

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mencari algoritma yang cocok untuk mencapai tujuan yang diinginkan tidaklah mudah. Pemilihan algoritma tergantung pada jenis masalah yang ingin dipecahkan, jenis data yang digunakan, serta tujuan analisis yang ingin dicapai. Algoritma populer seringkali digunakan tanpa mempertimbangkan alternatif lainnya, dengan anggapan bahwa algoritma tersebut sudah sangat optimal. Namun, penting untuk melakukan perbandingan antara algoritma-algoritma yang ada untuk melihat akurasi hasil yang diberikan oleh masing-masing algoritma pada data yang sama. Hal ini dilakukan untuk menentukan apakah terdapat algoritma lain yang lebih baik daripada yang sering digunakan dalam menyelesaikan masalah. Dalam bidang ilmu komputer, perbandingan algoritma dilakukan untuk membandingkan kelebihan dan kekurangan masing-masing algoritma dalam menyelesaikan masalah yang serupa. Perbandingan algoritma dapat dilakukan dalam berbagai macam permasalahan salah satunya untuk klasifikasi.

Algoritma *Hopfield*, *Random Forest*, dan *Support Vector Machine* (SVM) telah menjadi perhatian utama penulis. Ketiga algoritma ini cukup unik dibandingkan, karena algoritma *Hopfield* merupakan Jaringan Syaraf Tiruan (JST) yang masih jarang dilakukan untuk klasifikasi data numerik, sedangkan algoritma *Random Forest* dan SVM merupakan algoritma *machine learning* yang umum digunakan, karena hasilnya optimal dalam melakukan klasifikasi. Penelitian terkait perbandingan maupun implementasi masing-masing algoritma *Random Forest* dan SVM terhadap klasifikasi penyakit diabetes beberapa sudah dilakukan terkecuali algoritma *Hopfield*. Berdasarkan penelitian yang dilakukan oleh Kumiati & Hilda, (2020) ditemukan bahwa algoritma *Support Vector Machine* (SVM) memiliki akurasi yang lebih tinggi yaitu 96.2704% dibandingkan dengan algoritma *Naïve Bayes* sebesar 92.0746% [1]. Selain itu penelitian yang dilakukan oleh Witjaksana, Saedudin dkk., (2021)

menunjukkan jika algoritma *Random Forest* memiliki akurasi yang lebih tinggi yaitu 90.62% dibandingkan dengan algoritma *Artificial Neural Network* (ANN) sebesar 82.29% [2]. Adapun penelitian yang melakukan pendekatan algoritma *Hopfield* mengenai klasifikasi dilakukan oleh Budianita, Hustianto dkk., (2018) didapatkan hasil akurasi yang cukup baik sebesar 88.51% [3]. Berdasarkan penelitian terdahulu, dapat diketahui bahwa ketiga algoritma ini memiliki akurasi yang cukup besar untuk melakukan kasus klasifikasi.

Sampai saat ini, belum ditemukan ada penelitian yang secara komprehensif membandingkan ketiga algoritma ini dalam konteks deteksi dini penyakit diabetes. Oleh karena itu, penelitian ini bertujuan untuk melakukan perbandingan kinerja algoritma *Hopfield*, *Random Forest*, dan SVM untuk deteksi dini penyakit diabetes. Untuk mendukung penelitian ini penulis juga akan menggunakan metode klasifikasi data mining yang dikenal dengan sebutan *Knowledge Discovery in Database* (KDD). Proses KDD ini bertujuan untuk mengumpulkan dan memanfaatkan data historis dengan tujuan menemukan pola-pola keteraturan yang terdapat dalam kumpulan data berukuran besar [4]. Diharapkan penelitian ini dapat memberikan wawasan yang berharga dalam pemilihan algoritma yang paling efektif untuk klasifikasi penyakit diabetes. Hasil penelitian ini diharapkan dapat memberikan kontribusi positif dalam pengembangan sistem klasifikasi penyakit diabetes yang lebih efisien dan dapat digunakan secara luas.

1.2 Rumusan Masalah Penelitian

Berdasarkan hasil pemaparan latar belakang diatas, didapat rumusan masalah sebagai berikut: Bagaimana tingkat akurasi dari algoritma *Hopfield*, *Random Forest*, dan *Support Vector Machine* (SVM) dalam melakukan deteksi dini penyakit diabetes?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan didapat tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui performa dari algoritma *Hopfield*, *Random Forest*, dan *Support Vector Machine* (SVM) dalam deteksi dini penyakit diabetes.

1.4 Batasan Masalah Penelitian

Agar penelitian ini menjadi lebih terarah, maka perlu diadakannya batasan masalah. Pada penelitian ini batasan masalah yang dibahas terdiri dari:

1. Algoritma yang diuji hanya algoritma *Hopfield*, *Random Forest* dan *Support Vector Machine* (SVM).
2. Studi data yang menjadi fokus pada penelitian ini adalah data 'Pima Indians Diabetes Database' yang bersumber dari *National Institute of Diabetes and Digestive and Kidney Diseases*.

1.5 Kerangka Pemikiran Penelitian

Gambar 1.1 mewakili kerangka pemikiran yang akan digunakan dalam penelitian ini. Permasalahan yang diangkat adalah kebingungan yang sering muncul akibat banyaknya pilihan algoritma yang tersedia untuk tujuan klasifikasi. Untuk mengatasi hal ini, pendekatan yang diusulkan melibatkan perbandingan antara algoritma-algoritma tersebut. Penelitian ini selanjutnya menerapkan tiga algoritma berbeda untuk klasifikasi, dengan menggunakan metode *Knowledge Discovery in Database* (KDD) dalam pengembangan sistem. Tahapan metode ini akan diimplementasikan menggunakan bahasa pemrograman Python, dengan tujuan akhirnya untuk mencapai akurasi yang paling optimal di antara ketiga algoritma yang digunakan.



Gambar 1.1 Kerangka Pemikiran

1.6 Sistematika Penulisan

Pada penelitian ini, penulis membagi sistematika penulisan menjadi 5 bab sebagai berikut:

BAB I: PENDAHULUAN

Dalam BAB I, penulis memaparkan mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, kerangka berpikir, dan memaparkan bagaimana sistematika penulisannya.

BAB II: KAJIAN LITERATUR

Dalam BAB II, penulis memaparkan mengenai landasan teori serta penelitian terdahulu yang menjadi acuan dalam penelitian ini.

BAB III: METODE PENELITIAN

Dalam BAB III, penulis memaparkan metode yang akan digunakan dalam penelitian ini, meliputi metode penelitian analisi sumber data dan metode pengembangan sistem.

BAB IV: HASIL DAN PEMBAHASAN

Dalam BAB IV, penulis memaparkan hasil penelitian yang sudah dijalankan serta menjelaskan secara rinci mengenai evaluasi model yang sudah didapatkan.

BAB V: KESIMPULAN DAN SARAN

Dalam BAB V, penulis memaparkan mengenai kesimpulan yang sudah didapatkan serta memberi saran untuk penelitian selanjutnya.