1 Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi yang sangat pesat saat ini membuat jumlah data yang disimpan menjadi sangat besar. Besarnya jumlah ini membuat data yang tersimpan tidak bisa digunakan secara langsung. Diperlukan suatu metode *data mining* tertentu yang digunakan untuk mengolah data yang ada menjadi suatu informasi. *Data mining* sendiri berarti suatu eksplorasi dan analisis baik yang bersifat otomatis maupun semi-otomatis terhadap suatu data baik dalam jumlah yang besar ataupun tidak untuk menemukan pola yang mengandung informasi [3]. Salah satu metode yang dilakukan dalam proses data mining adalah dengan klasifikasi data atau pengelompokan data berdasarkan suatu kelas tertentu yang sudah didefinisikan. Dengan klasifikasi data dimungkinkan dapat dipahami pola-pola yang muncul dari suatu data sehingga dapat digunakan untuk menentukan strategi bisnis ke depannya. Dalam *data mining* sendiri terdapat berbagai macam algoritma atau metode untuk melakukan klasifikasi data. Contoh algoritma yang dapat digunakan dalam klasifikasi data adalah algoritma SLIQ yang merupakan turunan dari metode pohon keputusan dan juga Fuzzy SLIQ Decision Tree yang merupakan turunan dari algoritma SLIQ.

Algoritma SLIQ atau kepanjangan dari Supervised Learning In Quest adalah suatu algoritma klasifikasi yang dikembangkan oleh IBM's Quest pada tahun 1996. Pengembangan algoritma ini pada awalnya bertujuan untuk mengatasi permasalahan keterbatasan memori dataset dalam klasifikasi data dengan jumlah data yang besar pada algoritma-algoritma *decision tree* yang lain pada saat itu. Selain dapat digunakan dalam klasifikasi data yang sangat besar, algoritma SLIQ sendiri juga dapat digunakan untuk melakukan klasifikasi pada data yang terbatas atau data yang sedikit jumlahnya tanpa kehilangan tingkat akurasi[6]. Tetapi algoritma SLIQ memiliki kelemahan dalam klasifikasi pada data-data *continuous*, karena *split point* pada algoritma SLIQ berupa diskret, sehingga klasifikasi yang dihasilkan akan kasar dan kaku, tidak sesuai dengan karakteristik data *continuous*[4].

Algoritma Fuzzy SLIQ Decision Tree (FS-DT) adalah suatu algoritma pengembangan dari algoritma SLIQ. Pada FS-DT, *decision boundary* ditentukan dengan *fuzzy*, hal ini didasari pada susahnya menentukan batasan yang pasti dalam kasus klasifikasi yang sebenarnya. Pemilihan metode *fuzzy* sendiri didasarkan karena karakteristik sistem *fuzzy* yang dapat menangani suatu kasus yang bersifat ketidakpastian. Selain itu penggunaan FS-DT juga dapat mengurangi ukuran dari pohon pencarian yang dilakukan. Hal ini tentu menjadi salah satu parameter yang sangat baik bagi *Decision Tree*, karena semakin dalam dan semakin lebar ukuran pohon yang dihasilkan, maka data yang dihasilkan juga akan semakin tidak komprehensif dan memungkinkan munculnya *error* [4].

1.2 Perumusan Masalah

- Apakah penambahan *fuzzy* dalam proses penentuan *decision tree* pada algoritma SLIQ akan dapat meningkatkan performansi dalam klasifikasi data?
- Bagaimana performansi Algoritma FS-DT dibanding Algoritma SLIQ dilihat dari ukuran pohon, akurasi dan waktu proses?

