

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit diabetes adalah salah satu masalah kesehatan masyarakat yang signifikan dan mempengaruhi jutaan orang di seluruh dunia [1],[2]. Diabetes adalah gangguan metabolik kronis yang ditandai dengan kadar glukosa darah yang tinggi (hiperglikemia) karena masalah pada produksi insulin atau resistensi insulin. Identifikasi dan diagnosis dini diabetes sangat penting untuk mencegah komplikasi serius dan memberikan perawatan yang tepat bagi penderita. Dalam beberapa tahun terakhir, perkembangan teknologi informasi telah membawa perubahan dalam dunia kedokteran dan kesehatan. Penggunaan algoritma machine learning dan khususnya algoritma *machine learning* telah menunjukkan potensi besar dalam membantu diagnosis dan pemantauan penyakit, termasuk diabetes.

Penelitian sebelumnya telah mengusulkan pendekatan machine learning berbeda untuk mendeteksi diabetes. Penelitian [3] mengimplementasikan metode Support Vector Machine dan mencapai akurasi sebesar 80,73%. Penelitian [4] menggunakan Decision Tree C4.5 dan mencapai akurasi sebesar 70,32%, sedangkan penelitian [5] menggunakan metode K-Nearest Neighbor dan mencapai akurasi sebesar 65%. Meskipun penelitian sebelumnya telah menunjukkan hasil yang menjanjikan, namun masih ada potensi untuk meningkatkan akurasi dan efektivitas diagnosa diabetes dengan menggunakan algoritma ensemble learning. Algoritma ensemble learning, seperti Random Forest, AdaBoost, dan XGBoost, telah terbukti efektif dalam mengatasi permasalahan klasifikasi yang kompleks dengan menggabungkan beberapa model pembelajaran.

Oleh karena itu, untuk menyelesaikan permasalahan yang ada, penelitian ini akan melakukan analisis perbandingan dari tiga algoritma ensemble learning (Random Forest, AdaBoost, dan XGBoost) dalam mengklasifikasikan diagnosa penyakit diabetes. Penelitian ini akan mengevaluasi kinerja ketiga algoritma berdasarkan akurasi, presisi, dan area under the curve (AUC) dari kurva ROC. Dengan membandingkan hasil penelitian sebelumnya dengan metode ensemble learning yang diajukan, diharapkan penelitian ini dapat memberikan kontribusi dalam meningkatkan akurasi diagnosa penyakit diabetes.

Berdasarkan penjelasan yang telah disebutkan diatas, untuk memberikan kontribusi dalam dunia Kesehatan, maka peneliti memutuskan untuk mengangkat judul penelitian **“Analisis Perbandingan Algoritma Ensemble Learning dalam Mengklasifikasi Diagnosa Awal Penyakit Diabetes”**.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah dijelaskan, maka diangkat rumusan masalah penelitian yaitu “Bagaimana kinerja tiga algoritma ensemble learning (*Random Forest*, *AdaBoost*, dan *XGBoost*) dalam mengklasifikasikan diagnosa penyakit diabetes?”.

1.3. Batasan Masalah

Batasan masalah yang telah ditetapkan dalam penelitian ini adalah:

1. Dataset diambil dari repositori dataset penyakit diabetes milik UCI Machine Learning.
2. Melakukan klasifikasi biner (YA atau TIDAK) terhadap diagnose penyakit diabetes.
3. Penelitian berfokus pada penentuan gejala yang paling berpengaruh dan mencari model yang memiliki nilai akurasi tertinggi.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membuat sebuah model algoritma klasifikasi berbasis *Ensemble Learning* untuk mengklasifikasi diagnose penyakit diabetes

1.5. Manfaat Penelitian

Manfaat yang terdapat dalam penelitian ini adalah:

1. Sebagai sarana bagi instansi kesehatan dalam meningkatkan akurasi diagnosa diabetes dan mengembangkan sistem pendukung keputusan (SPK)
2. Mendapatkan gejala yang paling berpengaruh dalam diagnosa penyakit diabetes.
3. Menambah *insight* mahasiswa tentang ranah kesehatan.