar dari prediksi yang diberikan[5].

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan penerapan algoritma Decision Tree C4.5 dalam klasifikasi penderita diabetes. Algoritma C4.5 dipilih karena kemampuannya dalam menangani data dengan atribut numerik dan kategorikal, serta menghasilkan model yang mudah diinterpretasi dalam bentuk pohon keputusan. Selain itu, algoritma ini efektif dalam

mengidentifikasi variabel-variabel penting yang mempengaruhi diagnosis diabetes, sehingga dapat membantu memahami faktor-faktor kunci yang perlu diperhatikan. Dalam penelitian yang dikemukakan oleh Andy Supriyadi pada tahun 2023, penelitian ini menggunakan data dari sumber daya manusia (SDM) Universitas Sebelas Maret serta wawancara yang dilakukan oleh peneliti yang berjumlah 8 atribut. Penelitian ini mengulik mengenai perbandingan antara metode *decision tree* dan *naïve bayes*, penelitian ini mengukur tingkat klasifikasi dosen berprestasi dalam lingkup Universitas Sebelas Maret. Metode *decision tree* menunjukkan kinerja yang lebih baik dibanding dengan metode *naïve bayes*. Metode *decision tree* mencapai akurasi sebesar 95,80%, *precision* 96,90% dan *recall* 95,90%. Sedangkan metode *naïve bayes* hanya mencapai akurasi sebesar 94,80% dengan *precision* sebesar 95,25% dan *recall* sebesar 95,90%. Hal ini mengindikasikan bahwa *decision tree* lebih handal dalam mengklasifikasikan dosen berprestasi dibandingkan dengan *naïve bayes*[6].

Penerapan *decision tree* algoritma C4.5 dalam klasifikasi penderita diabetes dapat meningkatkan akurasi diagnosis. Selain itu, waktu pemrosesan yang dibutuhkan untuk menghasilkan keputusan dapat diminimalisir, sehingga memungkinkan tenaga medis untuk mengambil tindakan yang cepat dan tepat. Dengan model yang mudah diinterpretasi, tenaga medis dapat lebih memahami dasar dari setiap keputusan yang dihasilkan oleh sistem, sehingga meningkatkan kepercayaan dan penerimaan terhadap teknologi ini dalam praktik klinis sehari-hari.

1.2. Perumusan Masalah

Terdapat tiga poin penting dalam perumusan masalah, yaitu apa permasalahannya, apa Solusi yang digagas (metode usulan), dan apa pertanyaan penelitiannya. Berikut telah dijabarkan:

1.2.1 Permasalahan (*Problem Statement*)

Diabetes merupakan penyakit kronis yang mengalami peningkatan jumlah penderita setiap tahunnya. Salah satu hal yang bisa digunakan untuk mengatasi peningkatan yang semakin tinggi yaitu dengan adanya identifikasi dini terhadap penderita diabetes, hal ini sangat penting dilakukan untuk mencegah komplikasi

yang lebih serius. Namun, proses klasifikasi penderita diabetes sering kali mengalami kendala dalam hal akurasi dan efisiensi. Ketika menggunakan metode konvensional.

1.2.2 Metode Usulan

Salah satu teknik yang dapat digunakan dalam proses klasifikasi data adalah dengan menggunakan metode *Decision Tree* dengan algoritma C4.5. metode ini dapat digunakan untuk menentukan apakah seseorang menderita diabetes atau tidak berdasarkan beberapa variabel seperti gula darah, tekanan darah, BMI, usia dan faktor lainnya. Dengan menggunakan metode ini, penelitian menghasilkan model klasifikasi yang lebih akurat serta mudah dipahami.

1.2.3 Pertanyaan Penelitian

1. Seberapa akurat model klasifikasi yang dihasilkan oleh algoritma C4.5 dalam menentukan penderita diabetes berdasarkan data yang digunakan?
2. Faktor apa saja yang paling berpengaruh dalam proses klasifikasi penderita diabetes menggunakan algoritma C4.5?

