

1. Pendahuluan

1.1 Latar belakang

Hampir di semua bidang kehidupan nyata berhadapan dengan pertumbuhan data dari suatu pengukuran sehingga apabila data tersebut tidak diolah maka informasi yang terdapat didalamnya akan terbuang sia-sia. Misalnya data nilai mahasiswa yang hanya menampilkan nilai mahasiswa bisa saja diolah untuk mendapatkan informasi mengenai bidang peminatan berdasarkan nilai rata-rata terbaik. Oleh karena itu, diperlukan suatu teknik pengolahan data menjadi suatu informasi yang nantinya dapat berfungsi sebagai pendukung suatu keputusan.

Salah satu teknik yang sering digunakan adalah klasifikasi. Klasifikasi adalah teknik pembelajaran terhadap suatu data yang dilakukan untuk memprediksi kelas atau properti dari setiap *instance* kelas [CHA00]. Salah satu teknik klasifikasi yang menjadi fokus utama pada Tugas Akhir ini adalah tentang induksi pohon keputusan (*decision tree*). Suatu pohon keputusan sangat mudah diinterpretasikan, efisien dan cepat dalam mengklasifikasikan suatu *record* yang belum diketahui kelasnya. Namun, pohon keputusan bersifat rentan terhadap kesalahan dalam klasifikasi masalah dengan berbagai kelas dan jumlah data latih yang digunakan relatif kecil.

Tugas Akhir yang berjudul “Analisis dan Implementasi Klasifikasi Data Menggunakan *Soft Decision Tree* - ID3 (SDT - ID3)” ini merupakan teknik induksi pohon keputusan dengan menggunakan teori *fuzzy set*. Induksi pohon keputusan seperti ID3, C4.5 dan CART menghasilkan struktur pohon keputusan yang mempartisi secara rekursif suatu ruang atribut menjadi beberapa bagian dari ruang keputusan sehingga tidak ada yang *overlap* [JAN96]. Apabila atribut yang digunakan bertipe *continuous* (numerik) maka harus dilakukan diskritisasi untuk membagi range domain pada atribut menggunakan titik potong. Titik potong yang digunakan akan membedakan nilai suatu domain dengan batasan yang jelas sehingga bisa saja terjadi *missclassification*. Namun, pada contoh kasus di dunia nyata, terdapat masalah dengan informasi yang kurang presisi, tidak lengkap dan memiliki kebenaran parsial maka untuk menangani masalah tersebut digunakan pendekatan dengan menggunakan *fuzzy set* dimana proses *fuzzy reasoning* memungkinkan dua atau lebih aturan akan divalidasi secara bersamaan dan bertahap sehingga hasil akhirnya akan menjadi hasil dari penggabungan beberapa hasil [NGU07].

Pada Tugas Akhir ini, atribut domain didiskritisasi dalam proses induksi pohon keputusan dengan menggunakan *fuzzy set*. Tahap yang dilakukan selanjutnya adalah penentuan kriteria untuk pemilihan atribut dilanjutkan dengan membangun suatu pohon keputusan sebagai model untuk proses pembelajaran sistem. Implementasi sistem dibangun dengan bahasa pemrograman Java 6.0.

1.2 Perumusan masalah

Dengan mengacu pada latar belakang masalah di atas, maka permasalahan yang dibahas dan diteliti adalah :

1. Bagaimana melakukan klasifikasi dengan menggunakan algoritma *Soft Decision Tree - ID3* (SDT - ID3).
2. Bagaimana analisis perbandingan algoritma *Soft Decision Tree - ID3* dengan algoritma lain, jika dilihat dari perhitungan matriks evaluasi.

