

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Gagal jantung kongestif atau *Congestive Heart Failure* (CHF) merupakan penyebab kematian nomor satu di dunia (Maulida, 2018). Jumlah gagal jantung di Amerika Serikat kira-kira 5,7 juta orang dewasa dan 550.000 kasus baru didiagnosis setiap tahunnya diagnosis (Mozaffarian, et al., 2016). Gagal jantung berkontribusi terhadap 287.000 kematian per tahun. Sekitar setengah dari orang yang mengalami gagal jantung meninggal dalam waktu lima tahun setelah di diagnosis (Maulida, 2018). Penyakit gagal jantung merupakan penyakit dengan kematian tertinggi. Pengobatan yang lama dan sering keluar masuk rumah sakit akan memberikan dampak terhadap kualitas hidup pasien terhadap penyakit yang dialaminya. Dampak yang dialami merupakan reaksi psikologis terhadap dampak dari gagal jantung yang dihadapi oleh pasien (Ihdaniyati, 2009). Hampir semua pasien yang mempunyai penyakit jantung mengetahui jantung merupakan organ terpenting dan jantung yang rusak akan mengancam kesehatan. Hal ini yang menyebabkan pasien gagal jantung merasa cemas, kesulitan tidur, merasa depresi dan merasa putus asa akan penyakit yang dideritanya (Ihdaniyati, 2009). Seiring perkembangan zaman dalam dunia kesehatan terjadi perkembangan pula. Salah satunya penggunaan *Machine Learning* dan *Data Mining* dalam dunia kesehatan.

*Data Mining* merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut, selain itu *Data Mining* merupakan salah satu teknik yang dapat melakukan klasifikasi dengan teknik pengelompokan data. Dalam dunia kesehatan *Data Mining* dapat dimanfaatkan untuk mendapatkan informasi seperti penentuan kriteria suatu penyakit, misalnya tingkat kebutuhan transfusi darah dari penderita *Thalassaemia*. Kebutuhan waktu transfusi yang berbeda-beda dari setiap penderita menjadi sebuah masalah dalam mempersiapkan pemberian jumlah obat terafi kelasi besi dan kesiapan pendonor darah (Sulastri,2017). Ada beberapa teknik yang

diterapkan dalam *Data Mining*, salah satunya teknik klasifikasi. Teknik klasifikasi adalah sebuah metode dari *Data Mining* yang digunakan untuk memprediksi kategori atau kelas dari suatu *data instance* berdasarkan sekumpulan atribut-atribut dari data tersebut. Teknik klasifikasi sendiri memiliki banyak algoritma yang dapat digunakan dalam menyelesaikan permasalahan tergantung data yang sedang diteliti. *K-Nearest Neighbor* (K-NN) merupakan salah satu teknik klasifikasi. *K Nearest Neighbor* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan atribut dan sampel latih. Dari dataset atau sampel latih pada penelitian sering muncul permasalahan-permasalahan diantaranya *imbalanced data* atau ketidakseimbangan data. *Imbalanced data* (ketidakseimbangan data) adalah salah satu masalah utama yang muncul dalam deteksi anomali pada dataset yang bersifat *real time*. Dataset dianggap tidak seimbang jika salah satu kelasnya memiliki dominasi yang sangat besar dibandingkan dengan kelas lainnya (Somasundaram,2016). Beberapa teknik digunakan untuk mengatasi permasalahan data yang tidak seimbang, salah satunya adalah teknik *Synthetic Minority Oversampling Technique (SMOTE)*. *Synthetic Minority Oversampling Technique (SMOTE)* adalah salah satu metode yang digunakan untuk menangani kasus ketidakseimbangan data pada dataset (Santoso, et al., 2017). Luasnya penggunaan teknik *Data Mining* dalam dunia kesehatan dapat dilihat dari penelitian terdahulu yang telah dihimpun dalam penelitian ini sebagai bahan pertimbangan dan landasan oleh peneliti.