1.3. Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengukur tingkat akurasi model klasifikasi yang dihasilkan oleh algoritma C4.5 dalam menentukan penderita diabetes.
2. Mengidentifikasi variabel-variabel yang berpengaruh dalam klasifikasi penderita diabetes berdasarkan hasil penelitian.

1.3.2 Manfaat:

Adapun beberapa manfaat yang diharapkan dari penelitian ini sebagai berikut:

1. Menambah wawasan mengenai penerapan metode decision tree dengan algoritma C4.5 dalam klasifikasi penyakit, khususnya diabetes.
2. Menyediakan model klasifikasi dengan algoritma C4.5 yang dapat digunakan sebagai alat bantu dalam pengambilan keputusan.
3. Meningkatkan efisiensi dalam analisis data penderita diabetes sehingga

dapat digunakan untuk perencanaan tindakan medis yang lebih tepat.

1.4 Batasan-batasan

Supaya penelitian ini lebih terfokus, terdapat beberapa Batasan masalah yang ditetapkan, yaitu:

1. Data yang digunakan dalam penelitian ini merupakan dataset yang sudah tersedia pada dataset *Kaggle* yang berjudul “Klasifikasi Data Penyakit Diabetes”.(<https://www.kaggle.com/datasets/sitirahmahbasri/data-penyakit-diabetes>)
2. Metode yang digunakan dalam membangun model klasifikasi ini adalah algoritma C4.5 tanpa membandingkannya dengan metode lain.
3. Klasifikasi yang dilakukan hanya terfokus pada dua kategori, yaitu positif diabetes dan negatif diabetes.
4. Faktor-faktor yang dianalisis didalam penelitian ini berdasarkan atribut yang tersedia dalam dataset yang digunakan, yaitu glukosa, tekanan darah, BMI, dan usia.
5. Output yang dihasilkan dari penelitian ini berupa hasil pengujian yaitu nilai akurasi, presisi, *recall*, dan *f1-score*.

Dengan Batasan ini, diharapkan penelitian bisa berjalan lebih terarah serta menghasilkan kesimpulan yang lebih akurat dalam penerapan metode *decision tree* menggunakan algoritma C4.5 untuk klasifikasi penderita penyakit diabetes.

1.5 Sistematika Skripsi

Dalam menyusun laporan ini, digunakan sistematika penulisan yang telah disesuaikan berdasarkan format penulisan skripsi pada Program Studi Sistem Informasi Fakultas Teknik Universitas Trunojoyo Madura:

BAB I: PENDAHULUAN

Bab ini berisi pemaparan mengenai latar belakang, rumusan masalah, tujuan dan manfaat, batasan masalah dan sistematika penulisan.

BAB II: TINJAUAN PUSTAKA

Bab ini akan menjelaskan mengenai penyakit diabetes, metode yang digunakan berupa metode *Decision Tree* serta algoritma C4.5. Literature yang digunakan meliputi jurnal dan buku referensi dengan tahun penerbitan kurang dari 5 tahun sebelum skripsi ini ditulis.

BAB III: METODOLOGI PENELITIAN

Bab ini menjelaskan tentang Langkah-langkah ataupun tahap yang akan dilakukan pada penelitian, pada penelitian ini akan menjelaskan mengenai Langkah-langkah membuat model klasifikasi menggunakan metode *decision tree* algoritma C4.5.

BAB IV: HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan tahapan-tahapan implementasi serta uji coba dari perancangan sistem serta analisis hasil.

BAB V: PENUTUP

Bab ini berisi kesimpulan dari pembahasan yang telah ditulis sebelumnya serta saran yang bermanfaat untuk pengembangan skripsi kedepannya.

REFERENSI

Pada referensi berisi daftar pustaka yang digunakan untuk menunjang penelitian yang dilakukan.