

BAB I

PENDAHULUAN

1.1 Latar Belakang

Negara Indonesia saat ini sedang mengalami masa transisi epidemiologi. Epidemiologi merupakan ilmu yang berfokus pada frekuensi dan distribusi masalah kesehatan di antara sekelompok orang (Sitorus, 2012). Beberapa bentuk epidemiologi salah satunya yaitu distribusi penyakit tidak menular seperti diabetes yang frekuensinya semakin meningkat seiring berjalannya waktu.

Diabetes merupakan penyakit yang mengganggu sistem metabolisme sehingga menyebabkan ketidakmampuan pankreas memproduksi insulin secara maksimal (Safitri dan Nurhayati, 2019). Insulin adalah hormon yang mengatur kadar gula dalam darah. Kondisi meningkatnya konsentrasi kadar gula dalam darah disebut *hyperglikemia* (Pusat Data dan Informasi Kementerian Kesehatan RI, 2014).

Konsentrasi kadar gula darah normal setelah 8 jam berpuasa pada pagi hari (tes gula darah puasa) berkisar antara 70-110 mg/dL. Konsentrasi kadar gula darah normal 2 jam setelah makan (tes toleransi glukosa oral) berkisar antara 120-140 mg/dL. Seseorang dikatakan menderita diabetes, jika hasil tes kadar gula darah puasanya > 126 mg/dL dan hasil tes toleransi glukosa oral > 200 mg/dL (Suryo, 2009).

Penyakit diabetes yang juga dikenal sebagai penyakit gula darah, seringkali dijuluki "*the silent killer*" karena rata-rata penderitanya tidak menyadari kondisi ini, sampai terjadinya komplikasi (Kementerian Kesehatan RI, 2014). Diabetes dapat

menyebabkan komplikasi yang menyerang hampir seluruh bagian tubuh, mulai dari kulit hingga jantung.

International Diabetes Federation (IDF) pada tahun 2021 menyebutkan bahwa sebanyak 10,5% populasi orang dewasa usia 20 hingga 79 tahun menderita diabetes dan hampir setengahnya tidak menyadari hal tersebut. Proyeksi IDF pada tahun 2045 menunjukkan bahwa sekitar 783 juta jiwa atau satu dari delapan orang dewasa akan hidup dengan diabetes, serta dapat mengalami peningkatan sebesar 46%. Lebih dari 90% penderita diabetes saat ini mengidap diabetes tipe II yang disebabkan oleh faktor sosial, ekonomi, demografi, lingkungan, dan genetik.

Indonesia menempati peringkat ke-34 dari 204 negara di dunia, serta peringkat pertama sebagai penderita diabetes tipe I terbanyak di wilayah ASEAN, menurut laporan *International Diabetes Federation* (IDF). Pada tahun 2022, jumlah penderita diabetes tipe I di Indonesia mencapai 41,8 ribu jiwa. Mayoritas penderita diabetes tipe I di Indonesia berusia 20 hingga 59 tahun, tetapi jumlah penderita yang berusia muda juga cukup banyak. Proporsi penderita diabetes tipe I menurut Kementerian Kesehatan hanya sekitar 10% dari total penderita diabetes, sehingga jumlahnya lebih sedikit dibandingkan penderita diabetes tipe II.

Diabetes merupakan masalah kesehatan yang serius di Indonesia, jumlah penderita setiap tahunnya terus meningkat seiring dengan penambahan penduduk, gaya hidup kurang sehat, dan tingginya angka obesitas (Aryastami dan Tarigan, 2017). Penyakit diabetes disebabkan oleh dua faktor yaitu faktor yang dapat dimodifikasi dan tidak dapat dimodifikasi (Rovy, 2018). Faktor yang dapat dimodifikasi mencakup pola makan, kebiasaan merokok, hipertensi, stres, aktivitas fisik, kebiasaan meminum

alkohol, dan sebagainya. Faktor yang tidak dapat dimodifikasi terdiri dari umur, jenis kelamin, dan faktor keturunan (Ujani, 2016).

Berbagai kasus penyakit diabetes disebabkan oleh banyak faktor, sehingga perlu dilakukannya analisis berdasarkan faktor yang mempengaruhinya. Salah satu teknik analisis yang dapat diterapkan yaitu *data mining*. *Data mining* merupakan penggabungan dari disiplin ilmu untuk menemukan pola baru dari data yang besar, mencakup metode dari *Artificial Intelligence* (AI), *machine learning*, *statistics*, dan *database systems* (Aris dan Benyamin, 2019).

Keuntungan mengolah data dengan memanfaatkan teknik *data mining* agar mendapatkan sumber pengetahuan dan informasi baru. Pemanfaatan tersebut tidak sebatas pada ilmu teknologi, melainkan dapat dimanfaatkan dalam bidang kesehatan. Teknik *data mining* yang sesuai dengan data kesehatan yaitu klasifikasi, karena tingkat penyakit yang diderita oleh pasien dapat diklasifikasikan sehingga memudahkan dokter maupun tenaga medis dalam memberikan solusi pengobatan yang tepat.

