BAB I

PENDAHULUAN

1.1 Latar Belakang

Gagal Jantung memiliki arti sebagai suatu keadaan dimana jantung tidak mampu memompa darah yang cukup pada seluruh tubuh dengan merasakan ciri-ciri seperti sesak nafas ketika melakukan kegiatan atau ketika tidur dengan posisi terlentang tanpa menggunakan bantal, serta membengkak pada bagian tungkai bawah (PERKI, 2015). Gagal jantung berkontribusi terhadap 287.000 kematian per tahun. Sekitar setengah dari orang yang mengalami gagal jantung meninggal dalam waktu lima tahun setelah mendapat diagnosa. (Maulida, 2018).

Data dari WHO pada tahun 2016 menunjukkan bahwa tahun 2015 terdapat 23 juta atau sebanyak 54% dari total kematian disebabkan oleh penyakit gagal jantung. hasil dari penelitian yang dilakukan di Amerika Serikat menunjukkan bahwa resiko berkembangnya penyakit gagal jantung adalah 20% untuk usia lebih dari atau sama dengan 40 tahun dengan jumlah kasus lebih dari 650.000 (Rispawati, B.H, 2019). Estimasi tingkat penderita penyakit gagal jantung tertinggi berada di Provinsi Jawa Barat sebanyak 96.487 orang (0,3%), sementara tingkat penderita paling rendah berada di Provinsi Kepulauan Bangka Belitung, yaitu sebanyak 945 orang (0,1%) (Kemenkes RI, 2014). Jumlah kematian pasien penderita gagal jantung yang telah mendapat perawatan berkisar di angka 2%-17% (Maggioni et al., 2013). Meskipun gagal jantung tidak dapat disembuhkan, banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup. Berbagai cara pencegahan yang dapat dilakukan ialah dengan memberikan pemahan dan kesadaran terhadap masyarakat tentang cara mengenali gejalanya dan memberikan pemahanan secara sederhana agar masyarakat dapat mengetahuinya. Selain itu diperlukan sikap profesional dalam penanganan kesehatan agar dapat melakukan tindakan dengan cepat dan tepat (Ponikowski et al., 2014). Seiring perkembangan

zaman dalam dunia kesehatan terjadi pula perkembangan yang cukup pesat. Salah satunya penggunaan *Machine Learning* dan Data Mining dalam dunia kesehatan untuk mengklasifikasi sebuah data yang berjumlah besar.

Data mining merupakan suatu proses pengumpulan informasi penting dari sebuah data yang jumlahnya cukup besar yang bertujuan menemukan nilai tambah berupa pengetahuan yang sejauh ini belum dapat ditemukan secara manual dari suatu kumpulan data (Bustami, 2014). Salah satu teknik yang dapat diterapkan ialah teknik klasifikasi. Klasifikasi mirip dengan estimasi, hanya saja variabel target bersifat kategorik dari pada numerik. Teknik klasifikasi mempunyai banyak algoritma yang dapat digunakan dalam menyelesaikan permasalahan tergantung kesesuaian data yang telah di teliti (Larose & Larose, 2014). Adapun beberapa algoritma yang dapat digunakan dalam melakukan sebuah klasifikasi salah satunya adalah Algoritma Gaussian Naive Bayes dimana algoritma tersebut hanya membutuhkan data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi serta algoritma tersebut dapat diterapkan pada sebuah data yang berlanjut dan bersifat numerik.

Pada penelitian dengan judul Analisis Metode *K-Nearest Neighbors* terhadap Klasifikasi Data Pasien Penderita Gagal Jantung menggunakan algoritma *KNN* dan teknik *SMOTE* menghasilkan akurasi sebesar 63,33% (Hidayah Nur Umi & dkk, 2021). Pada penelitian dengan judul Klasifikasi Penderita penyakit Parkinson menggunakan Metode *Gaussian Naive Bayes* dengan menggunakan data *continue* yang didapat dari UCI hasil akurasi sebesar 64,91% dan presisi sebesar 72,72% (Purwatiningsih & Nilogiri, 2020). Pada penelitian dengan judul *Cancer Classification Using Gaussian Naive Bayes Algorithm* dengan data *continue* yang didapat dari UCI menghasilkan akurasi sebesar 98% pada kanker payudara dan 90% pada kanker paru-paru (Kamel, dkk, 2019). Pada penelitian dengan judul Perbandingan Kinerja Algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbors(KNN)* Untuk Mengklasifikasi Penyakit Hepatitis C Virus(HCV) berupa data *continue* menghasilkan nilai akurasi 90,98%, presisi 69,91%, *Recall* 61,57%, dan *f1*

score 63,24% pada algoritma *Gaussian Naive Bayes* dan akurasi sebesar 91,80%, presisi 68,96%, recall 51,85%, dan f1 *score* sebesar 54,3% pada algoritma *KNN* (Raharja, dkk 2021).

Berdasarkan penelitian yang dilakukan sebelumnya yakni pada penyakit parnkinson dan kanker dengan menggunakan metode *Gaussian Naive Bayes* menghasilkan akurasi yang cukup tinggi. Oleh karena itu dari serangkaian fakta tentang algoritma *Gausssian Naive Bayes* dan kemiripan data pada penelitian sebelumnya, maka pada penelitian ini penulis tertarik untuk meniliti dengan judul Analisis Algoritma *Gaussian Naive Bayes* Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung.

1.2 Rumusan Masalah

Sesuai dengan latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah yang dapat diambil dari penelitian ini yaitu, berapakah nilai akurasi, presisi, dan recall yang diperoleh metode *Gaussian Naive Bayes* dalam mengklasifikasi data pasien penderita penyakit gagal jantung?

1.3 Tujuan Penelitian

Sebagaimana telah dijelaskan pada rumusan masalah yang ada, maka tujuan yang ingin dicapai dalam penelitian ini yaitu, untuk mengetahui nilai akurasi, presisi, dan recall yang diperoleh metode *Gaussian Naive Bayes* dalam mengklasifikasi data pasien penderita penyakit gagal jantung.

1.4 Manfaat Penelitian

1. Sebagai alternatif terhadap klasifikasi data pasien penderita gagal jantung dengan menggunakan algoritma *Gaussian Naive Bayes*.
2. Dapat digunakan sebagai bahan referensi untuk melakukan penelitian selanjutnya.

1.5 Batasan Penelitian

1. Pada penelitian ini data diperoleh melalui situs media sosial yaitu *Kaggle* yang terdiri dari data pasien penderita gagal jantung pada tahun 2020 yang dapat diakses melalui link berikut ini: https://www.kaggle.com/jackleenrasmybareh/heart-failure
2. Jumlah data yang digunakan sebanyak 1000 record data pasien penderita gagal jantung. Variabel yang digunakan terdiri dari 29 variabel yaitu los, *age, gender, cencer, cabg, crt, defib, dementia, diabetes, hypertensi, ihd, mental healty, arrhythmias, copd, obesity, pvd, renal disease, vulvular disease, metastatic cencer, pacemaker, pneumonia, prior_appts_attended, prior_dnas, pci, stroke, senile, quintile, ethnichgroup, fu_time dan death eventh* yang terdapaat 2 parameter sebagai kelas *ouput*.
3. Hasil akhir pengujian berupa informasi nilai akurasi, presisi, dan recall.
4. Bahasa pemrograman yang digunakan dalam penelitian tersebut adalah *Python*.
5. *Tools* yang digunakan adalah *Jupyter Notebook*.