

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes adalah salah satu penyakit kronis yang cukup mengkhawatirkan, karena bisa menyebabkan berbagai komplikasi serius seperti kebutaan, gagal ginjal, serangan jantung, bahkan kematian. Penyakit ini terjadi ketika tubuh tidak bisa memproduksi *Insulin* yang cukup atau tidak dapat memanfaatkan *Insulin* dengan baik. Ada dua tipe utama diabetes: Tipe 1, yang disebabkan oleh gangguan autoimun yang menyerang pankreas sebagai penghasil *Insulin*, dan Tipe 2, yang lebih umum, di mana tubuh tidak merespons *Insulin* dengan optimal. Tipe 2 sendiri menyumbang lebih dari 90% kasus diabetes di seluruh dunia (IDF Diabetes Atlas, "IDF Diabetes Atlas 10th Edition," 2021). Meskipun begitu, diabetes bisa dicegah atau setidaknya ditunda dengan cara mengetahui kondisi tubuh lebih awal.

Saat ini, diabetes masih bisa dideteksi lewat pemeriksaan medis, sayangnya prosesnya cukup memakan waktu dan banyak pasien yang enggan melakukannya. Pada tahap ini teknologi *machine learning* bisa jadi solusi untuk mempercepat proses deteksi dan prediksi diabetes menggunakan metode klasifikasi, sehingga pasien bisa lebih cepat mendapatkan penanganan yang tepat (Febrian dkk, 2023).

Salah satu pendekatan yang banyak digunakan untuk mendiagnosis penyakit lewat data adalah *data mining* menggunakan *machine learning*, yang merupakan teknik untuk menggali pengetahuan dari kumpulan data yang sudah ada. *Data mining* merupakan sebuah Teknik yang relative baru untuk mengekstraksi pengetahuan dari sebuah data setelah dilakukan pemrosesan data yang ada untuk dilakukan penilaian (Rasjid & Setiawan, 2017). Pada bidang kesehatan, *data mining* sering digunakan untuk memprediksi apakah seseorang berisiko mengidap diabetes atau tidak. Beberapa algoritma *machine learning* yang bisa digunakan untuk klasifikasi diabetes, seperti *Naive Bayes*, *Support Vector Machine (SVM)*, *Logistic Regression (LR)*, *Adaboost*, *Random Forest*, *K-Nearest Neighbors (KNN)*, *Decision Tree (DT)*, hingga *Neural Networks (NN)* (Rakesh & Sanjay, S., 2019).

Decision tree (Pohon Keputusan) telah menjadi salah satu arsitektur klasifikasi paling populer karena proses pengambilan keputusan yang dilakukan olehnya transparan dan hasil klasifikasinya mudah diinterpretasikan (Han dkk, 2023). *Decision tree* yang dibangun berdasarkan penggunaan entropi informasi dan banyak digunakan untuk tugas klasifikasi (Han dkk, 2023). *Decision Tree* adalah salah satu model kecerdasan buatan tertua paling berpengaruh untuk klasifikasi dan regresi. Kemudahan dan efektivitasnya adalah alasan mengapa algoritma ini masih digunakan baik sebagai model tunggal maupun sebagai bagian dari model *ensemble* salah satu contohnya adalah Random Forest (Palczyński dkk, 2023).

Random Forest (Hutan Acak) adalah metode pembelajaran *ensemble* untuk klasifikasi dan regresi yang membangun sejumlah pohon keputusan acak selama fase pelatihan dan melakukan prediksi dengan rata-rata hasilnya. Salah satu faktor utama yang menyumbang popularitas *random forest* adalah kemampuannya untuk diterapkan pada berbagai masalah prediksi dan memiliki sedikit parameter yang perlu diatur. Selain kemudahan penggunaannya, metode ini secara umum diakui karena keakuratannya dan kemampuannya dalam mengatasi ukuran sampel yang kecil, ruang fitur berdimensi tinggi, dan struktur data yang kompleks (Scornet dkk., 2015).

Dalam *implementasi Random Forest*, akurasi yang dicapai oleh algoritma bisa di tingkatkan salah satunya dengan memberikan *improve* pada algoritma. Penelitian yang dilakukan Maria Ali, Muhammad Nasim Hadier, dan Saima Anwar Lashari melakukan penelitian tentang teknik *ensemble* yang disebut "*Stacking Classifier*" yang dirancang untuk meningkatkan kinerja pengklasifikasi individu yang fleksibel dan mengurangi kemungkinan salah mengklasifikasi. Dengan menggunakan berbagai pengklasifikasi, termasuk *Naïve Bayes*, *KNN*, *Linear Discriminant Analysis*, dan *Decision Tree*, dengan *Random Forest* berfungsi sebagai meta pengklasifikasi. Dari penelitian tersebut hasil dari *F1-measure*, *Recall*, Akurasi, dan Presisi, pengklasifikasi *stacking* yang diusulkan mencapai akurasi lebih tinggi yaitu 97.35% dibandingkan dengan model saat ini seperti *Naïve Bayes*, *KNN*, *Decision Tree*, dan *Linear Discriminant Analysis*, yang masing-masing adalah 74.60%, 78.57%, dan 77.35% (Ali dkk., 2022).

