

BAB I

Pendahuluan

1.1. Latar Belakang

Berdasarkan data Organisasi Kesehatan Dunia (WHO), Indonesia merupakan negara yang memiliki tingkat kematian keempat tertinggi didunia yang disebabkan oleh *diabetes mellitus* [1]. Setiap tahun jumlah penderita *diabetes mellitus* selalu meningkat. Hasil riset kesehatan pada tahun 2008 menyatakan bahwa Indonesia memiliki jumlah penderita *diabetes mellitus* sebanyak 5,7% [3]. Diperkirakan pada tahun 2030 jumlah penderita *diabetes mellitus* semakin bertambah yaitu mencapai 21,257 juta jiwa.

Diabetes mellitus merupakan penyakit jangka panjang yang disebabkan gangguan kerja insulin yang menyebabkan gula dalam darah meningkat [2]. Penyakit ini termasuk dalam kategori penyakit tidak menular. Terdapat beberapa jenis Diabetes Millitus diantaranya yaitu *diabetes mellitus tipe I*, *diabetes mellitus tipe II* dan *diabetes mellitus gestasional* [2]. Dari ketiga tipe diabetes yang ada, tipe diabetes yang sering diderita oleh pasien yaitu *diabetes mellitus* tipe II. Dalam beberapa kasus, *diabetes mellitus* tipe II biasanya diderita oleh penderita yang berusia lebih dari 30 tahun. Sekitar 2.1% dari seluruh kematian disebabkan oleh *diabetes mellitus* tipe II.

Diebetes mellitus memiliki beberapa gejala yang harus diwaspadai oleh semua orang, antara lain: berapa kali hamil, tekanan darah minimum, keadaan kulit, konsentrasi insulin, index massa tubuh (kelebihan berat badan), keturunan serta usia. Dari gejala- gejala yang ada tersebut pihak kesehatan dapat mendiagnosa diabetes seorang penderita. Namun gejala atau fitur yang banyak, dapat mempengaruhi efisiensi suatu diagnosa. Dibutuhkan suatu proses agar cepat dan akurat dalam mendiagnosa penyakit *diabetes mellitus*

Klasifikasi merupakan salah satu pendekatan yang dapat digunakan untuk mengelompokkan suatu data menurut kategori tertentu [4]. Algoritma yang dapat digunakan dalam pengelompokan atau klasifikasi penyakit *diabetes mellitus* yaitu Algoritma *Decision Tree*. *Decision tree* termasuk fleksibel dan antraktif karena dapat memberikan pemecahan masalah yang diolah menggunakan teknik data mining yang membuat protokol dari prediksinya dapat diamati. *Decision tree* ini sudah

banyak digunakan dalam bidang kesehatan, salah satunya untuk diagnosa penyakit pasien, ilmu komputer pada struktur data dan lain-lain.

Salah satu penelitian yang mengkaji tentang perbandingan kinerja algoritma C4.5 dan *naïve bayes* untuk ketepatan pemilihan konsentrasi mahasiswa dilakukan oleh Wiwiwt Supriyanti dkk pada tahun 2016. Dalam penelitiannya menggunakan algoritma C4.5 dengan seleksi fitur dan *naïve bayes* dengan seleksi fitur yang sama. Hasil dari penelitian menunjukkan bahwa algoritma C4.5 dengan seleksi fitur memiliki akurasi lebih baik dari pada menggunakan algoritma *naïve bayes* [5].

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau pengelompokan yang bersifat prediktif, yaitu memprediksi peristiwa yang akan datang. Dalam kasus ini algoritma C4.5 digunakan untuk pengelompokan berdasarkan kategori positif diabetes atau negatif diabetes. Dalam penelitian ini, Seleksi fitur juga digunakan untuk mengurangi dan menghilangkan fitur atau gejala *diabetes mellitus* yang kurang relevan. Terdapat beberapa penelitian yang menggunakan seleksi fitur *F-score* dalam menentukan suatu klasifikasi. Hasil penelitian yang dilakukan oleh Elvira Sukma tahun 2016 dengan judul *Penerapan Metode Seleksi Fitur Untuk Meningkatkan Hasil Diagnosis Kanker Payudara* menunjukkan bahwa performa algoritma C4.5 dan *naïve bayes* meningkat setelah menggunakan seleksi fitur *F-score* [6]. Penelitian yang juga membahas tentang seleksi *F-Score* yaitu penelitian yang berjudul *Support Vektor Machines Combined with Feature selection for breast cancer diagnosis*. Dalam penelitiannya seleksi *f-score* memiliki kemampuan yang sangat baik dalam memilih fitur-fitur yang signifikan terhadap klasifikasi.

Oleh sebab itu penelitian ini dilakukan dengan tujuan untuk mengetahui nilai akurasi yang dihasilkan dari klasifikasi diabetes menggunakan algoritma C4.5 dengan seleksi fitur *F-Score* dan tanpa seleksi fitur pada diagnosa penyakit *diabetes mellitus*.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, dapat dibuat rumusan masalah yaitu berapakah nilai akurasi yang dihasilkan dari klasifikasi diabetes menggunakan algoritma C4.5 dengan seleksi fitur *F-Score* dan tanpa seleksi fitur?

1.3 Tujuan dan Manfaat Penelitian

Adapun tujuan dari penelitian ini yaitu mengukur nilai akurasi yang dihasilkan dari klasifikasi diabetes menggunakan algoritma C4.5 dengan seleksi fitur *F-Score* dan tanpa seleksi fitur.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

- 2.1. Sumber dataset bersal dari *Pima Indians Diabetes Database* dan diambil dari *UCI Machine Learning Repository*. Dataset terdiri dari 8 fitur dan 768 data

1.5 Metode Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan, yaitu:

a. Studi Pustaka

Pada tahap ini peneliti melakukan pengumpulan data dengan cara mencari studi literatur. Pengumpulan data dengan studi literatur berupa buku, jurnal, paper dan situs situs penunjang lainnya yang berkaitan dengan penyakit *diabetes mellitus*. Hal ini dilakukan agar mendapat data keterangan lebih lengkap dengan mempelajari literatur tersebut

b. Analisa Kebutuhan Sistem

Pada tahap ini peneliti melakukan analisa kebutuhan sistem berupa dataset tentang penyakit *diabetes mellitus* yang dapat didownload di *UCI machine learning repository*. Terdapat 8 atribut dan 768 dataset. Dataset yang di dapat, harus dibagi menjadi 2. Pertama sebagai data pelatihan (data *training*), yang kedua adalah sebagai data pengujian (data *testing*).

c. Implementasi Sistem

Mengimplementasikan metode kedalam program sesuai dengan rancangan yang telah di lakukan sebelumnya.

d. Uji Coba Sistem

Tahap ini dilakukan setelah fungsi program telah dibuat, sistem yang telah dibuat diuji menggunakan *confusion matrix*.

e. Analisa dan Evaluasi

Pada tahap ini hasil dari uji coba sistem di evaluasi bagaimana akurasi dari sistem serta mengetahui kekurangan dan kelebihan dari metode yang telah diimplementasikan pada sistem.

f. Dokumentasi

Dilakukan penyusunan laporan tugas akhir yang berlangsung mulai dari tahap studi pustaka hingga tahap evaluasi.

1.6 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian ini, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai *Diabetes mellitus* dan atributnya, seleksi fitur, klasifikasi dan *cross validation*,

BAB III PERANCANGAN SISTEM

Pada bab ini akan menjelaskan rancangan dari seluruh sistem yang akan dibuat serta alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi sistem dari rancangan yang telah dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan selanjutnya.