

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung adalah sebuah kondisi yang menyebabkan jantung tidak dapat melaksanakan tugasnya dengan baik. Penyakit ini digambarkan sebagai rasa sakit di bagian dada yang berkelanjutan selama setengah jam, menjalar kearah tangan kiri dan rahang disertai sukar bernafas dan perasaan takut yang berlebihan. Hal ini disebabkan matinya sebagian otot jantung yang disebabkan oleh tersumbatnya saluran darah dibagian tersebut. Penyakit yang mengenai jantung biasa disebut sebagai penyakit *kardiovaskular* (Nurchahyo,2008).

Pada saat ini penyakit jantung merupakan penyebab kematian nomor satu di dunia. Pada tahun 2005 sedikitnya 17,5 juta atau setara dengan 30,0 % kematian diseluruh dunia disebabkan oleh penyakit jantung. Berdasarkan penelitian badan kesehatan dunia (WHO), di kalangan pria 30% kematian penyakit *kardiovaskular* disebabkan oleh penyakit jantung sedangkan wanita 25%. Di Indonesia, penyakit jantung juga cenderung meningkat sebagai penyebab kematian. Data survei kesehatan rumah tangga (SKRT) tahun 1996, menunjukkan bahwa proporsi penyakit ini meningkat dari tahun ke tahun sebagai penyebab kematian (Darmojo, 1999).

Penyakit jantung dapat memberikan perbedaan pengaruh pada pasien yang berbeda untuk tingkatan penyakit yang berbeda pula. Sebuah gejala juga bisa mengindikasikan beberapa penyakit jantung yang berbeda. Salah satu penelitian mengenai tingkatan penyakit jantung adalah data Haberman's Survival. Data ini disumbangkan oleh David W. Aha. Dataset ini berisi data yang berasal dari penelitian mengenai klasifikasi penyakit jantung di V.A. Medical Center, Longbeach dan Cleveland Clinic Foundation. Data Haberman's survival terdiri dari 303 kasus pasien penyakit jantung. Data ini telah digunakan dalam beberapa penelitian sebelumnya. Salah satu penelitian yang menggunakan data ini adalah menggunakan metode *fuzzy expert system* yang memperoleh tingkat akurasi sebesar 94% (Adeli, 2010). Selain metode yang disebutkan di atas, metode lain yang dapat digunakan untuk mengklasifikasi

penyakit jantung pada data Haberman's survival adalah dengan menggunakan penggabungan representasi *fuzzy* dengan *decision tree*.

Klasifikasi penyakit jantung sebenarnya dapat dilakukan dengan menggunakan logika tegas. Akan tetapi hal ini sangat kaku, karena dengan adanya perubahan yang kecil saja terhadap nilai dapat mengakibatkan perbedaan kategori. Logika *fuzzy* digunakan untuk mengantisipasi hal tersebut, karena dapat memberikan toleransi terhadap nilai, sehingga dengan adanya perubahan sedikit pada nilai tidak akan memberikan perbedaan yang signifikan. Dengan memanfaatkan kelebihan logika *fuzzy* dalam toleransi terhadap hal ambigu, diharapkan dapat menjadi pendukung keputusan dalam mengklasifikasikan data penyakit jantung berdasarkan pada faktor umur, tekanan darah, kolesterol, denyut jantung dan *oldpeak*.

Metode *decision tree* sangat terkenal daripada metode klasifikasi yang lainnya, karena metode ini tidak membutuhkan pengetahuan yang lebih atau pengaturan parameter (Han dan Khamber, 2001). Akan tetapi kebanyakan metode *decision tree* menghasilkan keputusan dari data kategorial untuk proses pelatihan dan klasifikasi data-data baru. Penggunaan data numerik pada *decision tree* memungkinkan dengan membentuk partisi pada data tersebut.

