

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Salah satu penyakit yang serius dan tidak menular adalah penyakit jantung, penyakit jantung merupakan penyakit dimana kondisi pembuluh darah utama yang mentransfer darah ke jantung mengalami kerusakan dan tidak dapat bekerja dengan baik. Tumpukan kolesterol pada pembuluh darah serta proses peradangan diduga menjadi salah satu faktor penyebab penyakit ini. Ketika terjadi penumpukan kolesterol (plak), pembuluh darah akan menyempit sehingga aliran darah dan transfer oksigen menuju jantung akan terhambat sehingga mempengaruhi kestabilan proses bekerja jantung [1].

Federasi Jantung Dunia memperkirakan angka kematian akibat penyakit gagal jantung di Asia Tenggara mencapai 1,8 juta kasus pada tahun 2014. Di Indonesia sendiri pada tahun 2013 tercatat ada setidaknya 883.447 orang yang terdiagnosis penyakit gagal jantung dengan mayoritas penderita berusia 55-64 tahun. Angka kematian akibat penyakit jantung pun menjadi cukup tinggi, yakni sekitar 45% dari seluruh angka kematian di Indonesia [2].

Salah satu bagian penting dari pengobatan atau tindakan medis adalah pengambilan keputusan dan proses klasifikasi atau prediksi pada suatu hal yang menjadi fokus seperti pendeteksi penyakit, namun klasifikasi medis atau prediksi biasanya merupakan proses yang sangat kompleks dan sulit dilakukan jika tidak dapat mengetahui metode yang tepat dan terbaik dalam memberikan solusinya [3].

Masalah yang dihadapi oleh pihak tenaga kesehatan adalah mendiagnosa pasien dengan benar, prediksi yang buruk dapat menyebabkan konsekuensi yang mendatangkan permasalahan yang kemudian tidak dapat diterima. Untuk menjawab masalah tersebut beberapa penelitian telah dilakukan dalam bidang kesehatan untuk mendapatkan prediksi penyakit dengan lebih akurat, namun belum diketahui metode apa yang paling akurat dalam memprediksi penyakit pasien. Berbagai macam teknik analisa yang secara konvensional dan manual yang selama ini digunakan tidak lagi begitu efektif untuk hal mendiagnosa suatu penyakit. Oleh karena itu, saat ini merupakan waktu yang tepat untuk mengembangkan sistem pengetahuan berbasis komputer yang modern, efektif dan efisien dalam mendiagnosa masalah penyakit gagal jantung [4].

Pada masa sekarang ini kebutuhan terhadap analisis data sangat penting dan dibutuhkan. Kemajuan perkembangan data yang semakin pesat mendorong untuk memanfaatkan data dalam pencarian informasi maupun pengetahuan. Salah satu cara pendekatan yang bisa dilakukan untuk menganalisis sekumpulan data yaitu dengan cara mengklasifikasikan data tersebut. Klasifikasi merupakan salah satu metode data mining yang merupakan teknik dari *Machine Learning*, metode ini digunakan untuk membangun suatu model dari sampel data yang belum terklasifikasi untuk digunakan serta mengklasifikasi sampel data baru ke dalam kelas-kelas yang sama. Adapun beberapa algoritma klasifikasi yang sering digunakan diantaranya adalah *Support Vector Machine*, *K-Nearest Neighbor*, *Neural Network*, *decision Tree* atau *C4.5*, *Naive Bayes* dan lain-lain. Tujuan utama data mining adalah untuk menemukan atau menggali pengetahuan dari data maupun

informasi yang kita miliki. Data mining memiliki beberapa teknik dalam menghasilkan informasi penting dari sekumpulan data, salah satu di antaranya adalah *C4.5* dan *Naïve Bayes*. Algoritma *C4.5* yang merupakan algoritma untuk membentuk pohon keputusan (*decision Tree*), pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Dan Algoritma *Naïve Bayes* yang merupakan teknik prediksi berbasis probabilistik sederhana berdasar pada penerapan teorema atau aturan Bayes dengan asumsi independensi yang kuat pada fitur.