Kinerja *Random Forest* sangat bergantung pada *hyperparameter*, pada penelitian Maria Ali (Ali dkk., 2022), pemilihan *hyperparameter* dilakukan dengan menggunakan 4 algoritma yang berbeda yang kemudian dilakukan “*stacking classifier*” untuk mencari hasil yang optimal. Dalam pemilihan *hyperparameter* yang optimal algoritma *Grey Wolf Optimizer (GWO)* dapat digunakan untuk mengoptimalkan *hyperparameter Random Forest* dengan melakukan pencarian global yang efisien dalam ruang *hyperparameter*.

Grey Wolf Optimizer (GWO) dikembangkan oleh Mirjalili pada tahun 2014 (Almufti dkk, 2021). GWO adalah algoritma *metaheuristic* yang termasuk dalam kategori ketiga (terinspirasi dari alam). Algoritma ini terinspirasi oleh proses berburu yang ditemukan pada serigala abu-abu. Hal ini unik karena mengikuti hierarki kepemimpinan serigala abu-abu. Serigala abu-abu terkenal dengan berburu dalam kelompok, dan tidak ada metode kecerdasan kawanan lainnya yang mengikuti perilaku berburu hierarkis ini (Almufti dkk, 2021).

Pada penelitian yang dilakukan Siva Shankar dan Manikandan, peneliti membandingkan akurasi *Fuzzy rule set Optimization* dengan *Grey Wolf Optimization*. Model dasar bekerja berdasarkan konsep *ant colony optimization* dan rule, yang tidak memberikan akurasi efektif karena algoritma ini hanya mengoptimalkan fitur lokal dan memberikan akurasi sebesar 71%. Model yang diusulkan didasarkan pada algoritma optimasi serigala abu-abu (*Grey Wolf Optimization*) dan mampu mengoptimalkan fitur secara global serta memberikan akurasi yang lebih tinggi dibandingkan dengan ACO (Siva & Manikandan, 2019).

Berdasarkan penelitian oleh Siva dan Manikandan (2019), terdapat kekurangan penelitian yang bisa diperbaiki. Meskipun jurnal tersebut berhasil membuktikan *Grey Wolf Optimization (GWO)* sebagai optimasi fitur yang unggul dibandingkan *Ant Colony Optimization*, model klasifikasinya yang berbasis *Fuzzy Rule Set* memiliki keterbatasan akurasi yang hanya mencapai 81.15% dan belum pernah mengeksplorasi penerapan GWO pada algoritma yang lebih kuat seperti *Random Forest*. Oleh karena itu, penelitian ini berjudul "Pemodelan Klasifikasi Penyakit Diabetes menggunakan *Random Forest* dengan *Grey Wolf Optimization*"

memiliki keunggulan signifikan. Keunggulannya terletak pada penggunaan *Random Forest* sebagai classifier yang lebih unggul dan robust, serta pendekatan hybrid inovatif yang mengintegrasikan kemampuan optimisasi global GWO yang telah terbukti. Lebih lanjut, penelitian ini menawarkan potensi optimisasi ganda yang lebih komprehensif, di mana GWO dapat diterapkan tidak hanya untuk seleksi fitur tetapi juga untuk optimasi *hyperparameter* pada *Random Forest*. Dengan demikian, penelitian ini berpotensi menghasilkan model klasifikasi diabetes yang lebih akurat dan andal dibandingkan pendekatan pada penelitian sebelumnya.

Pada penelitian pertama hasil maksimal *random forest* dapat diperoleh dengan melakukan klasifikasi awal menggunakan beberapa algoritma untuk mencari *hyperparameter*, kemudian pada penelitian kedua optimasi algoritma serigala abu-abu (*Grey Wolf Optimizer*) dinilai mampu mengoptimalkan fitur secara global serta memberikan akurasi yang lebih tinggi. Berdasarkan latar belakang penelitian ini mengusulkan Pemodelan Klasifikasi Penyakit Diabetes menggunakan *Random Forest* dengan *Grey Wolf Optimization*.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian tesis "Pemodelan Klasifikasi Penyakit Diabetes menggunakan *Random Forest* dengan *Grey Wolf Optimization*" dapat dirumuskan sebagai berikut:

1. Bagaimana membangun model klasifikasi penyakit diabetes menggunakan algoritma *Random Forest*?
2. Bagaimana menerapkan algoritma *Grey Wolf Optimization* untuk optimasi *hyperparameter* pada model klasifikasi *Random Forest*?
3. Bagaimana performa model klasifikasi penyakit diabetes sebelum dan sesudah dioptimalkan menggunakan *Grey Wolf Optimization*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah, maka dapat diperoleh tujuan penelitian sebagai berikut:

1. Menyusun model klasifikasi diabetes dengan algoritma *Random Forest*.

2. Implementasi algoritma *Grey Wolf Optimization* untuk optimasi *hyperparameter* pada model klasifikasi *Random Forest*.
3. Menganalisis performa model sebelum dan sesudah dioptimasi menggunakan *Grey Wolf Optimization*.

1.3.2 Manfaat Penelitian

Adapun beberapa manfaat dari penelitian ini sebagai berikut:

1. Bagi para peneliti, penelitian ini diharapkan bisa menjadi metode baru yang bisa digunakan untuk memprediksi penyakit di Indonesia. Sehingga dalam praktiknya prediksi penyakit di Indonesia menjadi lebih mudah dan cepat dalam penanganannya.
2. Bagi penelitian selanjutnya, penelitian ini diharapkan dapat menjadi referensi penelitian sejenis tentang prediksi penyakit menggunakan algoritma lain selain *Random Forest*.