Penggabungan metode *decision tree* dengan *fuzzy* memungkinkan untuk menggunakan nilai-nilai numerik yang dihubungkan dengan atribut kuantitatif yang masing-masing memiliki nilai derajat keanggotaan. Menurut Liang, 2005, menggunakan teknik *fuzzy* dalam *decision tree* dapat meningkatkan kemampuan atribut-atribut kuantitatif dalam melakukan penggolongan pada saat pelatihan. Selain itu, proses pengujian menggunakan metode *fuzzy* inferensi Mamdani adalah untuk mendapatkan output yang baik berdasarkan nilai derajat keanggotaan masing-masing atribut.

Pada skripsi ini, algoritma yang digunakan dalam *fuzzy decision tree* adalah C4.5. Algoritma C4.5 merupakan suksesor dari ID3 menggunakan *gain ratio* untuk memperbaiki *information gain*. Pendekatan ini menerapkan normalisasi pada *information gain* dengan menggunakan *split info* (Khairina, 2007). Metode *fuzzy decision tree* menggunakan algoritma C4.5 ini telah dilakukan pada berbagai riset. Rumusan metode ini dituliskan pada riset yang

dilakukan Ahmad Saikhu tahun 2011 pada data diabetes Indian pima dan menghasilkan tingkat akurasi 78,91%.

Berdasarkan latar belakang yang telah dikemukakan, maka judul yang diambil dalam skripsi ini adalah “**Penerapan *Fuzzy Decision Tree* dengan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung**”.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah, maka dalam skripsi ini dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana menerapkan metode *Fuzzy Decision Tree* menggunakan algoritma C4.5 untuk mengklasifikasi penyakit jantung pada data Haberman's *Survival*.
2. Bagaimana tingkat akurasi metode *Fuzzy Decision Tree* menggunakan algoritma C4.5 untuk klasifikasi penyakit jantung pada data Haberman's *Survival*.

1.3 Batasan Masalah

Untuk membatasi masalah yang dikaji, maka penulis membatasi dan memfokuskan permasalahan sebagai berikut:

1. Terdapat lima parameter dalam melakukan klasifikasi penyakit jantung, yaitu: umur, tekanan darah (*tresbps*), kolesterol (*chol*), denyut jantung (*thalach*) dan *oldpeak*.
2. Hasil prediksi atau keputusan yang dihasilkan terdiri dari lima kelas yaitu penderita sehat (*healthy*), penderita dengan penyakit jantung kategori ringan (*sick1*), sedang (*sick2*), berat (*sick3*) dan sangat berat (*sick4*).
3. Data yang dijadikan pelatihan dan pengujian pada skripsi ini adalah dataset Haberman's survival yang diambil UCI Machine Learning Repository diambil dari www.archive.ics.uci.edu/ml/datasets/Heart+Disease.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menerapkan metode *Fuzzy Decision Tree* menggunakan algoritma C4.5 untuk mengklasifikasi penyakit jantung pada data Haberman's *Survival*.

2. Menghitung hasil tingkat akurasi metode *Fuzzy Decision Tree* menggunakan algoritma C4.5 untuk klasifikasi penyakit jantung pada data Haberman's *Survival*.

1.5 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah dapat mengukur keakuratan hasil klasifikasi sehingga dapat diketahui kelayakan penerapan metode ini pada data Haberman's *Survival*.

1.6 Sistematika Penulisan

Skripsi ini disusun berdasarkan sistematika penulisan sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini menguraikan teori-teori yang berhubungan dengan penyakit jantung, data mining, logika *fuzzy*, *decision tree*, algoritma C4.5 dan metode *fuzzy decision tree*.

3. BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini membahas tentang data penelitian, spesifikasi perangkat lunak, perancangan sistem, perancangan basis data dan perancangan antarmuka.

4. BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini berisi implementasi sistem dan pengujian perangkat lunak yang dibangun.

5. BAB V PENUTUP

Bab ini berisi kesimpulan dari seluruh rangkaian penelitian serta saran untuk kemungkinan pengembangan lebih lanjut.