

## ANALISIS PERBANDINGAN ALGORITMA FUZZY SLIQ DECISION TREE DAN ALGORITMA SLIQ DALAM PROSES KLASIFIKASI DATAMINING

Ardana Riswanto<sup>1</sup>, Shaufiah<sup>2</sup>, Arie Ardiyanti Suryani<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

### Abstrak

Pohon keputusan merupakan salah satu metode klasifikasi dalam data mining. Pohon keputusan mengelompokkan data ke dalam kelas-kelas tertentu yang sudah didefinisikan sebelumnya. SLIQ merupakan salah satu contoh metode pohon keputusan yang menggunakan Gini Index dan dapat digunakan untuk mengklasifikasikan data bertipe numeric dan kategorial. Namun SLIQ memiliki keterbatasan dalam mengklasifikasikan data-data continuous, karena split point pada SLIQ menggunakan nilai diskret sehingga kurang sesuai dengan karakteristik data continuous. Hal ini mengakibatkan data klasifikasi yang dihasilkan akan cenderung kasar, kaku dan kurang sesuai dengan karakter data yang diberikan. Oleh karena itu diberikan metode Fuzzy SLIQ Decision Tree dengan menambahkan fuzzy dalam penghitungan split point algoritma SLIQ. Dengan penambahan fuzzy ini penentuan split point akan didasarkan pada keanggotaan suatu data terhadap titik potong tertentu. Dengan metode ini, data continuous tidak diklasifikasikan secara diskret, tetapi dengan memperhitungkan nilai keanggotaan terhadap titik-titik potong yang terdapat dalam model pohon yang dihasilkan. Dari hasil pengujian didapat bahwa Fuzzy SLIQ Decision Tree menghasilkan akurasi yang secara umum lebih baik dibandingkan dengan algoritma SLIQ dan mempunyai waktu proses pembentukan model yang lebih cepat dibandingkan dengan algoritma SLIQ.

Kata Kunci : Data Mining, Klasifikasi, Pohon Keputusan, Fuzzy SLIQ Decision Tree, SLIQ

---

### Abstract

Decision tree is one of classification method in data mining. Decision tree collects data into certain class that has been defined before. SLIQ is one decision tree method that use Gini Index and can use to classify categorical and numerical data. But SLIQ have limitation to classify continuous data, because split point in SLIQ use discrete value so not fit with continuous data characterization. It can make classification data that produce with SLIQ is rough, rigid and less appropriate with given data characterization. Because of that, give the Fuzzy SLIQ Decision Tree algorithm with add fuzzy in split point calculation SLIQ algorithm. With add fuzzy in the algorithm, split point decision will base with membership value the data to split point. With this method, continuous data doesn't classify use discrete split point but use calculation membership value split points that describe in tree model that has been produced in training process. From the test result we can see that Fuzzy SLIQ Decision Tree Algorithm generally give better accuracy than SLIQ Algorithm and give the time model building process faster than SLIQ algorithm.

Keywords : Data Mining, Classification, Decision Tree, Fuzzy SLIQ Decision Tree, SLIQ

---

# 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.

## 5 Kesimpulan dan Saran

### 5.1 Kesimpulan

Dari hasil pengujian dan analisis yang sudah dilakukan dapat diambil kesimpulan sebagai berikut ini :

