

DAFTAR ISI

**STUDI PERBANDINGAN METODE POHON KEPUTUSAN DAN1-10 POHON
KEPUTUSAN FUZZY PADA KLASIFIKASI PENUTUP LAHAN**
Dwi Wahyu dan M. Rahmat Widyanto

**IDENTIFIKASI SINYAL *ELEKTRODE ENCHEPALO GRAPH* UNTUK11-18
MENGERAKKAN KURSOR MENGGUNAKAN TEKNIK *SAMPLING*DAN JARINGAN
SYARAF TIRUAN**
Hindarto, Moch. HariadidanMauridhi Hery Purnomo

**OPTIMASI FUNGSI MULTI-OBYEKTIIF BERKENDALA MENGGUNAKAN19-26
ALGORITMA GENETIKA ADAPTIF DENGAN PENGKODEAN *REAL***
Wayan Firdaus Mahmudy dan Muh. Arif Rahman

**PERINGKAT WEBSITE PERGURUAN TINGGI BERBASISANALISA 27-36HYPERLINK
MENGGUNAKAN *FACTOR ANALYSIS***
Yuhefizar, Mochamad Hariadi, dan Yoyon K Suprapto

**DETEKSI KEBERADAAN KALIMAT SAMA SEBAGAI INDIKASI37-44PENJIPLAKAN
DENGAN ALGORITMA *HASHING* BERBASIS *N-GRAM***
Diana Purwitasari, Putu Yuwono Kusmawan, dan Umi Laili Yuhana

**IINUSAT-1: SATELIT-NANO PERDANA DI INDONESIA UNTUK 45-54
PENELITIAN DAN PENDIDIKAN**

Tri Kuntoro Priyambodo, Agfianto Eko Putra, Muh. Asvial, Ridanto Eko Putro, Gamantyo, Endra Pitowarno, Son Kuswadi, dan Gunawan S. Prabowo

**PERILAKU TAKTIS UNTUK *NON-PLAYER CHARACTERS* DI GAME55-
64PEPERANGAN MENIRU STRATEGI MANUSIA MENGGUNAKAN FUZZY LOGIC
DAN *HIERARCHICAL FINITE STATE MACHINE***

Supeno Mardi Susiki Nugroho, Yunifa Miftachul Arif, Mochamad Hariadi, danMauridhi Hery Purnomo

STUDI PERBANDINGAN METODE POHON KEPUTUSAN DANPOHONKEPUTUSAN FUZZY PADA KLASIFIKASI PENUTUP LAHAN

^aDwi Wahyu, ^bM. Rahmat Widyanto

Fakultas Ilmu Komputer, Universitas Indonesia, Depok 16424, Jawa Barat, Indonesia
E-Mail: ^aflash_worked@yahoo.com

Abstrak

Penelitian ini melakukan studi perbandingan kinerja metode pohon keputusan dengan pohon keputusan *fuzzy* pada klasifikasi penutup lahan untuk membantu mengelola sumber daya alam/hasil buminya. Metode pohon keputusan merupakan teknik yang lebih sederhana dan memberikan hasil yang optimal, namun sering mengalami kesulitan pada data yang tidak lengkap (*missing value*), data enumerasi, maupun data samar yang sering terdapat pada dataset penutup lahan. Sedangkan metode pohon keputusan *fuzzy* menerapkan konsep *fuzzy* pada pembentukan cabang sehingga menghasilkan pohon yang lengkap dan diharapkan mampu memenuhi kekurangan tersebut. Beberapa kumpulan data penutup lahan dilakukan untuk mengukur kinerja kedua metode tersebut pada beberapa pengujian dengan mengklasifikasikan data masukan dan hasilnya dibandingkan dengan klasifikasi sebenarnya. Hasil penelitian menunjukkan bahwa metode pohon keputusan *fuzzy* mampu memberikan kenaikan tingkat pengenalan sebesar 4.96% dan membutuhkan waktu komputasi 38.13% lebih dibanding metode pohon keputusan. Metode klasifikasi dengan kemampuan pengenalan data yang tidak lengkap maupun samar dengan waktu komputasi yang cepat menjadi kontribusi yang baik untuk diterapkan pada aplikasi Sistem Informasi Geografis.

