

BAB I

PENDAHULUAN

1.1 Latar Belakang

Akurasi selalu menjadi isu penting dalam klasifikasi data. Semakin tinggi tingkat akurasi suatu metode klasifikasi, semakin baik hasil klasifikasinya. Maka wajar jika kemudian, berbagai pendekatan diteliti guna memperoleh tingkat akurasi klasifikasi yang lebih tinggi. Namun, suatu pendekatan klasifikasi terkadang menghasilkan dampak yang berbeda pada himpunan data (*data set*) yang berbeda. Dengan kata lain, terkadang suatu pendekatan klasifikasi mampu meningkatkan akurasi ketika diterapkan pada suatu himpunan data, namun tidak demikian ketika diterapkan pada himpunan data yang lain. Dibutuhkan pendekatan klasifikasi yang sesuai untuk memperoleh peningkatan akurasi yang diharapkan.

K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi yang paling sederhana di antara sekian banyak metode klasifikasi yang ada. Hal ini dapat dipahami, karena KNN merupakan metode klasifikasi yang termasuk jenis *lazy learning*. Dalam melakukan klasifikasi, KNN membandingkan suatu sampel *query* dengan setiap sampel yang ada di dalam himpunan pelatihan (*training set*). Hasil klasifikasi KNN akan bergantung pada sampel-sampel pelatihan yang memiliki jarak terdekat dari sampel *query* tersebut.

Di balik kesederhanaannya, KNN juga memiliki tingkat akurasi yang patut diperhitungkan. Dalam kasus diagnosis penyakit liver, KNN menunjukkan tingkat akurasi yang tinggi [1]. Tentu akan sangat baik apabila terdapat suatu pendekatan yang dapat meningkatkan akurasi klasifikasi KNN. Dengan begitu, diharapkan bahwa diagnosis penyakit liver yang dihasilkan oleh KNN akan semakin akurat.

Beragam pendekatan untuk meningkatkan akurasi klasifikasi KNN telah diteliti, baik pada tahap pra-pemrosesan (*preprocessing*) maupun pengujian (*testing*). Salah satu pendekatan tersebut adalah diskretisasi berbasis *entropy*. Diskretisasi berbasis *entropy* termasuk metode diskretisasi data, yang mana merupakan salah satu bentuk pra-pemrosesan data. Dalam banyak penelitian, diskretisasi berbasis *entropy* mampu meningkatkan akurasi klasifikasi.

Berdasarkan fakta-fakta tersebut, dibangun sebuah hipotesis bahwa penerapan diskretisasi berbasis *entropy* sebagai metode pra-pemrosesan akan berdampak pada peningkatan akurasi klasifikasi oleh KNN dalam kasus diagnosis penyakit liver. Oleh karena itu, dalam penelitian ini diskretisasi berbasis *entropy* akan diterapkan sebagai metode pra-pemrosesan, sedangkan KNN akan digunakan sebagai metode klasifikasi. Nantinya, tingkat akurasi yang dihasilkan oleh pendekatan ini akan dibandingkan dengan tingkat akurasi yang dihasilkan oleh KNN tanpa menggunakan metode pra-pemrosesan apapun. Hal ini dilakukan untuk melihat dampak penerapan diskretisasi berbasis *entropy* terhadap akurasi klasifikasi KNN, sekaligus untuk membuktikan hipotesis yang telah disebutkan.

1.2 Perumusan Masalah

Dari latar belakang yang telah diuraikan sebelumnya, dapat dirumuskan masalah-masalah yang akan dipecahkan dalam penelitian Tugas Akhir ini, yaitu:

1. Bagaimana penerapan diskretisasi berbasis *entropy* sebagai metode pra-pemrosesan pada himpunan data penyakit liver?
2. Bagaimana penerapan KNN sebagai metode klasifikasi pada himpunan data penyakit liver?
3. Bagaimana melihat dampak penerapan diskretisasi berbasis *entropy* terhadap tingkat akurasi klasifikasi dengan KNN pada himpunan data penyakit liver?

