

BAB I PENDAHULUAN

1.1 Latar Belakang

Diabetes atau yang sering disebut dengan penyakit gula darah adalah penyakit yang mempunyai tanda awal berupa meningkatnya kadar gula darah akibat adanya gangguan sistem metabolisme di dalam tubuh. Diabetes juga salah satu penyakit kronis yang jika tidak diobati dapat menyebabkan banyak komplikasi kesehatan[1]. International Diabetes Federation (IDF) mencatat Indonesia menempati urutan ke-7 dari 10 negara dengan jumlah penderita diabetes tertinggi yang menyebabkan Indonesia berstatus waspada diabetes. Prevalensi pasien pengidap diabetes di Indonesia mencapai 6,2 persen, dengan demikian dapat diartikan ada lebih dari 10,8 juta orang menderita diabetes per tahun 2020. Selain itu Kementerian Kesehatan dengan data yang didapat dari Sample Registration Survey pada tahun 2014 menunjukkan bahwa penyebab kematian terbesar nomor 3 di Indonesia dengan persentase 6,7% adalah penyakit diabetes setelah penyakit stroke (21,1%), dan penyakit jantung koroner (12,9%) [2][3].

Pada masa pandemi ini pengidap penyakit diabetes bukanlah masalah yang dapat dianggap remeh. Penelitian yang dilakukan oleh Raden Muhammad Ali Satria pada studi kasus pasien Covid-19 di RS Bhakti Husada Surabaya menyebutkan Komorbid Diabetes dan penyakit Jantung yang diderita pasien Covid-19 dapat menjadi faktor risiko kematian [4].

Data mining adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika [5]. Teknik *data mining* juga dapat digunakan untuk memprediksi suatu penyakit menggunakan data dari rekam medis pasien. Dengan berbagai jenis data seperti jenis kelamin, tekanan darah, umur serta atribut lain menggunakan metode klasifikasi pada *data mining* dapat dimanfaatkan sebagai faktor pendukung untuk mengetahui kemungkinan pasien terkena suatu penyakit. [6].

Pohon keputusan (*decision tree*), klasifikasi *Bayessian*, jaringan saraf tiruan dan algoritma genetika adalah 3 dari sekian banyak metode yang dapat

digunakan untuk mengklasifikasikan suatu penyakit. Di antara metode klasifikasi di atas *Decision tree* telah menjadi metode yang populer untuk beberapa alasan. Jika dibandingkan dengan *neural network* atau pendekatan berbasis *bayesian*, *decision tree* lebih mudah diinterpretasikan oleh manusia. salah satu algoritma dari *decision tree* adalah C4.5 [7].

Dalam penelitian sebelumnya yang dilakukan oleh Achmad Bisri, dkk [8] menyebutkan perolehan hasil dari metode *decision tree* akurasi sebesar 87,18%, AUC sebesar 0,846, dan RSME sebesar 0,320 sedangkan hasil dari *decision tree* dengan *adaboost* perolehan akurasi sebesar 90,45%, AUC sebesar 0,951, dan RSME sebesar 0,273. Dari hasil tersebut dapat disimpulkan bahwa *decision tree* dengan *adaboost* menghasilkan akurasi yang lebih baik dibandingkan menggunakan *decision tree* biasa. Penelitian yang dilakukan oleh Mirqotussa'adah dkk menyebutkan bahwa prediksi algoritma C4.5 dengan teknik *ensemble bagging* pada data penyakit diabetes mendapatkan peningkatan akurasi sebesar 6,2% dengan akurasi awal 68,61%, setelah diterapkan *discretization* dan teknik *bagging* menjadi 74,87% [6].