1. Algoritma Fuzzy SLIQ Decision Tree menghasilkan akurasi yang secara umum lebih baik dibandingkan dengan algoritma SLIQ tanpa pruning dan algoritma SLIQ dengan menggunakan Pruning pada dataset yang memiliki distribusi kelas yang relatif seimbang.
2. Ketiga algoritma akan menghasilkan akurasi yang secara umum kurang baik jika kelas dalam dataset tersebut besar, dan jika distribusi kelasnya tidak seimbang, maka akurasi yang dihasilkan oleh algoritma FSDT akan lebih buruk dibandingkan dua algoritma yang lain, karena adanya mekanisme mengkalkulasikan seluruh node dalam proses inferensi.
3. Algoritma SLIQ dengan menggunakan proses Pruning akan menghasilkan akurasi klasifikasi terhadap data uji yang lebih baik jika dibandingkan dengan Algoritma SLIQ tanpa Pruning.
4. Waktu proses pembentukan model algoritma SLIQ dengan Pruning jauh lebih lama jika dibandingkan dengan algoritma SLIQ tanpa Pruning dan algoritma Fuzzy SLIQ Decision Tree karena *split point* pada FSDT hanya dihitung pada saat terjadi perubahan nilai data.
5. Waktu proses inferensi algoritma Fuzzy SLIQ Decision Tree lebih lama jika dibandingkan dengan algoritma SLIQ tanpa Pruning dan algoritma SLIQ dengan Pruning karena proses inferensi pada Fuzzy SLIQ mengkalkulasikan semua node pada pohon.
6. Ukuran pohon yang dihasilkan dengan menggunakan algoritma SLIQ with Pruning relatif kecil karena adanya metode pruning, tetapi tidak jauh berbeda dengan pohon yang dihasilkan dengan menggunakan Fuzzy SLIQ Decision Tree.
7. Algoritma Fuzzy SLIQ Decision Tree memberikan hasil klasifikasi yang secara umum lebih baik jika dibandingkan algoritma SLIQ with Pruning dan juga waktu proses pembentukan model yang jauh lebih cepat tetapi dengan waktu proses inferensi yang lebih lama.
8. Dari hasil pengujian dapat disimpulkan bahwa algoritma Fuzzy SLIQ Decision Tree cocok digunakan apabila memerlukan waktu proses yang cepat untuk membentuk model pohon keputusan dan dataset yang digunakan memiliki distribusi kelas yang seimbang. Hal ini karena jika dibandingkan algoritma uji yang lain, Fuzzy SLIQ Decision Tree menghasilkan akurasi yang tidak kalah baik tetapi memiliki waktu proses yang jauh lebih cepat.

### 5.2 Saran

Dari penulisan tugas akhir ini dapat diberikan beberapa saran untuk pengembangan dan penelitian berikutnya antara lain :

1. Dapat dilakukan pengujian perbandingan performansi antara algoritma Fuzzy SLIQ dengan algoritma klasifikasi pohon keputusan yang lain seperti C4.5, IDR, Sprint, dan lain sebagainya
2. Dapat dilakukan pengujian perbandingan performansi antara algoritma Fuzzy SLIQ Decision Tree dengan algoritma *Fuzzy Decision Tree* yang lain seperti Fuzzy C4.5, Fuzzy IDR, dan lain sebagainya.

3. Dapat dilakukan percobaan dan pengujian terhadap perbandingan algoritma dengan menggunakan dataset yang sangat besar, dataset yang memiliki tipe data lain seperti kategorial, dan lain sebagainya.
4. Dapat ditambahkan suatu metode tertentu untuk mencari nilai parameter fuzzy yang paling optimal secara otomatis, misalkan dengan menambahkan Algoritma Genetika atau JST.



## Daftar Pustaka

- [1] Suyanto. 2007. *Artificial Intelligence Searching, Reasoning, Planning dan Learning*. Informatika : Bandung
- [2] Saufiah. 2005. *Klasifikasi dalam Data Mining Menggunakan Algoritma SLIQ*. Bandung. Institut Teknologi Telkom. Tugas Akhir
- [3] Pang-Ning Tan, Michael Steinbach and Vipin Kumar. 2005. *Introduction to Data Mining*
- [4] B. Chandra and P.P. Varghese. *Fuzzy SLIQ Decision Tree Algorithm*. presented at IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2008, pp.1294-1301.
- [5] B. Chandra and P.P. Varghese. *On Improving Efficiency of SLIQ Decision Tree Algorithm*. presented at proceeding International Joint Conference on Neural Networks, Orlando, Florida, USA, 2007.
- [6] M. Mehta, R. Agrawal, and J. Rissanen. *SLIQ : A Fast Scalable Classifier for Data Mining*. In Proc. Extending Database Technol., Avignon, France, 1996, pp.18-32
- [7] B. Santosa. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu : Yogyakarta
- [8] B. Santosa. 2007. *Data Mining Terapan dengan Matlab*. Graha Ilmu : Yogyakarta
- [9] C. Z. Janikow. *Fuzzy Decision Trees: Issues and Methods*. presented at IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 28, no. 1, pp. 1-14, Feb. 1998
- [10] J. H. Halimsetiawan. 2009. *Struktur Data Tree/Pohon dalam Bahasa Java*. diunduh dari [www.tutorialpemrograman.wordpress.com](http://www.tutorialpemrograman.wordpress.com) pada tanggal 16 Januari 2011