Kata kunci: Pohon Keputusan, Pohon Keputusan *Fuzzy*, Klasifikasi Penutup Lahan.

Abstract

Comparative Study of Decision Tree Method and Fuzzy Decision Tree on the Land Cover Classification. This research is a comparative study of the performance of decision tree method with fuzzy decision tree classification of land cover to help manage natural resources. The decision tree method is a simple technique and provides optimal results, but often have difficulty on incomplete data (missing value), the data enumerated, or vague, often found on land cover datasets. While the fuzzy decision tree method for implementing fuzzy concepts in the branch formation so as to produce a complete and expected to accommodate the shortage. Some land cover datasets are selected. Then two methods of performance measurement in some test by grouping the input data and the results compared with the actual classification. The results showed that the fuzzy decision tree method can provide the level of recognition by the 4.96%, but it takes much longer computation 38.13%. The classification method with data recognition capabilities that are incomplete or not clear with time computing is fast becoming a good contribution to be implemented on the application of Geographic Information System.

Keywords: Decision Tree, Fuzzy Decision Tree, Land Cover Classification.

PENDAHULUAN

Citra *remote sensing* merupakan hasil yang *up to date*. Pendekatan statistik biasanya didasarkan hanya pada informasi spektral, sedangkan seorang ahli penerjemah foto menggabungkan informasi spektral, spasial dan informasi kontekstual sebagai informasi tambahan. Untuk dapat mengklasifikasi secara akurat diperlukan suatu metode yang dapat menganalisa mendekati hasil analisa seorang ahli penerjemah foto *remote sensing*.

Pengetahuan tentang geografi dari manusia dalam menginterpretasikan suatu obyek adalah samar (*fuzzy*) dan tak menentu. Pengambilan data di lapangan sering kali tidak lengkap serta mencakup luas daerah yang besar dan sulit dijangkau, sehingga membuat semakin sulit dalam membangun pengetahuan untuk suatu sistem cerdas. Sebagai tambahan, beberapa sistem yang berdasar pengetahuan yang ada merupakan sistem yang khusus, dimana sangat sulit untuk memindahkan mereka ke proyek yang lain dengan sasaran yang berbeda dan kondisi geografis yang berbeda. Hal ini membuat biaya untuk membangun suatu sistem cerdas semakin mahal. Meskipun para ahli dapat menuangkan pengetahuan mereka dalam membuat keputusan, tetapi mereka umumnya tidak dapat menuangkan atau merumuskan pengetahuan mereka secara tepat dalam suatu bentuk secara teratur yang cukup, benar dan lengkap dalam suatu bentuk aplikasi komputer [1]. Sehingga pada perkembangan selanjutnya dilakukan penerapan konsep *fuzzy* pada pohon keputusan yang akan dibangun.

Pohon keputusan yang digunakan dalam klasifikasi di sini merupakan teknik yang lebih sederhana dan memberikan hasil yang terbaik [2], namun sering mengalami kesulitan pada data yang tidak lengkap (*missing value*), berlanjut (*continue*), data *string*, maupun samar yang sering terdapat pada *dataset* penutup lahan. Sedangkan untuk menampung kenyataan di lapangan, telah berkembang menjadi pohon keputusan *fuzzy* yang menerapkan konsep *fuzzy* pada pembentukan cabang, sehingga menghasilkan pohon yang lengkap dan diharapkan mampu menutup kekurangan tersebut.