Klasifikasi merupakan teknik pembelajaran data dalam *machine learning* yang melibatkan penggunaan dataset yang sudah diberi label (Retnoningsih dan Pramudita, 2020). Klasifikasi berguna untuk memprediksi dan menemukan pola yang berharga dari data yang besar. Ketepatan pengklasifikasian objek sangat penting karena teknik klasifikasi yang baik adalah meminimalkan kesalahan prediksi serta mampu melakukan pembelajaran data klasifikasi dengan baik. Salah satu algoritma klasifikasi yaitu CART.

Algoritma CART (*Classification and Regression Trees*) dikembangkan pada tahun 1984 oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J.

Stone untuk kebutuhan analisis klasifikasi pada variabel dependen baik berskala nominal, ordinal, maupun rasio. Algoritma CART dibagi menjadi dua metode yaitu metode klasifikasi dan regresi. Apabila variabel dependen bertipe kategorik, maka algoritma CART akan menghasilkan pohon klasifikasi (*classification tree*). Namun, apabila variabel dependen bertipe kontinu atau numerik, maka algoritma CART akan menghasilkan pohon regresi (*regression tree*).

Penelitian Pratama *et al.* (2021) yaitu melakukan seleksi fitur dan penanganan data tidak seimbang menggunakan RFECV dan ADASYN pada algoritma CART. Hasil penelitian tersebut, terjadi penurunan akurasi pada penanganan seleksi fitur dan data tidak seimbang menggunakan RFECV dan ADASYN pada algoritma CART. Hasil akurasi sebelumnya sebesar 85,10% namun setelah dilakukan penanganan seleksi fitur dan data tidak seimbang hasil akurasi menurun menjadi 79,10%.

Penelitian yang dilakukan Sulistiyanto (2018) yaitu penerapan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO) dalam memprediksi siswa lolos seleksi perguruan tinggi. Hasil akurasi algoritma C4.5 sebesar 84,84%, sedangkan algoritma C4.5 berbasis PSO dapat meningkatkan hasil akurasi menjadi 94,56%. Penelitian Pratiwi *et al.* (2020) yaitu membandingkan klasifikasi algoritma C5.0 dengan CART studi kasus data sosial kepala keluarga masyarakat Desa Teluk Baru Kecamatan Muara Ancalong tahun 2019. Hasil akurasi algoritma CART lebih baik dari algoritma C5.0 dengan rata-rata nilai akurasi sebesar 84,63%. Oleh sebab itu, penelitian ini menggunakan algoritma CART karena dinilai sesuai untuk mengklasifikasikan data kesehatan, seperti data penderita diabetes.

Berdasarkan penelitian terdahulu, permasalahan yang terdapat pada algoritma CART yaitu ketidakstabilan terhadap perubahan kecil dalam data latih karena dapat memberikan perubahan yang signifikan terhadap pohon yang dihasilkan (Sutton, 2015), kemungkinan terjadinya *overfitting*, dan kurang efisien terhadap data berdimensi tinggi. Banyaknya variabel independen yang digunakan dalam pemodelan dapat mengurangi kinerja algoritma CART, sehingga hasil akurasi menjadi rendah serta pohon yang terbentuk menjadi kompleks.

Dalam mengatasi permasalahan tersebut, perlu dilakukannya optimasi dengan algoritma PSO untuk meningkatkan hasil akurasinya. PSO merupakan sebuah algoritma metaheuristik yang terinspirasi dari kawanan burung atau ikan dalam mencari makanannya. Algoritma PSO dianggap lebih efisien karena kumpulan partikel yang dioptimalkan dalam ruang fitur tidak hanya mengingat pengalaman individu saja, tetapi dapat mengingat pengalaman global partikel (Arifin, 2017).

Berdasarkan uraian tersebut, peneliti akan membahas analisis data penderita diabetes menggunakan algoritma CART berbasis optimasi dengan judul “Optimasi Algoritma *Classification and Regression Trees* (CART) Menggunakan *Particle Swarm Optimization* (PSO) Pada Penderita Diabetes”.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah yang diambil pada penelitian ini sebagai berikut:

1. Bagaimana perbandingan hasil akurasi algoritma CART dengan algoritma CART berbasis PSO?

2. Bagaimana pohon keputusan beserta *rule* yang terbentuk dari algoritma terbaiknya?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah penerapan algoritma CART dengan metode *Classification Tree* karena variabel dependen bertipe kategorik. Optimasi PSO digunakan untuk seleksi fitur terhadap variabel independen pada algoritma CART. Data diperoleh dari Kaggle sebanyak 952 data penderita diabetes tipe II dengan 14 variabel independen dan satu variabel dependen.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, tujuan dari penelitian ini adalah sebagai berikut:

1. Dapat menerapkan pola klasifikasi pada penderita diabetes baik dengan algoritma CART maupun algoritma CART yang dioptimasi dengan algoritma PSO.
2. Dapat menerapkan optimasi PSO dalam pemilihan fitur, serta menilai apakah dapat meningkatkan hasil akurasi algoritma CART.
3. Dapat membentuk pohon keputusan beserta *rule* dari algoritma terbaiknya (algoritma dengan ukuran kinerja klasifikasi yang lebih baik).