Industri kesehatan memiliki data kesehatan yang cukup besar, namun sebagian besar data tersebut tidak diolah agar mengetahui informasi tersembunyi untuk dijadikan pengambilan keputusan yang efektif oleh para praktisi kesehatan. Pengambilan keputusan data dan informasi yang akurat akan menghasilkan prediksi penyakit menjadi tepat sasaran. Penyakit jantung di Indonesia merupakan penyakit nomor satu dengan angka kematian yang cukup tinggi [5].

Pada tahun 2016 data dari *World Health Organization* (WHO) tercatat sebanyak 17,5 juta orang di dunia meninggal akibat penyakit jantung. Berdasarkan data Riset Kesehatan Dasar (*Riskesdas*) di Indonesia pada tahun 2018 jumlah keseluruhan penyakit jantung sebesar 1,5% atau sekitar 1.017.290 penduduk. Penyebab utama gagal jantung adalah penyakit kardiovaskuler seperti hipertensi, kardiomiopati dan penyakit jantung koroner, yang dapat menyebabkan disfungsi ventrikel kiri berupa penurunan kemampuan kontraksi, relaksasi atau keduanya yang berdampak pada penurunan curah jantung [6].

Menurut Abdul dan M.Rochcham [5], dalam memprediksi penyakit jantung dengan algoritma *C4.5* berdasarkan 867 kasus dan 14 atribut yang diperiksa

sebanyak 364 pasien terdeteksi sakit, sehingga 503 pasien terdeteksi sehat. Model yang dihasilkan diuji untuk mendapatkan nilai akurasi, dan AUC dari setiap algoritma sehingga nilai akurasi nya adalah 86,59 % dengan nilai AUC adalah 0.957, dan masuk kategori kelompok klasifikasi yang sangat baik, karena nilai AUC antara 0.90 sampai 1.00.

Penelitian yang dilakukan oleh Agil dan Agung [7], dalam melakukan perbandingan algoritma *Naïve Bayes* dengan algoritma *K-Nearest Neighbor* prediksi penyakit jantung dengan berdasarkan 300 data yang diperiksa dan sebanyak 163 pasien terdeteksi sehat dan 137 terdeteksi sakit jantung. Algoritma *Naïve Bayes* lebih akurat dan lebih baik dalam klasifikasi penyakit jantung dibandingkan algoritma *K-Nearest Neighbor* dengan nilai persentase akurasi yaitu 90.00%, dan nilai presisi sebesar 86,67%.

Oleh karena itu, penelitian ini dilakukan untuk membantu menyelesaikan permasalahan tersebut dengan data mining klasifikasi untuk prediksi penyakit jantung, diperlukan suatu metode atau teknik yang dapat mengolah data yang sudah ada. Penggunaan data mining Algoritma *C4.5* dan *Naive Bayes* dalam Klasifikasi Penderita Gagal Jantung sebagai pilihan untuk prediksi penyakit gagal jantung dapat menjadi alternatif pilihan yang tepat, tetapi sampai saat ini belum diketahui algoritma klasifikasi yang paling akurat. Dalam melakukan penelitian ini penulis memilih algoritma *C4.5* dan *Naïve Bayes* sebagai perbandingan tingkat akurasinya, Algoritma *C4.5* membentuk pohon keputusan yang merupakan metode klasifikasi dan prediksi yang terkenal. Algoritma *Naïve Bayes* yang merupakan prediksi berbasis probabilistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Maka

penulis tertarik untuk mengangkat penelitian dengan judul “**KOMPARASI DALAM PREDIKSI GAGAL JANTUNG DENGAN MENGGUNAKAN METODE C4.5 DAN NAÏVE BAYES**” Untuk mengetahui tingkat akurasi dari masing-masing metode.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah di uraikan di atas, maka yang menjadi rumusan masalah dalam penelitian ini adalah bagaimana komparasi dalam memprediksi serta mengukur tingkat akurasi penyakit gagal jantung menggunakan metode *C4.5* dan *Naïve Bayes*.