Dengan adanya beragam macam *dataset* penutup lahan, maka perlu dilakukan perbandingan kedua metode tersebut untuk mengetahui kemampuan pengenalan terhadap

data penutup lahan. Setelah itu baru akan dibandingkan kinerjanya dan diharapkan dapat diterapkan pada aplikasi Sistem Informasi Geografis. Penelitian ini akan membandingkan kinerja kedua metode pembelajaran terbimbing berbasis pohon keputusan tersebut sehingga didapat klasifikasi penutup lahan yang akurat.

METODE POHON KEPUTUSAN

Istilah lain dari pembelajaran terbimbing adalah klasifikasi. Kinerja klasifikasi sangat tergantung dari karakteristik dari data yang diklasifikasikan. Tidak ada sebuah jenis klasifikasi pun yang dapat bekerja dengan sangat baik pada semua permasalahan yang diberikan, dengan kata lain tidak ada yang bersifat menyeluruh atau universal [3]. Berbagai macampengujian telah dilakukan pengukuran kinerjanya dan karakteristik dari data menentukan kinerja klasifikasi. Menentukan metode klasifikasi yang sesuai untuk masalah yang diberikan kadangkala lebih merupakan seni daripada ilmu pengetahuan.

Pengujian *fuzzy* Perilaku *NPC Scout*

Tujuan dari pembelajaran terbimbing dari sebuah contoh menyeluruh adalah menemukan sebuah fungsi g , diberikan suatu set dari bentuk $(x, g(x))$ [3]. Diasumsikan g di kenal sebagai contoh data yang bebas/berdiri sendiri dan variabel yang secara acak terdistribusi secara identik, digambarkan menurut kemungkinan distribusi p sebagai sesuatu yang besar, *possibly infinite population*. Lebih jauh lagi, dimisalkan keberadaan *task-specific loss function L* seperti pada Persamaan (1).

$$L : Y \times Y \rightarrow R^+ \quad (1)$$

di mana Y adalah codomain dari g dan L memetakan kedalam *nonnegative real numbers* (untuk pembatasan lebih lanjut digantikan oleh L). Banyaknya $L(z, y)$ adalah kerugian yang terjadi oleh perkiraan z sebagai nilai dari g pada titik yang diberikan ketika nilai sebenarnya adalah y .

Resiko yang berhubungan dengan fungsi f kemudian didefinisikan sebagai *loss function* yang diharapkan, seperti yang dirumuskan pada Persamaan (2).

$$R(f) = \sum_i L(f(x_i), g(x_i))P(x_i) \quad (2)$$

```

fungsi ID3 (R: himpunan bukan atribut
categorical,
C: atribut categorical,
S: himpunan pelatihan) kembali
ke pohon keputusan;
mulai
Jika S kosong, kembali ke titik utama
dengan nilai Gagal;
Jika S terdiri dari catatan dengan
semuanya nilainya sama untuk
atribut categorical, kembali ke
titik utama dengan nilai tersebut;
Jika R kosong, kembali ke titik
utama sebagai nilai yang paling
banyak frekwensi nilai atribut
categorical yang ada di catatan S;
[perhatikan, akan ada kesalahan,
sehingga, kesalahan akan
terkласifikasi secara tepat];
Biarkan D menjadi atribut dengan
Gain(D,S) terbesar di antara
atribut dalam R;
Biarkan {dj| j=1,2, .., m} menjadi
nilai atribut D;
Biarkan {Sj| j=1,2, .., m} menjadi
subset dari S terdiri dari catatan
dengan nilai dj untuk atribut D;
Kembali ke pohon dengan akar
berlabel D dan arcs berlabel d1,
d2, .., dm pergi ke pohon masing-
masing
ID3(R-{D}, C, S1), ID3(R-{D}, C, S2),
.., ID3(R-{D}, C, Sm);
selesai ID3;

```

Gambar 1. Algoritma ID3.

Jika kemungkinan penyebaran p adalah *discrete* (*analog continue*) akan menyebabkan sebuah *definite integral* dan sebuah *probability density function*.

