

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Diabetes melitus merupakan gangguan metabolisme yang dikenali melalui lonjakan kadar gula dalam darah (*hiperglikemia*) yang terjadi akibat ketidakseimbangan produksi atau fungsi insulin. Keadaan ini tidak hanya mengganggu proses pemecahan karbohidrat, namun juga berpengaruh terhadap pengolahan lemak dan protein dalam tubuh. Penderita diabetes melitus umumnya mengalami beberapa gejala yang biasa ditemukan, antara lain rasa haus yang berlebihan (*polidipsia*), frekuensi buang air kecil yang meningkat (*poliuria*), nafsu makan yang meningkat secara drastis (*polifagia*), penurunan berat badan tanpa sebab yang jelas, serta sensasi kesemutan di beberapa bagian tubuh.

Diabetes melitus sering dikenal sebagai '*silent killer*' karena penderitanya tidak merasakan gejala dan baru diketahui ketika telah terjadi komplikasi yang serius [1]. Berdasarkan data yang berasal dari World Health Organization (WHO), diabetes termasuk salah satu penyakit dengan jumlah kasus terbanyak di dunia, dengan sekitar 422 juta kasus pada tahun 2019. Sebagian besar di antaranya merupakan diabetes tipe 2. Dalam beberapa dekade terakhir, prevalensi diabetes mengalami peningkatan yang meningkat pesat, khususnya di negara-negara dengan pendapatan rendah dan menengah.

Diabetes menjadi pemicu utama berbagai komplikasi kesehatan serius, termasuk kebutaan, gagal ginjal, serangan jantung, stroke, dan amputasi pada bagian tubuh bawah. Pada tahun 2019, penyakit ini serta gangguan ginjal yang diakibatkannya menyebabkan sekitar 2 juta kematian. Sementara itu, laporan International Diabetes Federation (IDF) tahun 2021 mencatat bahwa 537 juta orang dewasa berusia 20-79 tahun di dunia hidup dengan diabetes, dan jumlah ini diperkirakan akan meningkat menjadi 783 juta pada tahun 2045.

Pada tahun 2021 sekitar 10,6% atau 19,47 juta dari populasi dewasa (usia 20-79 tahun) di Indonesia hidup dengan diabetes, sehingga Indonesia menempati peringkat kelima di seluruh dunia, dan jumlah ini diprediksi akan terus bertambah seiring waktu. Dengan semakin meningkatnya jumlah penderita diabetes secara global dan di Indonesia, penting untuk memiliki model prediksi yang akurat untuk membantu diagnosis dini dan penanganan yang lebih baik.

Salah satu penyebab meningkatnya prevalensi penyakit diabetes adalah pola hidup yang tidak sehat, seperti mengonsumsi makanan yang berlemak dan manis secara berlebihan baik secara sadar maupun tidak sadar [2]. Seiring dengan meningkatnya jumlah penderita diabetes melitus, maka kebutuhan akan metode diagnostik yang cepat dan akurat menjadi semakin penting [3].

Di sinilah peran teknologi, khususnya kecerdasan buatan dan *machine learning* dapat membantu menangani berbagai penyakit dan mengembangkan model prediktif yang akurat [4]. Sebagai bagian dari kecerdasan buatan, *machine learning* memungkinkan komputer untuk belajar dari data tanpa perlu instruksi eksplisit. Dalam konteks kesehatan, algoritma *machine learning* dapat digunakan untuk menganalisis data pasien, seperti kadar gula darah, berat badan, dan faktor risiko lainnya, yang kemudian dapat digunakan untuk memprediksi apakah seseorang berisiko terkena diabetes melitus.

Berbagai penelitian telah memanfaatkan *machine learning* untuk memprediksi penyakit diabetes dengan tingkat akurasi yang sangat tinggi. Salah satu penelitian yang membandingkan algoritma *support vector machine (SVM)* dan *modified balanced random forest (MBRF)* dalam mengklasifikasikan data pasien diabetes dari *dataset* Gula Karya Medika menunjukkan bahwa *MBRF* memiliki tingkat akurasi sebesar 97,8%, sementara *SVM* sebesar 91,48% [5]. Penelitian lain menunjukkan bahwa algoritma *Naïve Bayes* menghasilkan akurasi sebesar 74,87%, *logistic regression* 75,78%, dan *neural network* 69,27% [6].

Masing-masing algoritma memiliki kelebihan dan kekurangannya tersendiri, seperti algoritma *neural network* memiliki kelebihan mampu menganalisis data dan memberikan hasil dengan cepat dan akurat, sehingga risiko kesalahan dan *error* menurun, tetapi membutuhkan waktu dan sumber daya komputasi yang lebih lama dibandingkan dengan *random forest*. Sedangkan algoritma *random forest* memiliki ketahanan yang baik terhadap *outlier* dan *noise*, tetapi kurang efektif terhadap data yang berdimensi tinggi dibandingkan algoritma *support vector machine*. Kemudian algoritma *support vector machine* memastikan bahwa model tidak terlalu menyesuaikan data latihan, sehingga dapat melakukan generalisasi pada data yang baru, tetapi algoritma ini lebih lambat pada *dataset* yang besar dibandingkan dengan algoritma *random forest*.

Proses penelitian yang akan dilakukan dalam penelitian ini adalah mengumpulkan data, mengeksplor data untuk memahami data lebih komprehensif, membagi data dalam bentuk *training* dan *testing*, kemudian data dilatih dan dievaluasi berdasarkan latihan yang sudah diberikan sebelumnya sehingga bisa melihat manakah algoritma yang memiliki tingkat akurasi yang paling baik. Selain itu, *dataset* yang digunakan pada penelitian ini memiliki atribut *gender*, *age*, *hypertension*, *heart disease*, *smoking history*, *BMI*, *level HbA1c*, *blood glucose* dan *diabetes* yang juga akan digunakan sebagai label.

Peneliti bermaksud untuk membandingkan tingkat akurasi dalam memprediksi penyakit diabetes menggunakan algoritma *neural network*, *support vector machine* dan *random forest* dengan menggunakan *dataset* yang didapat dari situs *Kaggle*. Belum adanya penelitian yang melakukan perbandingan langsung antar algoritma dengan menggunakan ketiga algoritma di atas, sehingga penulis memutuskan untuk melakukan penelitian dengan judul **“PERBANDINGAN TINGKAT AKURASI ALGORITMA NEURAL NETWORK, SUPPORT VECTOR MACHINE, DAN RANDOM FOREST DALAM MEMPREDIKSI PENYAKIT DIABETES MELITUS MENGGUNAKAN MACHINE LEARNING”**.

1.2 Ruang Lingkup Masalah

- a. Menganalisis dan membandingkan tingkat akurasi dari algoritma *neural network*, *support vector machine*, dan *random forest* dalam memprediksi penyakit diabetes melitus.
- b. Penyakit yang diprediksi pada penelitian ini hanya diabetes melitus.
- c. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* yang disediakan oleh Kaggle berjumlah 100.000 yang akan dibagi menjadi training dan testing.
- d. Evaluasi algoritma akan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk membandingkan performa masing-masing algoritma.

1.3 Rumusan Masalah

- a. Bagaimana perhitungan nilai metrik pada algoritma neural network, support vector machine, dan random forest dapat membantu untuk memilih algoritma terbaik dalam memprediksi penyakit diabetes melitus?
- b. Algoritma manakah yang memberikan performa terbaik berdasarkan tingkat akurasi, precision, recall dan f1-score?

1.4 Tujuan Penelitian

- a. Menemukan algoritma machine learning dengan nilai akurasi dan performa terbaik dalam memprediksi diabetes melitus.
- b. Membuat model prediksi berdasarkan *dataset* diabetes menggunakan algoritma neural network, support vector machine dan random forest.
- c. Mengevaluasi performa setiap model menggunakan metrik akurasi, precision, recall, dan f1-score.

1.5 Manfaat Penelitian

- a. Menjadi referensi bagi peneliti lain untuk memilih algoritma machine learning dengan akurasi dan performa terbaik dalam memprediksi diabetes melitus.
- b. Memberikan wawasan mengenai algoritma yang memiliki performa evaluasi terbaik berdasarkan metrik seperti precision, recall, dan F1-score, sehingga bisa digunakan sebagai dasar untuk pengembangan model prediksi lebih lanjut.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini yaitu:

Bab I Pendahuluan. Memuat tentang latar belakang masalah, ruang lingkup masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka. Bab ini membahas teori-teori yang mendasar dan memberikan pemahaman terkait materi penelitian yang bersumber dari berbagai referensi ilmiah, seperti buku dan jurnal.

Bab III Metodologi Penelitian. Bab ini membahas tentang metode-metode pendekatan yang digunakan untuk menyelesaikan permasalahan yang telah dirumuskan dalam penelitian ini.

Bab IV Hasil dan Pembahasan. Bab ini menguraikan hasil analisis yang diperoleh serta pembahasan terkait dengan teori-teori yang relevan.

Bab V Penutup. Bab ini memuat kesimpulan dari keseluruhan penelitian, menjawab pertanyaan dari rumusan masalah, serta memberikan saran-saran untuk perbaikan dan pengembangan lebih lanjut.