Dari beberapa penjelasan diatas peneliti ingin menggunakan algoritma C4.5 dengan *ensemble Adaboost*. Data yang digunakan pada penelitian ini menggunakan pima indians diabetes datasets (PIDD). Untuk mengatasi *missing value* dan atribut kontinu pada data set peneliti menggunakan imputasi K-NN dan diskritisasi berbasis *Entropy*. Algoritma *boosting* adalah algoritma iteratif yang memberikan bobot yang berbeda pada distribusi *data training* pada setiap iterasi. Setiap iterasi *boosting* menambahkan bobot pada contoh-contoh kesalahan klasifikasi dan menurunkan bobot pada contoh-contoh klasifikasi benar [8].

1.2 Perumusan Masalah

Berdasarkan pada uraian latar belakang di atas maka dapat dipaparkan rumusan masalah sebagai berikut :

1.2.1 Permasalahan

Tingkat akurasi dari Algoritma C4.5 untuk mendiagnosis penyakit diabetes dapat dipengaruhi adanya *missing value* dan atribut kontinu selain itu untuk meningkatkan akurasi dibutuhkan algoritma peningkatan akurasi agar akurasi dari

suatu algoritma dapat ditingkatkan.

1.2.2 Metode Usulan

Solusi yang ditawarkan dalam penelitian ini adalah dengan mengatasi *missing value* dan atribut kontinu menggunakan imputasi KNN dan diskritisasi berbasis *entropy* serta penggunaan metode *ensemble adaboost* pada algoritma C4.5 untuk klasifikasi penyakit diabetes dengan akurasi tinggi.

1.2.3 Pertanyaan Peneliti

1. Bagaimana pengaruh metode *ensemble adaboost* pada algoritma C4.5 pada hasil klasifikasi penyakit diabetes?
2. Bagaimana pengaruh penggantian nilai *missing value* dengan imputasi KNN?

1.3 Tujuan Dan Manfaat

Dalam penelitian yang akan dilakukan terdapat beberapa tujuan dan manfaat yaitu :

1.3.1 Tujuan

Meningkatkan akurasi pada hasil pengklasifikasian penyakit diabetes menggunakan metode *ensemble adaboost* pada *decision tree* C4.5.

1.3.2 Manfaat

Manfaat yang diperoleh dari penelitian ini bagi peneliti adalah dapat mengetahui tingkat akurasi algoritma C4.5 dengan *ensemble adaboost* serta diharapkan dapat berguna bagi dunia kesehatan dengan mempermudah diagnosis penyakit diabetes.

1.4 Batasan

Adapun batasan dalam usulan penelitian ini adalah :

1. Dataset yang digunakan adalah *Pima Indians Diabetes Database* berjumlah 768 data dengan atribut *outcome* sebagai atribut klasifikasi.
2. Atribut klasifikasi terdiri dari 2 yakni 0 (negatif) dan 1 (positif).
3. Pra-proses dilakukan dengan mengatasi *missing value* dan menggantinya menggunakan imputasi KNN
4. Menggunakan *ensemble adaboost* dengan C4.5 sebagai algoritma pembentuk

stump.

1.5 Sistematika Laporan Skripsi

Skripsi ini ditulis dengan menggunakan sistematika penulisan seperti berikut :

- **BAB I PENDAHULUAN**

Bab ini membahas mengenai latar belakang tentang ruang lingkup permasalahan yang dibahas, solusi yang diusulkan agar dapat menyelesaikan permasalahan, manfaat dan tujuan dengan diterapkannya solusi yang diusulkan, serta sistematika penulisan proposal.

- **BAB II KAJIAN PUSTAKA**

Bab ini membahas teori mengenai metode yang digunakan, pengukuran akurasi, serta penelitian terkait yang telah ada sebelumnya.

- **BAB III METODE USULAN**

Bab ini membahas tentang keseluruhan analisa dan rancangan sistem yang dilakukan, arsitektur sistem, serta arsitektur pelatihan dan pengujian.

- **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil implementasi sistem dengan metode yang digunakan. Hasil implementasi terdiri dari lingkungan uji coba, tahapan pembuatan program, dan hasil uji coba.

- **BAB V PENUTUP**

Bab ini berisi kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian dan saran terkait pengembangan aplikasi selanjutnya.