Tujuannya sekarang adalah menemukan sebuah fungsi f^* di antara beberapa subkelas tetap dari fungsi, dimana resiko $R(f^*)$ adalah minimal. Kadangkala, sejak g hanya diketahui hanya terbatas untuk suatu *set of points* $(x_1, y_1), \dots, (x_n, y_n)$, hanya dapat diperkirakan resiko sebenarnya, sebagai contoh dengan *empirical risk* yang dirumuskan dalam Persamaan (3).

$$\tilde{R}_n(f) = \frac{1}{n} \sum_{i=1}^n L(f(x_i), y_i) \quad (3)$$

Pemilihan fungsi f^* yang meminimalkan *empirical risk* dikenal sebagai prinsip dari *empirical risk minimization*. Teori pembelajaran statistik digunakan untuk menyelidiki seberapa kondisi dibawah *empirical risk minimization* dapat diterima dan seberapa baik perkiraan masih dapat diharapkan.

Metode Pohon Keputusan

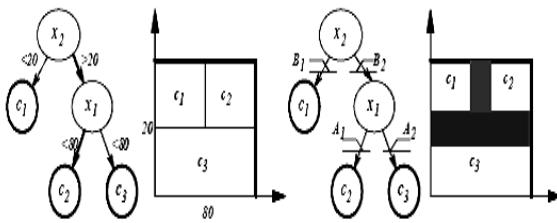
Konsep pohon keputusan pada dasarnya adalah mengubah data menjadi pohon keputusan dan aturan hirarki (aturan-aturan keputusan) [4] yang pada perkembangan selanjutnya dapat disederhanakan dengan menghilangkan cabang-cabang atau aturan-aturan yang tidak perlu. Sedangkan atribut-atribut yang menyertai data tersebut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Salah satu atribut merupakan atribut yang menyatakan data penyelesaian *per-item* data yang disebut dengan klasifikasi atau kelas. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut cuaca mempunyai *instance* berupa cerah, berawan, dan hujan.

Pembelajaran dari contoh (*inductive learning*) yang digunakan didasarkan pada *Quinlan's Learning Algorithm* dari Ross Quinlan. Perkembangan teknik pohon keputusan ini berawal dari Algoritma *ID3* yang disempurnakan menjadi *C4.5* dan *C5* (untuk kepentingan komersial) [5]. Adapun Algoritma *ID3* seperti yang ditunjukkan dalam Gambar 1.

Titik percabangan terakhir (dari cabang) disebut *leaf*. Untuk setiap *leaf*, sebuah keputusan telah dibuat dan diaplikasikan ke semua observasi di dalam *leaf*. *Leaf* tersebut merupakan hasil akhir pengelompokan data dari pohon keputusan yang membagi ke dalam kelompok klasifikasi.

Pohon Keputusan Fuzzy

Potensi pohon keputusan *fuzzy* dalam mengembangkan kelebihan dan keumuman dari masalah klasifikasi didasarkan pada penggunaan pemikiran *fuzzy*. Pada Gambar 2 akan digambarkan *classical* dan pohon keputusan *fuzzy*. Dalam kedua pohon keputusan, setiap jalan dari titik *root* ke sebuah titik *leaf* membangun sebuah aturan klasifikasi.



Gambar 2. Pohon Keputusan *Classic* dan *Soft Discretisasi*.

Dengan menggunakan klasifikasi *crisp*, jarak keputusan dibagi dalam sebuah *set non-overlapping subspace*, seperti terlihat dalam Gambar 2 sebelah kiri, di mana setiap titik data merujuk pada suatu kelas tertentu. Sebaliknya, pohon keputusan *fuzzy* memberikan hasil antara [0, 1], tingkat kemungkinan suatu obyek memenuhi klasifikasi, sebagaimana terlihat pada Gambar 2 sebelah kanan. Pohon keputusan *fuzzy* menyediakan lebih banyak cara untuk menghindari kesalahan klasifikasi. Sebagai contoh, mungkin saja suatu obyek memiliki kelas c_1, c_2, c_3 dengan $\pi_1=0, \pi_2=0.52, \pi_3=0.48$. Menurut hasil ini, pengguna dapat menentukan pilihan akhir mereka atau melakukan penyelidikan lebih jauh. Dengan kata lain, rata-rata kesalahan klasifikasi dapat dikurangi.