Tugas Akhir ini memiliki batasan masalah sebagai berikut :

1. *Dataset* yang digunakan adalah *breast cancer*, *glass*, *wine* yang diperoleh dari *UCI Machine Learning Repository* dan data CDR (*Calling Detail Record*) yang merupakan data untuk klasifikasi daerah pelanggan Telekomunikasi.
2. Tipe *dataset* yang digunakan adalah *record* data.
3. Jenis atribut yang digunakan adalah atribut numerik (*continuous*)
4. Setiap *record* data terklasifikasi ke dalam satu kategori atau kelas.
5. Fungsi keanggotaan yang digunakan adalah trapesium, segitiga dan *gaussian*. Hal ini berdasarkan distribusi data dari keempat *dataset* tersebut.
6. Tidak menangani *preprocessing*.

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan dari Tugas Akhir ini antara lain:

1. Menerapkan algoritma SDT - ID3 sebagai suatu metode dalam mengklasifikasikan data.
2. Menganalisis jumlah variabel linguistik dan bentuk fungsi keanggotaan terhadap hasil klasifikasi. Hal ini bertujuan untuk menghasilkan suatu keputusan yang tepat serta tingkat akurasi tiap klasifikasi yang dihasilkan
3. Mengevaluasi hasil klasifikasi dengan menggunakan matriks evaluasi dan matriks performansi berupa tingkat keakuratan, *precision*, *recall* dan *F-Measure*.

1.4 Metodologi penyelesaian masalah

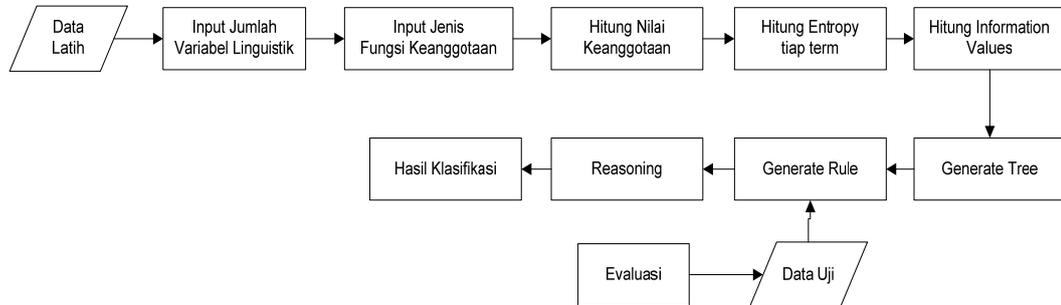
Untuk mencapai tujuan yang dimaksud, maka tahapan yang akan dilakukan dalam pelaksanaan Tugas Akhir ini adalah :

1. Pengumpulan data

Tahap ini bertujuan untuk mencari dan mengumpulkan data yang ingin diklasifikasikan serta jurnal ilmiah atau referensi lain berupa teori-teori dasar tentang *Classification*, *Fuzzy set* dan *Soft Decision Tree - ID3* (SDT - ID3) untuk mendukung pengerjaan Tugas Akhir. Dokumen yang ingin di analisis adalah *dataset breast cancer*, *wine*, *glass* dan CDR (*Calling Detail Record*) yang diperoleh dari *UCI Machine Learning Repository*.

2. Pembangunan model

Tahap ini bertujuan untuk membangun alur model sistem berupa *flowchart* atau *Unified Model Language (UML)*. Alur model pengklasifikasian data menggunakan *Soft Decision Tree - ID3* yaitu :



Gambar 1. 1 Alur Model Klasifikasi Data dengan *Soft Decision Tree - ID3*

3. Implementasi

Pada tahap ini dilakukan implementasi berdasarkan alur model sistem dan melakukan evaluasi kinerja metode yang digunakan sebagai kajian selanjutnya. Dari pemodelan yang telah dibuat kemudian dibuat ke dalam bahasa pemrograman menggunakan Java dengan editor Netbeans 6.0.

4. Analisis hasil

Pada tahap ini melakukan analisis hasil dari implementasi berupa performansi hasil klasifikasi berdasarkan perhitungan nilai akurasi, dan matriks evaluasi (*precision*, *recall* dan *F-Measure*).

5. Pembuatan laporan

Menyusun laporan hasil analisis yang dirangkum ke dalam sebuah buku Laporan Tugas Akhir.