1.3 Tujuan dan Manfaat

Tujuan dilaksanakannya penelitian Tugas Akhir ini, yaitu:

1. Menerapkan diskretisasi berbasis *entropy* sebagai metode pra-pemrosesan pada himpunan data penyakit liver,
2. Menerapkan KNN sebagai metode klasifikasi pada himpunan data penyakit liver, dan
3. Melihat dampak penerapan diskretisasi berbasis *entropy* terhadap tingkat akurasi klasifikasi dengan KNN pada himpunan data penyakit liver.

Adapun manfaat yang dapat diperoleh dari adanya penelitian Tugas Akhir ini, antara lain:

1. Dapat diketahui dampak dari penerapan diskretisasi berbasis *entropy* terhadap akurasi diagnosis penyakit liver menggunakan KNN, dan
2. Hasil akhir penelitian ini dapat digunakan oleh penelitian lain yang terkait sebagai landasan untuk mengembangkan apa yang telah dilakukan oleh penelitian ini.

1.4 Batasan Masalah

Dalam penelitian Tugas Akhir ini, ditetapkan beberapa batasan masalah, antara lain:

1. Performa klasifikasi yang diamati dalam penelitian ini hanyalah tingkat akurasi,
2. Algoritma KNN yang digunakan dalam penelitian ini adalah algoritma KNN dasar, dan
3. Fungsi jarak (*distance function*) yang digunakan dalam KNN adalah HVDM.

1.5 Metodologi Pengerjaan Tugas Akhir

Penelitian Tugas Akhir ini terbagi menjadi beberapa tahap, yaitu:

1. Studi Literatur
Pada tahap ini, konsep penelitian Tugas Akhir diperoleh dari referensi-referensi yang relevan. Referensi tersebut dapat berupa buku, makalah, halaman *web*, dan sebagainya. Dari tahap ini, diperoleh gambaran mengenai apa yang akan dilakukan dalam penelitian Tugas Akhir.
2. Penyiapan Data
Pada tahap ini, data yang dibutuhkan dalam penelitian Tugas Akhir diunggah dari sumbernya. Data yang telah diunggah kemudian dimasukkan ke dalam basis data sistem.
3. Perancangan Sistem
Pada tahap ini, dibuat rancangan basis data dan alur sistem. Rancangan basis data dilengkapi dengan penjelasan mengenai

struktur tabel-tabel yang ada di dalamnya. Alur sistem digambarkan dalam bentuk diagram alir (*flow chart*).

4. Pengembangan Sistem

Pada tahap ini, rancangan sistem yang telah dibuat pada tahap sebelumnya, diwujudkan ke dalam program aplikasi. Perubahan rancangan sistem dapat dilakukan sewaktu-waktu apabila dibutuhkan.

5. Uji Coba dan Evaluasi

Pada tahap ini, beberapa skenario uji coba dilakukan terhadap sistem yang telah dibangun. Proses *debugging* dapat dilakukan sewaktu-waktu apabila ditemukan kesalahan pada sistem.

6. Penyusunan Laporan

Pada tahap ini, seluruh rangkaian penelitian Tugas Akhir didokumentasikan dalam format yang telah ditetapkan.

1.6 Sistematika Penulisan

Laporan Tugas Akhir ini disusun dalam beberapa bab, yaitu:

1. Pendahuluan

Bab ini memuat latar belakang penelitian Tugas Akhir, perumusan masalah, tujuan dan manfaat penelitian Tugas Akhir, batasan masalah, metodologi penelitian, serta sistematika penulisan laporan Tugas Akhir.

2. Studi Pustaka

Bab ini terdiri atas 2 subbab, yaitu Tinjauan Pustaka dan Landasan Teori. Tinjauan Pustaka berisi gambaran singkat dari sumber-sumber pustaka yang dijadikan referensi, sedangkan Landasan teori berisi penjelasan tentang konsep-konsep dan teori-teori yang digunakan dalam penelitian Tugas Akhir.

3. Perancangan Sistem

Bab ini memaparkan tentang rancangan basis data dan alur sistem, baik secara umum maupun rinci.

4. Implementasi Sistem

Bab ini berisi tentang spesifikasi lingkungan uji coba sistem, tahap-tahap uji coba dan hasil akhir penelitian.

5. Penutup

Bab ini memuat kesimpulan penelitian Tugas Akhir, juga saran-saran dan peluang untuk pengembangan penelitian.