Klasifikasi dari obyek tak dikenal yang diberikan, didapat dari tingkat kecocokan dari setiap titik dari *root* ke *leaf*. Dalam contoh sebelumnya, kemungkinan sebuah obyek termasuk dalam kelas c_2 , dihitung sebagaimana Persamaan (4).

$$\pi_2 = \bigotimes [B_2(x_2), A_1(x_1)] \quad (4)$$

Dimana *circle plus* \bigotimes adalah operasi *fuzzy product* sedangkan $B_2(x_2)$ dan $A_1(x_1)$ adalah *membership degree* dari x_2 ke B_2 dan x_1 ke A_1 . Dengan cara yang sama, kemungkinan sebuah obyek memiliki setiap kelas klasifikasi dapat dihitung (misalnya suatu obyek memiliki kelas c_1, c_2, c_3 dengan $\pi_1=0, \pi_2=0.52, \pi_3=0.48$). Jika lebih dari satu *leaf* yang berhubungan dengan sebuah kelas yang sama, akan dipertimbangkan sebagai kemungkinan dari kelas yang sesuai, dimana maksimum operator digunakan sebagai *fuzzy sum operation* \bigotimes . Pada akhirnya, jika suatu nilai kemungkinan, semisal π_k , jauh lebih besar dari yang lain, kemudian kelas c_k akan ditunjuk sebagai kelas dari obyek, dapat dikatakan pohon keputusan meramalkan sebuah penyebaran pada semua kelas [6].

Diskritisasi lunak dapat dilihat sebagai sebuah pengembangan dari diskritisasi keras, dan pengukuran-pengukuran informasi klasik didefinisikan dalam probability domain yang telah diperluas menjadi definisi baru dalam *possibility domain* didasarkan pada teori *fuzzy set*. Sebuah *crisp set* A_c diperjelas dengan sebuah fungsi karakterisasi yang tajam $A_c(a): \Omega \rightarrow \{0, 1\}; a \in \Omega$, sebagai alternatif sebuah *fuzzy set* A adalah dikarakteristik oleh sebuah fungsi keanggotaan $A_c(a): \Omega \rightarrow \{0, 1\}; a \in \Omega$. Keanggotaan $A(a)$ disebut kemungkinan dari A untuk mengambil sebuah nilai $a \in \Omega$. Kemungkinan *fuzzy set* A telah didefinisikan, menurut Zadeh, dengan $P_F(A) = \int_{\Omega} A(a) dP$, di mana dP adalah sebuah pengukuran kemungkinan pada Ω , dan subscript *F* digunakan untuk bersekutu dengan *fuzzy term*. Khususnya, bila A didefinisikan pada *discrete domain* $\Omega = \{a_1, \dots, a_i, \dots, a_m\}$ dan kemungkinan $P(a_i) = p_i$ maka kemungkinannya seperti yang ditunjukkan pada Persamaan (5).

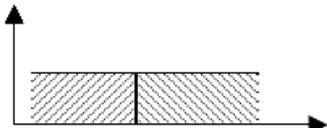
$$P_F(A) = \sum_{i=1}^m A(a_i) P_i \quad (5)$$

Jika $Q = \{A_1, \dots, A_\chi\}$ adalah bagian dari *fuzzy set* pada Ω , Q disebut bagian *fuzzy* dari Ω ketika memenuhi ketentuan pada Persamaan (6).