Penelitian terdahulu terhadap penggunaan *Data Mining* dalam dunia kesehatan yang berhasil dihimpun dan memiliki kedekatan atau kemiripan dengan penelitian yang diajukan di antaranya adalah pada penelitian dengan judul *Machine Learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone* menggunakan metode *Logistic Regression* memperoleh hasil yaitu metode ini mampu melakukan klasifikasi terhadap pasien untuk menentukan tingkat keberlangsungan pasien terhadap penyakit gagal jantung dengan nilai akurasi mencapai 83% (Davide,2020). Pada penelitian dengan judul *Case Based Reasoning Diagnosis Penyakit Difteridengan Algoritma K-Nearest Neighbor* disebutkan hasil yang diperoleh dalam penelitian ini adalah penelitian

penyakit Difteri ini disempurnakan dengan menggunakan penalaran berbasis kasus atau *Cased Based Reasoning* (CBR) agar hasil diagnosis lebih akurat. Output dari penelitian ini yaitu berupa hasil diagnosa penyakit Difteri berdasarkan gejala-gejala yang dialami dengan hasil akurasi pengujiannya sebesar 95,17% (Fatoni,2017). Dan pada penelitian dengan judul Penerapan Algoritma Klasifikasi *Nearest Neighbor* (K-NN) Untuk Mendeteksi Penyakit Jantung disebutkan dalam hasil penelitiannya yaitu pada penelitian ini digunakan algoritma K-NN dengan  $k = 9$  pada 100 data pasien penyakit jantung. Hasil penelitian diperoleh nilai akurasi sebesar 70% serta nilai *AUC* sebesar 0.875 yang masuk ke dalam klasifikasi baik, sehingga algoritma K-NN dapat digunakan dan diterapkan untuk mendeteksi penyakit jantung. Dari rangkaian fakta terhadap metode *K Nearest Neighbor* dan penyakit gagal jantung di atas maka penulis akan melakukan penelitian dengan judul Analisis Metode *K Nearest Neighbor* Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung.

## 1.2. Rumusan Masalah

Dari rangkaian pada latar belakang yang telah dipaparkan, penulis melakukan penelitian yang menitik beratkan pada beberapa permasalahan yaitu:

1. Berapakah nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung?
2. Berapakah nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung setelah melalui proses *SMOTE*?

## 1.3. Tujuan

Dari rangkaian titik permasalahan yang dijelaskan pada rumusan masalah berikut tujuan yang ingin diperoleh dalam penelitian ini yaitu:

1. Mengetahui nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung.
2. Mengetahui nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung melalui proses *SMOTE*.

#### 1.4. Manfaat

Hasil dalam penelitian ini diharapkan bermanfaat terhadap penulis, pembaca dan masyarakat luas. Berikut manfaat yang diharapkan dalam penelitian ini:

1. Pengembangan *Data Mining* dalam dunia kesehatan.
2. Menjadi bahan pertimbangan kesehatan dalam perawatan pasien gagal jantung.

#### 1.5. Batasan Masalah

Dalam pokok masalah yang telah dijabarkan pada rumusan permasalahan, penulis membatasi penelitian ini agar tetap dalam topik permasalahan yang telah ditentukan, berikut batasan-batasan masalah yang diterapkan:

1. Data yang dipakai dalam penelitian ini bersumber dari UCI *Machine Learning* pada laman <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records#>.
2. Jumlah data yang digunakan pada penelitian ini sebanyak 299 data pasien.
3. *Vector* pengukuran jarak yang digunakan dalam metode *K Nearest Neighbor* adalah *Euclidean Distance*.
4. Skenario uji yang digunakan dalam penelitian ini yaitu *Cross Fold Validation* dengan nilai  $k = 2, 5, 7$  dan  $10$ .
5. Nilai *K Nearest Neighbor* yang dipakai adalah  $3, 5, 7$  dan  $9$ .