

BAB I PENDAHULUAN

1.1. Latar Belakang

Penyakit jantung koroner biasa juga disebut dengan kardiovaskular, yaitu menyempit atau tersumbat pembuluh darah yang mengarah pada serangan jantung, nyeri dada atau stroke (Babič, et al., 2017). Penyakit jantung merupakan penyakit dengan tingkat kematian yang tinggi dengan setiap tahun lebih dari 12 juta kematian terjadi diseluruh dunia dikarenakan penyakit jantung (Umadevi & Snehapriya, 2017). Dengan demikian diagnosis dini sangat penting dan merupakan bidang penting dari penelitian medis, (Roohallah Alizadehsani, et al., 2012). Diagnosis yang benar dari penyakit jantung adalah tugas yang cukup menantang karena saling ketergantungan yang kompleks dari berbagai faktor (Chaki, et al., 2015). Permasalahan yang sering kali dihadapi adalah waktu di dalam proses serta kurangnya keakurasian pada proses klasifikasi (Kaur & Singh, 2014).

Fitur atribut merupakan suatu hal yang penting di dalam keakuratan hasil, sehingga perlu kiranya diketahui atribut utama pada sebuah penyakit (Alizadehsani, et al., 2013). Sering didapatkan hasil yang berbeda ketika mendiagnosis penyakit jantung, sehingga dipilih fitur atribut penting yang didasarkan pada efisiensi diagnosis (El-Bialy, et al., 2015). Keputusan diagnosis penyakit yang berbeda merupakan salah satu tantangan yang paling penting untuk analisis data penyakit jantung (Babič, et al., 2017). Umumnya keputusan diagnosis didasarkan kepada pemeriksaan kepada pasien dan juga pengalaman dokter. Pada pemeriksaan yang dilakukan tidak semua kondisi atribut terpenuhi, tetapi tetap harus dilakukan proses diagnosis.

Kesalahan pada proses diagnosis akan mempengaruhi hasil pada pasien ataupun dokter, jadi diagnosis merupakan tugas kompleks yang membutuhkan keahlian tinggi dan pengalaman. Sistem komputer dapat membantu para dokter sebagai alat memprediksi dan diagnosis penyakit jantung, penelitian medis selalu

berurusan dengan data yang besar, penanganan data yang besar dengan tidak benar dapat mempengaruhi keakuratan terhadap hasil. Dataset yang digunakan merupakan kumpulan dari pengalaman dokter dan juga dari beberapa rumah sakit serta perguruan tinggi (Umadevi & Snehapriya, 2017). Reduksi atribut merupakan pemilihan fitur digunakan untuk mengurangi dimensi data dengan menghapus atribut yang berlebihan dan tidak relevan dalam satu set data. Proses klasifikasi dipengaruhi oleh atribut, fitur redundansi yang tidak relevan memiliki efek negatif pada keakuratan klasifikasi dan setelah dilakukan redundansi pada atribut, tidak mengurangi tingkat akurasi terhadap dataset. (Janabi & Kadhim, 2018). Pengurangan atribut telah menjadi langkah penting dalam pengenalan pola dan tugas Machine Learning (Ding, et al., 2013). Hal ini menginspirasi penulis sehingga menambahkan proses reduksi atribut di dalam pengambilan keputusan klasifikasi pada dataset penyakit jantung.

Principal component analysis (PCA) adalah teknik untuk mengurangi dimensi atribut dari dataset, meningkatkan interpretabilitas tetapi pada saat yang sama meminimalkan hilangnya informasi, banyak penelitian dan buku yang telah ditulis dan juga bahkan ada buku tentang varian PCA untuk tipe data tertentu (Jolliffe & Cadima, 2016). PCA dapat digabungkan dengan metode lain yang akurasi dikenal keakurasiannya cukup tinggi, dimana PCA berfungsi untuk melakukan reduksi dimensionalitas dan dengan pemilihan fitur pada PCA yang hanya memilih nilai eigen yang paling besar maka ketika digabungkan dengan metode lain akan menaikkan tingkat akurasi klasifikasi dan terutama sekali mempercepat hasil pengenalannya (Syakhala, et al., 2015).