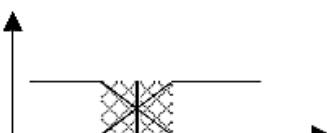
$$\sum_{\gamma=1}^{\chi} A_\gamma(a) = 1, \forall a \in \Omega \quad (6)$$

Sebuah diskritisasi keras seperti ditunjukkan pada Gambar 3, didefinisikan dengan sebuah *threshold*, dimana membangun di antara dua *crisp set*. Sedangkan sebuah diskritisasi lunak seperti ditunjukkan pada Gambar 4, didefinisikan oleh sebuah pasangan *fuzzy set* ulang membuat sebuah *fuzzy partisi*. Gambaran untuk metode klasik sebagai pembagian yang tidak tumpang tindih, dan diskritisasi lunak adalah tumpang tindih. Diskritisasi lunak didefinisikan dengan tiga parameter/fungsi, yaitu yang satu adalah titik silang T , sedangkan dua lainnya adalah fungsi keanggotaan dari pasangan *fuzzy set* A_1 dan A_2 . Titik silang T adalah tempat diskritisasi lunak, ditentukan dengan cara memaksimalkan *information gain*, dan fungsi keanggotaan dari pasangan *fuzzy set* ditentukan menurut karakteristik atribut data, seperti persekutuan atribut yang meragukan. Umumnya, tumpang tindih lebar digunakan untuk atribut yang keragu-raguannya tinggi, sebagai contoh, kita dapat menggunakan rata-

rata jarak titik data sebagai lebar tumpang tindih [6].



Gambar 3. Diskritisasi Keras dengan Sebuah *Threshold*.



Gambar 4. Diskritisasi Lunak dengan Partisi *Fuzzy*.

USULAN RANCANGAN SISTEM STUDI PERBANDINGAN

Tujuan klasifikasi adalah mengelompokkan obyek pengamatan berdasarkan serangkaian peubah-peubah yang disebut sebagai peubah bebas. Kedua sistem pengembangannya yang dibangun sama-sama menggunakan dasar teori pohon keputusan, yang kemudian diterapkan sesuai metode masing-masing. Metode pohon keputusan dan konsep *fuzzy* dilatihkan dengan data yang sama sehingga diharapkan memperoleh hasil perbandingan yang optimal.

Karakteristik data penutup lahan merupakan data yang unik, dan variatif karena mencakup wilayah pengambilan data yang luas, dilakukan oleh orang atau organisasi yang berbeda-beda dengan pengetahuan yang berbeda dan selang waktu yang lama. Selain itu data yang didapat juga kadangkala tidak lengkap, dan samar. Dengan demikian pengujian yang dilakukan didasarkan pada kemampuan menangani jenis-jenis data pada data penutup lahan, seperti ditunjukkan pada Gambar 5. Jenis pengujian yang dilakukan meliputi:

1. Pengujian data *integer*.

Data *integer* adalah data dengan nilai bilangan bulat [7]. Bilangan bulat ini juga terbagi dalam dua bagian yaitu bilangan bulat bertanda dan tidak bertanda. Nilai dari data ini berkisar dari -2,147,483.648 sampai 2,147,483.647. Data geografi yang bersifat *integer* antara lain tahun penanaman, curah hujan, jumlah benih, dan konversi warna (citra penginderaan jarak jauh) ke data *integer* (nilai 1-255).

2. Pengujian data berlanjut (*continue*).

Data *continue* adalah data yang nilainya ada pada suatu interval tertentu atau berada pada nilai yang satu ke nilai yang lainnya. Contohnya penggunaan kata sekitar, kurang lebih, kira-kira, dan sebagainya [8]. Dalam aplikasinya data berlanjut merupakan data yang satuannya bisa dalam pecahan atau bilangan real, misalnya minyak dalam 1/2 liter, panjang dalam 0,2 meter dan sebagainya [9]. Fungsi berlanjut digambarkan sebagai kurva. Data geografi yang didapat merupakan data real yang merupakan hasil pengukuran di lapangan, antara lain contohnya adalah diameter dasar terbaik dari suatu pohon, diameter batang, dan diameter mahkota.

