

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Gagal jantung adalah penyakit kardiovaskular yang disebabkan oleh jantung yang memompa darah ke seluruh tubuh dan mengganggu fungsi sistem sirkulasi fisiologis. Prevalensi gagal jantung di Asia Tenggara mencapai 3 kali lipat jika dibandingkan dengan negara Eropa dan Amerika yaitu sebesar 4.5–6.7% : 0.5–2% (Lam, 2015). Angka kematian akibat gagal jantung tinggi yaitu mencapai 50% dalam 5 tahun (Yancy dkk, 2013). Data di Indonesia tahun 2018 menunjukkan bahwa gagal jantung termasuk dalam 10 penyakit tidak menular di Indonesia dan mendapatkan hasil 229.696 (0,13%) orang menderita gagal jantung. Provinsi Jawa Tengah pada tahun 2018 jumlah penderita gagal jantung sebanyak 3.493 (1,6%) orang (Kristinawati dkk, 2019). Penting untuk mengetahui tentang faktor-faktor yang mempengaruhi kematian akibat penyakit ini agar dapat dilakukan pencegahan dini. Salah satu faktor yang dapat meningkatkan kejadian gagal jantung adalah gaya hidup yang tidak sehat dan kemampuan perawatan diri (Lewis, 2019). Penyakit gagal jantung tidak dapat disembuhkan, tetapi banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup. Seiring dengan perkembangan zaman dalam dunia kesehatan terjadi juga perkembangan yang cukup pesat. Salah satunya adalah pemanfaatan *Machine Learning* dan Data Mining dalam dunia kesehatan.

Klasifikasi merupakan bagian dari data mining, dimana data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database (Sumarlin, 2015). Klasifikasi adalah teknik pengolahan data yang membagi suatu data objek menjadi beberapa kelas sesuai dengan jumlah kelas yang dibutuhkan. Salah satu algoritma metode klasifikasi yaitu *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* adalah metode untuk melakukan klasifikasi pada objek berdasarkan data latih yang jaraknya paling dekat dengan ojekt (Devi, 2017). Keakuratan algoritma *K-Nearest Neighbor* dipengaruhi oleh atribut yang

tidak relevan atau jika bobot kriteria tersebut tidak setara dengan relevansinya terhadap klasifikasi. Algoritma *K-Nearest Neighbor* juga memiliki keunggulan pelatihan yang sangat cepat dan sederhana sehingga mudah dipelajari (Iriantoro et al., 2018). Namun ada kekurangan pada algoritma *K-Nearest Neighbor* yaitu perlu penentuan nilai K dan untuk pemilihan atribut terbaik (Bode, 2017). Untuk pemilihan atribut terbaik maka dibutuhkan optimasi yaitu dengan menggunakan seleksi fitur.

Backward Elimination merupakan salah satu tahap yang dapat dilakukan dalam melakukan seleksi atribut pada tahap *preprocessing*. *Backward Elimination* yang berfungsi untuk mengoptimalkan kinerja suatu model dengan sistem kinerja mundur yang digunakan untuk memilih atribut paling relevan. *Backward Elimination* digunakan untuk seleksi fitur, dibandingkan dengan *Forward Elimination* yaitu *Forward Elimination* mempertahankan semua atribut baik yang independen maupun tidak, sehingga atribut yang tidak independen menjadi tidak signifikan dan mengurangi jumlah kuadrat residual sedangkan *Backward elimination* membuang atribut – atribut yang tidak independen terhadap data yang diolah (Sulaehani, 2016). Kinerja terbaik dibuktikan dengan *sensitivitas*, *spesifisitas*, dan akurasi yang tinggi (Ali et al., 2014). Pada penelitian ini *Backward Elimination* digunakan untuk seleksi fitur yang bertujuan untuk meningkatkan nilai akurasi.

Pada penelitian terdahulu yang dilakukan oleh Ikhsan Wisnuadji Gamadarenda dkk (2018) yang berjudul Implementasi Data Mining Untuk Deteksi Penyakit Ginjal Kronis (PGK) Menggunakan *K-Nearest Neighbor* dengan *Backward Elimination*. Pada penelitian ini menggunakan atribut yang telah diseleksi menggunakan *Backward Elimination* berhasil menekankan biaya pemeriksaan hingga 73,36%. Selanjutnya untuk hasil pendeteksi penyakit menggunakan Algoritma *K-Nearest Neighbor* menghasilkan nilai akurasi sebesar 99,25%. Pada penelitian yang dilakukan oleh Hasran (2020) yang berjudul Klasifikasi Penyakit Jantung Menggunakan Metode *K-Nearest Neighbor*, pada penelitian ini menggunakan data penyakit jantung dan mendapatkan K terbaik pada $K=6$ dengan

akurasi sebesar 85%. Melihat dari penelitian- penelitian sebelumnya perbedaan dengan penelitian sebelumnya yaitu dapat dilihat dari data atribut yaitu 13 atribut yang digunakan dan pada tujuan penelitian.

Berdasarkan penjelasan diatas, maka penulis tertarik untuk mengklasifikasi penyakit gagal jantung dengan judul “Penerapan *Backward Elimination* Untuk Seleksi Fitur Pada Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Penyakit Gagal Jantung”.

1.2 Rumusan Masalah

Dari latar belakang yang sudah dijelaskan, maka diambil perumusan masalah dalam penelitian ini yaitu seberapa tinggi nilai akurasi, *presisi*, *recall* algoritma *K-Nearest Neighbor* apabila *Backward Elimination* diterapkan pada proses seleksi fitur?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini yaitu menerapkan *Backward Elimination* untuk seleksi fitur pada algoritma *K-Nearest Neighbor*.

1.4 Batasan Masalah

Agar pembahasan permasalahan dalam penelitian ini tidak terlalu luas, maka diberikan batasan masalah sebagai berikut:

1. Data kasus yang digunakan pada penelitian ini adalah data klasifikasi penyakit gagal jantung yang diambil dari situs *UCI Machine Learning Repository*.
2. Jumlah data yang digunakan berjumlah 299 data yang terdiri dari 12 fitur atribut dan 1 atribut output.
3. Penelitian ini dilakukan dengan menggunakan *tools Rapid Miner* yang merupakan salah satu *tools open source*.

1.5 Manfaat Penelitian

Adapun manfaat yang didapatkan dari penelitian ini adalah:

1. Dapat memberikan informasi tentang cara kerja dari Algoritma *K-Nearest Neighbor* menggunakan seleksi fitur *Backward Elimination* dalam mengklasifikasi data penyakit gagal jantung.
2. Dapat memberikan informasi yang dapat di gunakan sebagai acuan referensi untuk di gunakan dalam penelitian selanjutnya.