1.3 Hipotesis

- Penambahan *fuzzy* dalam proses penentuan *decision boundaries* pada algoritma SLIQ akan dapat meningkatkan akurasi klasifikasi yang dihasilkan pada klasifikasi data-data bertipe *continuous*
- Semakin besar nilai training sets yang digunakan dalam proses pembentukan model, maka hasil klasifikasi yang diberikan juga akan semakin besar

1.4 Tujuan

Tujuan yang ingin dicapai dari penulisan tugas akhir ini adalah:

1. Mengimplementasikan algoritma Fuzzy SLIQ Decision Tree ke dalam suatu aplikasi untuk klasifikasi data
2. Membandingkan ukuran *decision tree*, keakuratan hasil klasifikasi dan waktu proses klasifikasi antara algoritma SLIQ dan algoritma FS-DT

1.5 Batasan Masalah

Batasan masalah yang digunakan dalam pelaksanaan tugas akhir ini adalah

1. Perbandingan performansi kedua algoritma didasarkan pada besarnya pohon yang dihasilkan, waktu proses dan akurasi kedua algoritma tersebut
2. Jenis data yang digunakan adalah data numerik.
3. Dataset yang digunakan berupa *record* data (kumpulan data) yang disimpan dalam file bertipe *.txt*.
4. Fungsi keanggotaan *fuzzy* yang digunakan adalah : (B. Chandra and P. Paul Varghese)

$$fuzzy\ value = \begin{cases} \frac{lw}{lp + lw - val}, & val < lp \\ 1.0, & lp \leq val \leq rp \\ \frac{rw}{val - rp + rw}, & val > rp \end{cases}$$

Persamaan 1-1 Fungsi Keanggotaan Fuzzy SLIQ Decision Tree[4]

1.6 Metode Penyelesaian Masalah

Metodologi yang dilakukan dalam pelaksanaan tugas akhir ini adalah sebagai berikut ini

1. Pengumpulan Data

Pengumpulan data dilakukan dengan melakukan studi pustaka baik di internet atau dalam bentuk media cetak. Data yang dikumpulkan adalah yang berkaitan dengan FS-DT, SLIQ, kasus data mining yang akan diujicoba ataupun terhadap *Data Mining* itu sendiri.

2. Pembuatan Model

Model yang dibuat berkaitan dengan skenario pelaksanaan perbandingan dan pembuatan desain sistem yang akan dibangun dengan menggunakan algoritma Fuzzy SLIQ Decision Tree di atas.

3. Implementasi

Implementasi dilakukan dengan membangun sistem berdasarkan algoritma Fuzzy SLIQ Decision Tree yang ada, kemudian mengujicobakan terhadap contoh permasalahan yang ada.

4. Analisa Hasil

Analisis dilakukan terhadap hasil dari pengujian. Analisis ini nantinya akan menghasilkan sebuah kesimpulan perbandingan algoritma yang ada sesuai dengan parameter yang diujikan.

5. Pembuatan Laporan

Laporan dibuat berdasarkan hasil penelitian yang sudah dilakukan sebelumnya.

1.7 Sistematika Penulisan

Penulisan Tugas Akhir ini dibagi ke dalam enam bab sebagai berikut ini :

- Bab I Pendahuluan
Berisi penjelasan mengenai latar belakang pembuatan tugas akhir ini, perumusan masalah yang akan dianalisa dalam tugas akhir ini, batasan dari masalah yang akan dibahas, tujuan pembuatan tugas akhir ini, metodologi penyelesaian masalah dan sistematika penulisan tugas akhir ini.
- Bab II Landasan Teori
Berisi landasan teori mengenai data mining, klasifikasi, algoritma SLIQ dan algoritma Fuzzy SLIQ Decision Tree.
- Bab III Analisis Perancangan Sistem
Berisi perancangan dari perangkat lunak yang akan dibuat meliputi spesifikasi kebutuhan, perancangan diagram perangkat lunak, serta penjelasan mengenai spesifikasi perangkat lunak yang dibuat dalam pelaksanaan tugas akhir ini.
- Bab IV Implementasi dan Pengujian Sistem
Berisi implementasi dari algoritma Fuzzy SLIQ Decision Tree dalam suatu perangkat lunak dan pengujian terhadap hasil klasifikasi yang dilakukan oleh perangkat lunak.
- Bab V Kesimpulan dan Saran
Berisi kesimpulan dari keseluruhan pelaksanaan tugas akhir ini serta saran untuk pengembangan sistem yang lebih baik.