Algoritma C5.0 memiliki kemampuan untuk mengklasifikasikan sebagai pohon keputusan atau seperangkat aturan. Kekuatan algoritma ini dalam menangani nilai, serta lebih sedikit waktu diperlukan untuk mempelajarinya (Abdar, et al., 2015). C5.0 adalah classifier yang mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan classifier lain. Untuk menghasilkan pohon keputusan penggunaan memori minimum dan juga meningkatkan akurasi (Pandya & Pandya, 2015). C5.0 merupakan terusan dari algoritma C4.5, dimana C5.0 memperbaiki kelemahan dari C4.5 dari segi hasil

klasifikasi, kesalahan klasifikasi, prediksi akurasi, waktu, tingkat kesalahan, dan pemakaian memori (Revathy & Lawrance, 2017)

Berdasarkan dari penjelasan di atas maka penulis bermaksud untuk melakukan penelitian sebagai berikut, melakukan reduksi atribut pada dataset penyakit jantung. Setelah dilakukan reduksi atribut kemudian dilakukan klasifikasi kembali terhadap atribut-atribut yang digunakan selanjutnya dan apakah terdapat perubahan pada tingkat akurasi klasifikasi dari proses yang setelah terdapat reduksi atribut dengan proses klasifikasi sebelum diadakannya reduksi atribut. Pada proses reduksi atribut dan juga klasifikasi penelitian ini menggunakan dataset yang dibuat oleh para ahli dibidang penyakit jantung. Tetapi sebelum dilakukannya proses reduksi atribut dan juga klasifikasi, dilakukan terlebih dahulu preprosesing pada dataset. Dimana tujuan dilakukan preprosesing untuk menyamakan karakteristik dataset terhadap metode yang akan digunakan. Berdasarkan latar belakang masalah yang telah disebutkan di atas, maka penulis membuat judul **“Reduksi Atribut Pada Dataset Penyakit Jantung Dan Klasifikasi Menggunakan Algoritma C5.0”**

1.2. Masalah Penelitian

Berdasarkan latar belakang yang telah dijelaskan di atas, masalah yang ingin dicari solusinya dalam penelitian ini dibagi lagi ke dalam dua bagian, yaitu indentifikasi masalah dan rumusan masalah.

1.2.1. Identifikasi Masalah

Sesuai dengan latar belakang yang ditulis, masalah yang terdapat di dalam penelitian ini, yaitu:

1. Atribut yang terlalu banyak dimiliki oleh dataset untuk proses klasifikasi penyakit jantung.
2. Tidak diketahuinya atribut-atribut utama yang sebenarnya digunakan sebagai proses pengambilan keputusan pada penyakit jantung.

3. Waktu yang dibutuhkan lebih lama jika pada dataset memiliki atribut atau record yang terlalu banyak

1.2.2. Rumusan Masalah

Berdasarkan dengan latar belakang masalah yang sudah dijelaskan di atas, maka yang menjadi rumusan masalah dalam penelitian ini adalah pada proses diagnosis yang dilakukan sering pertanyaan atau atribut tidak terpenuhi secara keseluruhan. Sehingga perlu dilakukan reduksi atribut pada dataset penyakit jantung dengan menggunakan metode Principal Component Analysis (PCA), dimana reduksi atribut berguna untuk mengetahui atribut utama di dalam proses diagnosis penyakit jantung, kemudian algoritma C5.0 untuk mengklasifikasi dan mengetahui tingkat akurasi klasifikasi setelah dilakukan reduksi reduksi dan membandingkan dengan tingkat akurasi sebelum dilakukan reduksi atribut

