

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining merupakan penggalian dan analisis data berukuran besar menggunakan peranti otomatis atau semi otomatis yang bertujuan untuk mendapatkan pola yang memiliki makna (Muflikhah dkk., 2018). Di dalam data mining terdapat kegiatan mengumpulkan dan memakai data historis sebagai pengetahuan untuk menghasilkan data baru yang sudah diklasifikasikan dalam data berskala besar. Sehingga menghasilkan *output* yang bisa digunakan sebagai pilihan untuk pengambilan kebijakan di masa yang akan datang. (Buuololo, 2020)

Klasifikasi merupakan salah satu metode yang ada pada data mining. Metode ini digunakan untuk menemukan pola yang akan membedakan kategori atau kelas data, dengan tujuan agar dapat memperkirakan label yang belum diketahui dari suatu objek kelas data. (Wanto dkk., 2020)

Salah satu algoritma yang sering digunakan dalam klasifikasi adalah metode K-Nearest Neighbor (KNN). Algoritma ini mengklasifikasikan data baru yang belum diketahui kelasnya dengan cara mencari tetangga terdekat dari data tersebut (Aprilian dkk., 2020). KNN juga merupakan teknik klasifikasi paling dasar dan sederhana karena bisa dengan efektif melakukan pelatihan data dengan jumlah yang banyak dalam waktu yang relatif cepat. Di sisi lain metode KNN memiliki kelemahan yaitu dalam penentuan parameter K yang masih bias, sebab belum ada rumus untuk menentukan parameter tersebut. Sehingga, tingkat akurasi dari kinerja

KNN yang dihasilkan tidak optimal jika dalam menentukan parameter K kurang tepat (Zadlyka, 2021). Oleh karenanya, dibutuhkan suatu metode optimasi untuk mengoptimalkan penentuan parameter K.

Beberapa penelitian telah dilakukan untuk mengoptimalkan nilai parameter K dengan cara menggabungkan algoritma KNN dengan beberapa algoritma, diantaranya penelitian (Lesmono dkk., 2017) menggabungkan algoritma genetika dengan KNN dan terbukti bahwa algoritma genetika mampu membantu metode KNN menghasilkan nilai akurasi lebih tinggi dengan membantu dalam penentuan nilai parameter K yang bias. Algoritma ini dapat melakukan pencarian dengan sangat efektif untuk menyelesaikan permasalahan optimasi dengan mekanisme evolusi.

Berdasarkan beberapa pokok permasalahan tersebut, maka dilakukan penelitian untuk mengoptimasi parameter K pada KNN menggunakan algoritma genetika pada klasifikasi pemberian obat. Selanjutnya, digunakan *confusion matrix* untuk membandingkan hasil klasifikasi penggabungan metode KNN dan algoritma genetika dengan hasil klasifikasi yang seharusnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka diperoleh rumusan masalah untuk melakukan penelitian mengenai optimasi nilai K pada algoritma KNN menggunakan algoritma genetika untuk klasifikasi pemberian obat yaitu sebagai berikut :

1. Bagaimana menerapkan algoritma genetika dalam optimasi untuk penentuan parameter K pada algoritma KNN?
2. Bagaimana menerapkan algoritma KNN dengan parameter K yang telah teroptimasi dengan algoritma genetika untuk melakukan klasifikasi obat?

1.3 Ruang Lingkup

Adapun ruang lingkup untuk memfokuskan penelitian ini yaitu:

1. Metode klasifikasi yang digunakan adalah algoritma KNN
2. Metode optimasi yang digunakan adalah algoritma genetika dengan mengoptimalkan nilai K pada algoritma KNN
3. Skema pengkodean yang digunakan adalah *binary encoding*
4. Proses persilangan atau rekombinasi menggunakan metode *one-cut point crossover*
5. Proses mutasi menggunakan metode *random mutation*
6. Proses seleksi menggunakan metode *elitism*
7. Kualitas hasil klasifikasi dihitung menggunakan metode evaluasi *confusion matrix*
8. Dataset yang digunakan yaitu dataset obat (*drug*) sejumlah 200 *record* data yang diperoleh dari *platform* Kaggle. Dimana dari data tersebut diambil sebanyak 75% untuk data latih dan sebanyak 25% untuk data uji.
9. Dataset tersebut memiliki 5 atribut yaitu usia (*age*), jenis kelamin (*sex*), tekanan darah (*Blood Pressure*), kadar kolesterol (*Cholesterol level*), dan sodium-potasium

10. Data diklasifikasikan ke dalam 5 kategori yaitu obat A, obat B, obat C, obat X dan obat Y
11. Aplikasi yang dibuat berbasis web
12. Aplikasi yang dibuat menampilkan output dari hasil pencarian parameter K optimal dari algoritma genetika, hasil klasifikasi dari KNN, dan kualitas hasil klasifikasi dari *confusion matrix*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Menguji metode algoritma genetika dalam optimasi untuk penentuan parameter K paling optimal pada algoritma KNN
2. Menguji hasil optimasi parameter K pada KNN menggunakan algoritma genetika untuk melakukan klasifikasi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining yang kemudian dapat dikembangkan lebih lanjut maupun dijadikan referensi pada kasus penelitian lain dengan mempertimbangkan karakteristik pada penggunaan algoritma genetika untuk optimasi parameter K pada KNN.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini terdiri dari lima bab yang akan diuraikan pada penjelasan berikut:

a. Bab I Pendahuluan

Pada bab ini berisi pengantar sekaligus memberikan hipotesa awal mengenai penelitian yang sedang dilakukan. Bab I berisi latar belakang, rumusan masalah, ruang lingkup, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

b. Bab II Tinjauan Pustaka dan Dasar Teori

Pada bab II ini berisi tinjauan pustaka dan dasar teori. Dimana tinjauan pustaka digunakan sebagai acuan primer sedangkan dasar teori berisi definisi dan uraian dari berbagai sumber yang telah dipublikasikan sehingga dapat digunakan untuk memahami pengertian dasar dan istilah dalam penelitian ini.

c. Bab III Metode Penelitian

Pada bab III ini akan berisi langkah-langkah lengkap dalam menyelesaikan penelitian ini yang terdiri dari bahan dan data, prosedur dan pengumpulan data, serta analisis dan rancangan sistem.

d. Bab IV Implementasi dan Pembahasan

Pada bab IV ini berisi laporan hasil penelitian berupa implementasi sistem, implementasi antarmuka, uji coba sistem, hasil uji coba sistem, dan pembahasan sistem.

e. Bab V penutup

Pada bab V ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk pengembangan penelitian selanjutnya.