

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit diabetes merupakan penyakit yang cukup berbahaya. Penyakit ini menyebabkan komplikasi jangka panjang yang berkembang secara bertahap. Semakin lama penderita penyakit diabetes akan semakin tidak terkontrol gula darahnya, dan semakin tinggi risiko komplikasinya. Akhirnya, komplikasi diabetes dapat melumpuhkan dan bahkan mengancam jiwa. Menurut *International Diabetes Federation* yang mengidap penyakit diabetes pada tahun 2015 sebanyak 415 juta jiwa, dan diperkirakan meningkat sebanyak 227 juta jiwa atau menjadi 642 juta jiwa pada tahun 2040 pasien yang mengidap penyakit diabetes.

Guna menyikapi masalah ini, perlu dilakukan deteksi dini penyakit diabetes. Hal ini dilakukan agar dapat melakukan pencegahan terhadap penyakit diabetes. Salah satu upaya yang dapat dilakukan untuk mencapai tujuan tersebut adalah dengan dikembangkannya suatu sistem untuk memprediksi penyakit diabetes dengan memanfaatkan berbagai metode. Salah satu metode yang dapat digunakan yaitu metode *data mining* dengan prinsip klasifikasi (Anisa & Jumanto, 2022).

Data mining adalah suatu proses untuk mencari pola dari sekumpulan data yang terdapat di dalam *database* untuk kemudian dianalisis sehingga menghasilkan suatu informasi tertentu untuk dimanfaatkan pada proses selanjutnya (Devita, Herwanto, & Wibawa, 2018). Salah satu teknik data mining

yang umum digunakan yaitu teknik klasifikasi. Klasifikasi termasuk ke dalam *supervised learning* karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu, kemudian pola dari hasil analisis tersebut digunakan untuk pengklasifikasian data uji. Pada pembelajaran data *training* dianalisis menggunakan algoritma klasifikasi, selanjutnya pada klasifikasi digunakan data *testing* untuk memastikan tingkat akurasi dari *rule* klasifikasi yang digunakan (Sartika & Sensuse, 2017). Berbagai *tools* komersial maupun non-komersial beredar dan digunakan untuk berbagai tujuan yang berkaitan dengan *data mining*. Salah satunya adalah WEKA, yang merupakan alat bantu dalam *data mining* terutama dalam penerapan untuk menyelesaikan masalah klasifikasi. Sebagian besar penelitian terbaru pada data spasial menggunakan teknik klasifikasi dikarenakan oleh sifat dari data tersebut, algoritma yang digunakan dalam penelitian yakni algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*.

Algoritma *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Naive Bayes* didasarkan pada *teorema bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Lee et al., 2001).

Sementara algoritma *Support Vector Machine (SVM)* adalah salah satu metode *machine learning* yang dikenal cukup efektif untuk kasus klasifikasi. *Dataset* dibersihkan dan dinormalisasi terlebih dahulu sehingga siap untuk dimasukkan ke dalam model SVM. Model SVM diproses dan diuji sehingga

mendapatkan model terbaik untuk melakukan diagnosis. Model SVM dibagi menjadi dua jenis yaitu model *benchmark* yang diimplementasikan menggunakan algoritma *Sequential Minimal Optimization (SMO)* dan model *scratch* yang diimplementasikan menggunakan algoritma *Sequential Learning* (Nurchaya, 2017).

Melalui *diabetes prediction dataset* yang terdapat pada dataset [kaggle.com/dataset](https://www.kaggle.com/dataset) maka akan dilakukan perbandingan dua algoritma pada WEKA yaitu algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dengan *percentage split* yang diuji coba menjadi 3 bagian yaitu 60%, 70% dan 80% yang sebelumnya belum dijabarkan pada parameter *metric* tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalahnya adalah bagaimana membandingkan nilai *accuracy*, *precision* dan *recall* yang diuji pada algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* pada *Diabetes Prediction Dataset*.

1.3 Batasan Masalah

1. Dataset diperoleh dari *website* <https://www.kaggle.com/datasets/> dengan dataset yang bernama "*diabetes prediction dataset*".
2. Pemrosesan data pada tools WEKA dibatasi dengan *file size* sebesar 16 MB.
3. Nilai *accuracy*, *precision* dan *recall* dipilih sebagai parameter pengujian untuk menguji performa dari algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*.

4. Spesifikasi laptop yang digunakan pada penelitian ini menggunakan Acer Aspire A314-32 dengan sistem operasi windows 10 64-bit.

1.4 Tujuan Penelitian

Menentukan dengan cara membandingkan nilai *accuracy*, *precision* dan *recall* pada *diabetes prediction dataset* dengan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* pada *percentage split* 60%, 70% dan 80% menggunakan tools WEKA versi 3.9.6.

1.5 Manfaat Penelitian

Adapun manfaat dalam penelitian ini bagi perkembangan IPTEK diharapkan dapat bermanfaat bagi ilmu perkembangan di bidang teknologi informasi khususnya mengenai pengolahan dataset diabetes menggunakan *data mining*.

1.6 Metodologi Penelitian

1. Pengumpulan Data

Menjelaskan asal mula data yang diperoleh dari internet untuk selanjutnya dilakukan penelitian.

2. Analisis Permasalahan dan Pencarian Solusi

Proses kegiatan membandingkan data dari algoritma yang dipilih untuk digolongkan dan dikelompokkan sesuai kriteria yang diharapkan.

3. Implementasi Solusi

Proses pemecahan suatu masalah yang diteliti pada pengujian ini dan menghasilkan sebuah jawaban yang diharapkan.

4. Penarikan Kesimpulan

Pada tahap ini hasil dari penelitian yang telah dilakukan secara keseluruhan ditarik menjadi sebuah kesimpulan.

1.7 Sistematika Penulisan

Penulisan dalam laporan tugas akhir ini memakai sistematika pembahasan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini akan membahas tentang teori-teori dan konsep-konsep yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan masalahnya. Selain itu, bab ini juga memuat teori-teori dalam pelaksanaan pengumpulan dan pengolahan data serta melakukan penganalisaan.

BAB III METODOLOGI

Pada bab ini menjelaskan tentang metode yang dipilih dan digunakan yang terdiri dari beberapa tahapan yaitu pengumpulan data, analisis permasalahan dan pencarian solusi, implementasi solusi dan penarikan kesimpulan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan membahas mengenai analisa yang dilakukan terhadap hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Bab terakhir akan memuat kesimpulan dan saran keseluruhan dari bab sebelumnya sebagai hasil yang diperoleh dan diharapkan dapat bermanfaat dalam penelitian selanjutnya.