1.3. Tujuan dan Manfaat Penelitian

Berdasarkan latar belakang yang telah dijelaskan di atas, maka tujuan dan manfaat yang ingin dicapai dalam penelitian ini :

1.3.1. Tujuan Penelitian

Berdasarkan rumusan yang telah diuraikan maka tujuan yang ingin dicapai pada penelitian ini yakni mencari atribut yang paling berpengaruh pada proses diagnosis dan pengambilan keputusan dengan melakukan reduksi atribut dengan menggunakan metode Principal Component Analysis (PCA) untuk mengetahui atribut-atribut utama yang digunakan pada pengambilan keputusan jantung serta mempersingkat proses klasifikasi dan algoritma C5.0.

1.3.2. Manfaat Penelitian

Adapun manfaat yang diharapkan akan didapatkan dari penelitian ini adalah sebagai berikut :

1. Dengan melakukan reduksi pada atribut dataset dapat mempersingkat waktu proses pengklasifikasi pada dataset.
2. Mengetahui atribut-atribut utama pada pengambilan keputusan di dalam proses diagnosis penyakit jantung.
3. Sebagai bahan masukan bagi peneliti lainnya tentang reduksi atribut yang dilakukan pada dataset.

1.4. Batasan Masalah

Berdasarkan perumusan masalah yang telah diuraikan di atas, maka dapat disusun Batasan masalah dari penelitian ini sebagai berikut :

1. Penelitian ini menggunakan dataset penyakit jantung yang berasal dari UCI Machine Learning Repository.
2. Untuk melakukan pengujian pada penelitian ini menggunakan tools IBM SPSS Modeler.

1.5. Metodologi Penelitian

Dalam penyusunan tesis ini penulis melakukan beberapa penerapan metodologi penyelesaian untuk menyelesaikan masalah. Adapun langkah-langkah penyelesaian ini adalah:

1. Studi Literatur

Pada tahap ini dilakukan pengumpulan bahan referensi yang berhubungan dengan penelitian, seperti: penyakit jantung, reduksi atribut, metode Principal Component Analysis (PCA), Algoritma C5.0, dataset penyakit jantung dan bahan referensi lain terkait dengan penelitian mengenai reduksi atribut, klasifikasi dan penyakit jantung.

2. Analisis Masalah

Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada.

3. Preprocessing data

Pada tahapan preprosesing data digunakan untuk melakukan perubahan data / penyesuaian data dengan algoritma ataupun metode yang digunakan sehingga proses dapat dilakukan.

4. Tahap Pengujian

Melakukan pengujian terhadap algoritma C5.0 pada tools IBM SPSS Modeler dalam mengolah dataset penyakit jantung sehingga didapatkan hasil yang sesuai pada tahapan analisa.

5. Evaluasi

Pada tahap ini, dilakukan evaluasi terhadap hasil pengujian yang telah dilakukan untuk mengambil kesimpulan dan saran.

1.6. Sistematika Penulisan

Dalam laporan tesis ini, pembahasan disajikan dalam 5 BAB dengan sistematika pembahasan sebagai berikut. Pada BAB 1, akan membahas mengenai yang melatarbelakangi pemilihan masalah masalah dan tujuannya dilakukan penelitian reduksi atribut penyakit jantung serta klasifikasinya. Pada BAB 2, akan membahas mengenai kajian literatur dan juga tinjauan singkat mengenai permasalahan pada penelitian dan juga metode yang digunakan dalam proses reduksi atribut serta klasifikasi penyakit jantung. Pada BAB 3, membahas tentang identifikasi masalah, langkah-langkah dari metode ataupun algoritma yang digunakan, alat pengujian serta dataset yang akan digunakan untuk melakukan analisa. Pada BAB 4, akan menjelaskan hasil dari masalah penelitian yang telah dijelaskan pada BAB 1. Dan pada BAB 5 akan ditarik kesimpulan dari hasil yang didapat serta saran yang diberikan terhadap penelitian ini selanjutny