1.3 BATASAN MASALAH

Agar tidak menyimpang dalam penulisan tugas akhir ini dari permasalahan dan lebih terarah, maka perlu adanya batasan masalah pada ruang lingkup bahasan antara lain sebagai berikut:

1. Penelitian ini menggunakan algoritma klasifikasi *C4.5* dan *Naïve Bayes* dalam memprediksi penyakit gagal jantung.
2. Aplikasi yang digunakan untuk memproses data adalah *tools Rapid Miner*.
3. *Dataset* prediksi gagal jantung yang akan digunakan dalam penelitian ini diambil pada laman:

<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

4. Data yang diambil berjumlah 918 data dan 12 atribut. Atribut yang digunakan yaitu : *Age, Gender, Chest pain type, Resting BP, Cholesterol, Fasting BS, Resting ECG, Max HR, Exercise Angina, Old peak, ST Slope, Heart Disease.*

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang terjadi maka penelitian ini bertujuan untuk:

1. Melakukan Persiapan Data.
2. Melakukan *Exploratory Data Analysis* (EDA) menggunakan *tools Rapid Miner.*
3. Mengetahui bagaimana algoritma klasifikasi *C4.5* dan *Naïve Bayes* memprediksi penyakit gagal jantung.
4. Mengetahui perbandingan antara algoritma klasifikasi *C4.5* dan *Naïve Bayes* dalam mengukur tingkat akurasi penyakit gagal jantung.

1.4.2 Manfaat Penelitian

Adapun beberapa manfaat yang terdapat dalam penelitian ini adalah:

1. Penulis dapat menambah pengetahuan terkait bagaimana cara mencari informasi yang penting dalam suatu data dengan menggunakan metode Algoritma Klasifikasi *C4.5* dan *Naïve Bayes*
2. Dapat menjadi informasi bagi pembaca untuk mengetahui resiko pasien terkena penyakit gagal jantung dengan tingkat akurasi yang tinggi.

3. Dapat membantu pihak tenaga kesehatan dalam memprediksi pasien yang berkemungkinan terkena penyakit gagal jantung.
4. Menjadi bahan pertimbangan dalam perawatan pasien gagal jantung.
5. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya.

1.5 SISTEMATIKA PENULISAN

Untuk memberikan gambaran umum mengenai keseluruhan penulisan dapat dilihat melalui sistematika penulisan yang meliputi beberapa bab yaitu:

BAB I : PENDAHULUAN

Bab ini berisi pendahuluan yang membahas latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan yang digunakan dalam penulisan laporan tugas akhir ini.

BAB II : LANDASAN TEORI

Bab ini memuat landasan teori yang mendasari pembahasan laporan yang secara khusus berisi definisi-definisi yang melandasi penelitian yang dilakukan, yang didapat dengan melakukan studi pustaka sebagai dasar dalam melakukan penerapan data mining. Teori-teori yang dipakai antara lain mengenai algoritma *C4.5* dan *Naïve Bayes* serta *tools Rapid Miner*.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang penjelasan dari metode-metode penelitian yang digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian.

BAB IV : ANALISIS

Pada bab ini menjelaskan tentang gambaran umum pemilihan dataset pada kaggle.com. Kemudian menganalisis dataset dengan melakukan persiapan data, lalu penulis akan melakukan *Exploratory Data Analysis* untuk mengetahui prediksi tingkat akurasi data menggunakan metode klasifikasi algoritma *C4.5* dan *Naive Bayes* serta menggunakan *tools Rapid Miner*. Hasil dari analisis yang telah dilakukan diinterpretasikan menjadi sebuah informasi mengenai status prediksi gagal jantung.

BAB V : KESIMPULAN

Pada bab ini berisikan kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.