

BAB I PENDAHULUAN

1.1 Latar Belakang

Diabetes merupakan salah satu masalah kesehatan global yang terus meningkat secara signifikan dari tahun ke tahun. Berdasarkan laporan International Diabetes Federation (IDF), jumlah penderita diabetes di seluruh dunia pada tahun 2024 diperkirakan mencapai 11,3 % dari populasi dewasa, atau sekitar 800 juta orang. Peningkatan ini dipengaruhi oleh berbagai faktor seperti pola hidup yang kurang sehat, tingginya persentase obesitas, dan faktor risiko genetik. Kondisi tersebut tidak hanya berdampak pada kualitas hidup individu, tetapi juga memberikan beban besar bagi sistem kesehatan global, baik dari segi biaya perawatan maupun tingginya angka penderita penyakit diabetes dan korban dunia yang meninggal akibat diabetes[1].

Di tingkat nasional, Indonesia termasuk salah satu negara dengan jumlah penderita diabetes tertinggi di dunia. Berdasarkan data IDF, Indonesia menempati peringkat ketujuh dengan jumlah kasus mencapai 20,4 juta orang dewasa pada tahun 2024 [2]. Angka ini mengindikasikan adanya ancaman serius terhadap produktivitas nasional dan peningkatan beban pembiayaan kesehatan. Tidak hanya itu, komplikasi yang ditimbulkan seperti penyakit jantung, gagal ginjal, retinopati, dan neuropati semakin memperparah kondisi kesehatan masyarakat [3]. Oleh karena itu, upaya diagnosis dini dan akurat menjadi salah satu langkah krusial dalam mengendalikan perkembangan penyakit ini.

Metode diagnosis standar medis manual yang selama ini digunakan, seperti pengukuran glukosa darah puasa, tes HbA1c, dan tes toleransi glukosa oral, memang menjadi standar medis global. Namun, metode ini memiliki keterbatasan signifikan, di antaranya rendahnya sensitivitas pada tahap awal penyakit, tingginya risiko hasil negatif palsu maupun positif palsu, serta ketidakmampuan mengolah data medis yang besar dan kompleks secara efisien [4]. Pendekatan statistik tradisional, seperti regresi logistik, cenderung hanya

mampu menangkap hubungan linear sehingga sulit memodelkan interaksi non-linear yang umum terjadi pada data kesehatan [5]. Selain itu, perbedaan karakteristik klinis antar populasi juga membuat metode berbasis aturan tetap (rule-based) sulit beradaptasi [6]. Keterbatasan inilah yang mendorong pemanfaatan teknologi pembelajaran mesin, khususnya Artificial Neural Network (ANN) dan Random Forest (RF), yang memiliki kemampuan mengenali pola kompleks, beradaptasi dengan data beragam, dan meningkatkan akurasi prediksi dibandingkan metode konvensional [7].

Berbagai penelitian telah membuktikan efektivitas kedua algoritma tersebut. Misalnya, penelitian menggunakan dataset Pima Indian Diabetes menunjukkan bahwa ANN mampu mencapai akurasi sebesar 78,57 %, sedangkan RF memperoleh akurasi 76,30 % [8]. Ini mengindikasikan bahwa ANN dan RF keduanya berpotensi menjadi solusi yang lebih andal dibanding metode konvensional.

Namun demikian, masih terdapat beberapa celah penelitian yang perlu diisi. Pertama, sebagian besar studi terdahulu belum secara spesifik membandingkan ANN dan RF pada dataset diabetes yang telah melalui proses balancing menggunakan SMOTE dan normalisasi fitur, padahal kedua teknik preprocessing ini penting untuk mengurangi bias akibat ketidakseimbangan data [10]. Kedua, mayoritas penelitian masih berfokus pada akurasi saja, tanpa memperhatikan metrik evaluasi lainnya seperti presisi, recall, F1-score, dan ROC-AUC yang sangat penting dalam konteks diagnosis medis [11]. Ketiga, minimnya studi yang melakukan perbandingan kinerja ANN dan RF pada dataset dengan karakteristik berbeda dari Pima Indian Diabetes Dataset membuat analisis kestabilan hasil prediksi pada data yang telah diproses optimal menjadi jarang dilakukan. Oleh karena itu, penelitian ini hadir untuk mengisi gap tersebut dengan melakukan perbandingan kinerja ANN dan RF menggunakan dataset diabetes yang telah di-resample dan dinormalisasi, serta dievaluasi secara komprehensif

menggunakan berbagai metrik performa.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah didapatkan rumusan masalah sebagai berikut:

1. Bagaimana perbandingan kinerja antara Artificial Neural Network dan Random Forest dalam mengklasifikasi penyakit diabetes?
2. Bagaimana evaluasi kinerja model dapat membantu dalam memahami proses klasifikasi penyakit?

1.3 Batasan Masalah

Dalam penelitian ini menggunakan batasan masalah agar penelitian hanya fokus pada pokok permasalahan oleh karena itu memiliki batasan antara lain:

1. Fokus penelitian adalah pada analisis kinerja model ANN dan RF tanpa membandingkan dengan algoritma lain.
2. Evaluasi kinerja akan dibatasi pada metrik akurasi, presisi, recall, dan F1- score.
3. Penelitian hanya melakukan perbandingan antar dua algoritma serta membuat modelnya tanpa melakukan implementasi secara langsung pada sebuah sistem.
4. Penelitian dilakukan dengan menggunakan python, tanpa melibatkan uji klinis atau data real- time dari pasien rumah sakit.
5. Visualisasi hasil dibatasi pada confusion matrix, Kurva ROC, heatmap korelasi fitur, dan feature importance.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk membandingkan kinerja kombinasi ANN dan RF dalam mengklasifikasi penyakit diabetes.
2. Untuk mengetahui bagaimana evaluasi kinerja model dapat membantu

dalam memahami proses klasifikasi penyakit.

1.5 Manfaat Penelitian

Ada pula beberapa manfaat yang diharapkan oleh penulis dari penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi pada dunia medis terutama dalam pengembangan ilmu di bidang data mining dan kecerdasan buatan, khususnya terkait efektivitas algoritma ANN dan RF dalam klasifikasi penyakit diabetes.
2. Hasil penelitian ini dapat digunakan sebagai acuan dalam pemilihan algoritma klasifikasi yang optimal untuk diagnosis diabetes serta dapat menjadi referensi bagi peneliti selanjutnya untuk melakukan penelitian serupa terutama di bidang Kesehatan.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini sebagai berikut:

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang digunakan.

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang objek penelitian, analisis masalah, solusi yang ditawarkan, rancangan.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam mengembangkan aplikasi, testing hingga penerapan aplikasi di objek penelitian.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian