

BAB I

PENDAHULUAN

1.1. Latar Belakang

Diabetes mellitus, yang lebih dikenal sebagai penyakit kencing manis, adalah gangguan metabolisme kronis yang memengaruhi kerja pankreas, khususnya dalam produksi insulin. Gangguan ini ditandai dengan kondisi hiperglikemia atau tingginya kadar gula dalam darah yang dapat berlangsung seumur hidup. Penyakit ini memiliki dua jenis utama, yaitu diabetes tipe 1, yang disebabkan oleh reaksi autoimun terhadap sel beta pankreas, dan diabetes tipe 2, yang diakibatkan oleh kombinasi faktor genetik dan lingkungan, termasuk obesitas, pola makan yang tidak sehat, stres, dan penuaan.

Komplikasi akibat diabetes dapat dibagi menjadi dua kategori, yakni makrovaskular seperti penyakit kardiovaskular, serta mikrovaskular yang meliputi kerusakan pada ginjal (nefropati), sistem saraf (neuropati), dan mata (retinopati). Diabetes juga merupakan masalah kesehatan global yang terus meningkat. Data WHO mencatat bahwa pada tahun 2000 terdapat 171 juta kasus diabetes di seluruh dunia, dan angka ini diproyeksikan meningkat menjadi 366 juta pada tahun 2030. Sementara itu, Indonesia menduduki peringkat keempat dengan jumlah penderita diabetes terbanyak setelah Amerika Serikat, China, dan India, serta diperkirakan mengalami lonjakan kasus hingga dua hingga tiga kali lipat pada tahun 2030 dibandingkan tahun 2000 [1].

Di Indonesia sendiri, penyakit diabetes mellitus sedang mengalami kenaikan yang tak hanya pada orang dewasa, namun juga anak-anak. Menurut Data Ikatan Dokter Anak Indonesia (IDAI), telah terjadi lonjakan kasus diabetes pada anak usia 0-18 tahun sebesar 70 kali lipat dalam 13 tahun terakhir. Faktor penyebab utama meliputi pola makan tinggi gula dan kurangnya aktivitas fisik, yang berkontribusi pada obesitas. Nur Handayani, S.KM menjelaskan bahwa hal ini dapat terjadi karena banyaknya tempat-tempat yang menjual makanan ringan dan murah dengan kandungan gula yang cukup tinggi. Ini membuat anak-anak memiliki kebiasaan buruk dengan mengonsumsi makanan tersebut secara berlebihan [2]. Oleh karena itu, dibutuhkan sebuah pencegahan dengan mengenali gejala-gejala awal dari penyakit ini sehingga dapat melakukan diagnosis dini pada anak.

Teknologi kecerdasan buatan dan pembelajaran mesin (machine learning) telah memberikan kontribusi besar dalam bidang medis, termasuk untuk deteksi dan prediksi penyakit. Salah satu metode yang sering digunakan adalah algoritma *K-Nearest Neighbors* (KNN), sebuah algoritma berbasis pendekatan jarak untuk klasifikasi dan regresi. Algoritma ini memiliki keunggulan dalam kesederhanaan implementasi serta kemampuannya untuk memberikan hasil yang akurat dalam banyak kasus. Namun, performa KNN sangat bergantung pada pemilihan parameter, seperti jumlah tetangga terdekat (k) dan metrik jarak yang digunakan, yang memengaruhi akurasi prediksi [3].

Beberapa metode telah digunakan untuk melakukan prediksi penyakit diabetes pada penelitian-penelitian sebelumnya. Metode-metode tersebut yakni, algoritma *Multinomial Naïve Bayes* yang digunakan untuk prediksi penyakit diabetes mellitus pada wanita *reproductive* yang mendapat akurasi sebesar 93% [4], penggunaan algoritma C4.5 yang digunakan untuk deteksi dini penyakit diabetes mendapat akurasi 95,96% dan berhasil ditingkatkan ke 96,54% dengan optimasi algoritma genetika [5], algoritma *K-Nearest Neighbors* atau KNN yang memperoleh akurasi sebesar 76,56% pada scenario terbaiknya, metode *Decision Tree* dengan optimalisasi *Feature Selection* mendapat akurasi deteksi penyakit diabetes sebesar 96,25% [6].

Dari penelitian-penelitian dengan berbagai pendekatan yang telah dilakukan, hasil yang didapat memiliki akurasi yang terbilang bagus terutama pada pendekatan KNN. Namun jika dibandingkan dengan pendekatan lainnya, Algoritma KNN memiliki akurasi yang paling rendah. Hal ini dikarenakan Algoritma KNN memiliki kelemahan dalam penentuan parameter k yang mana hasil akurasi yang dihasilkan akan berbeda jika menggunakan parameter yang berbeda. Tentunya kelemahan ini dapat diatasi dengan penerapan penerapan algoritma optimasi, salah satu adalah Algoritma Genetika.

Pada penelitian ini, Algoritma *K-Nearest Neighbors* akan dioptimasi dengan Algoritma Genetik pada bagian seleksi fitur. Algoritma genetika (GA) digunakan dalam optimasi seleksi fitur untuk meningkatkan kinerja model dengan memilih subset fitur yang relevan dan mengabaikan fitur yang tidak signifikan. Selain itu, pendekatan ini juga dapat menentukan parameter k yang optimal dalam melakukan proses pelatihan pada model. Prosesnya mencakup representasi kromosom untuk setiap subset fitur dan parameter k , evaluasi menggunakan fungsi fitness, serta penerapan

operator genetika seperti seleksi, crossover, dan mutasi untuk menghasilkan populasi baru. Melalui iterasi, GA mengeksplorasi ruang solusi secara efisien untuk menemukan kombinasi fitur optimal dan parameter k yang memberikan performa terbaik pada dataset. Pendekatan ini bermanfaat dalam domain medis, di mana jumlah fitur besar dapat memengaruhi efisiensi dan akurasi model prediksi [7].

Optimasi dengan pendekatan Algoritma Genetika juga terbukti efektif dalam meningkatkan akurasi di penelitian-penelitian sebelumnya. Salah satunya adalah penelitian prediksi penyakit hepatitis menggunakan metode KNN dan dioptimasi dengan Algoritma Genetika, dimana penelitian tersebut membandingkan hasil yang didapat ketika tanpa melakukan optimasi dan hasil yang telah diterapkan optimasi Algoritma Genetika. Hasil akurasi yang didapat pada penelitian ini sebesar 73.81% jika tanpa menggunakan optimasi Algoritma Genetika dan sebesar 95.24% setelah dilakukan optimasi [8]. Peningkatan akurasi sebesar 21.43% cukup membuktikan bahwa optimasi yang dilakukan Algoritma Genetika mampu meningkatkan akurasi deteksi penyakit.

Dari latar belakang yang telah dipaparkan, tujuan penelitian yang berjudul “OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS” adalah untuk mengetahui performa Algoritma *K-Nearest Neighbor* setelah dilakukan optimasi dengan Algoritma Genetika dalam melakukan deteksi penyakit diabetes mellitus. Hasil luaran dari penelitian ini adalah tingkat akurasi yang didapat dari penerapan metode KNN dengan optimasi Algoritma Genetika lalu dievaluasi dengan cara membandingkan tingkat akurasi dari beberapa skenario.

1.2. Rumusan Masalah

Berikut beberapa permasalahan utama yang telah dirumuskan dari hasil paparan latar belakang penelitian ini:

1. Bagaimana proses optimasi Algoritma *K-Nearest Neighbor* dengan Algoritma Genetika pada deteksi penyakit diabetes mellitus?
2. Bagaimana perbandingan tingkat akurasi deteksi penyakit diabetes mellitus dengan Algoritma *K-Nearest Neighbor* dan Algoritma *K-Nearest Neighbor* yang telah dioptimasi dengan Algoritma Genetika?

1.3. Tujuan Penelitian

Dari beberapa permasalahan yang berhasil dirumuskan, menyelesaikan dan memperoleh solusi dari permasalahan tersebut menjadi tantangan dan tujuan penelitian ini. Berikut tujuan yang ingin dicapai dalam penelitian ini:

1. Melakukan proses optimasi Algoritma *K-Nearest Neighbor* dengan Algoritma Genetika pada deteksi penyakit diabetes mellitus.
2. Mengetahui perbandingan tingkat akurasi deteksi penyakit diabetes mellitus dengan Algoritma *K-Nearest Neighbor* dan Algoritma *K-Nearest Neighbor* yang telah dioptimasi dengan Algoritma Genetika.

1.4. Manfaat Penelitian

Penelitian ini dapat memberikan manfaat bagi penulis dan untuk penelitian-penelitian selanjutnya sebagai berikut:

1. Memberikan pengetahuan terkait proses optimasi Algoritma *K-Nearest Neighbor* dengan Algoritma Genetika pada deteksi penyakit diabetes mellitus.
2. Memberikan pengetahuan terkait perbandingan tingkat akurasi deteksi penyakit diabetes mellitus dengan Algoritma *K-Nearest Neighbor* dan Algoritma *K-Nearest Neighbor* yang telah dioptimasi dengan Algoritma Genetika.

1.5. Batasan Masalah

Pada penelitian ini, diterapkan beberapa batasan permasalahan yang dijabarkan sebagai berikut:

1. Dataset yang digunakan merupakan data sekunder yang didapat dari website Kaggle. Dataset tersebut berisikan diagnosis gejala pada pasien beserta output (diabetes atau tidak). Terdapat 2 Dataset yang digunakan pada penelitian ini.
2. Dataset pertama bersumber dari National Institute of Diabetes and Digestive and Kidney Diseases yang memiliki 8 fitur atribut dan Dataset kedua berasal dari hasil kuisioner dari pasien di rumah sakit Sylhet Diabetes Hospital di Sylhet, Bangladesh yang memiliki 16 fitur atribut.
3. Keluaran yang dihasilkan adalah fitur-fitur yang diikutsertakan, parameter k , dan hasil akurasi deteksi penyakit diabetes mellitus paling optimal