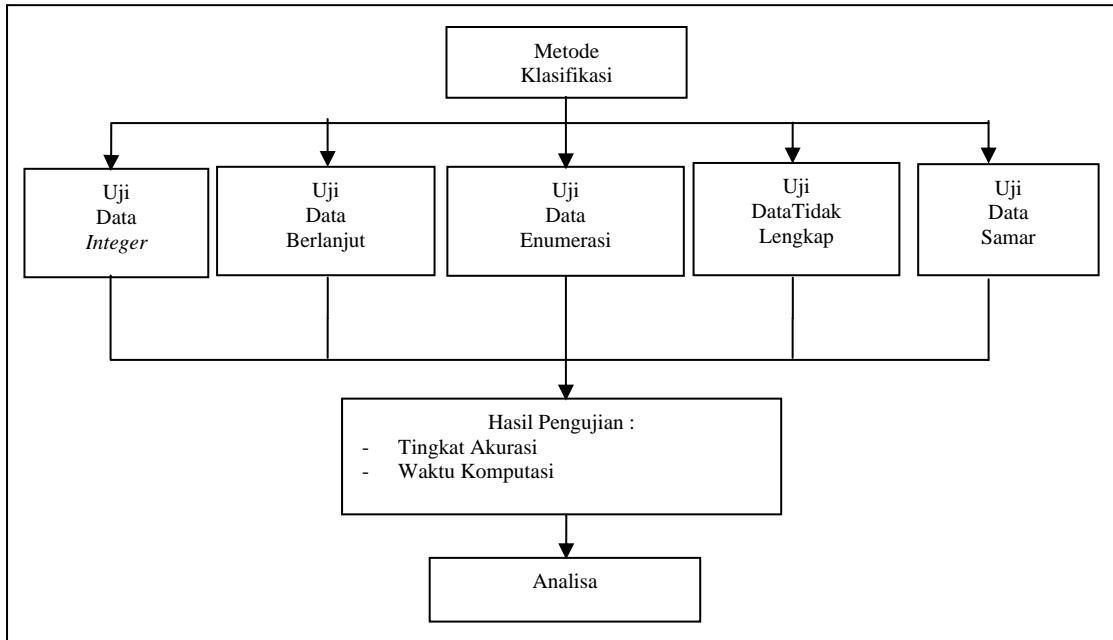
3. Pengujian data enumerasi.

Dalam pemrograman komputer, sebuah tipe enumerasi (juga disebut pencacahan atau enum) adalah jenis data yang terdiri dari satu himpunan nilai nama yang disebut unsur, anggota atau pencacahan dari jenis ini. Nama pengidentifikasi enumerator yang biasanya berperilaku sebagai konstanta dalam bahasa itu [10]. Sebuah variabel yang telah dinyatakan memiliki tipe *enumerated* dapat dinyatakan enumerasi sebagai suatu nilai. Sebagai contoh dalam data geografi adalah nama tempat yang disingkat, lokasi suatu daerah, lokasi di peta, maupun daerah ketinggian.

Beberapa jenis pencacahan dapat dibangun ke dalam bahasa tersebut. Jenis *Boolean*, misalnya, seringkali berupa satu *set* nilai pencacahan *FALSE* dan *TRUE*. Banyak bahasa memungkinkan pengguna untuk menentukan tipe enumerasi baru. Nilai dan peubah tipe enumerasi biasanya diterapkan sebagai string panjang tetap-bit, sering dalam format dan ukuran yang sesuai dengan beberapa jenis *integer*. Beberapa bahasa, terutama bahasa pemrograman sistem, memungkinkan pengguna untuk menentukan kombinasi bit yang akan digunakan untuk setiap pencacahan.

4. Pengujian data tidak lengkap (*missing value*).

Data tidak lengkap (*missing value*) kadang kala terjadi dalam pengamatan lapangan. Dalam statistik, nilai-nilai yang hilang terjadi ketika nilai-nilai data yang ada tidak



Gambar 5. Rancangan Sistem.

Tabel 1. Perbandingan (%) Tingkat Akurasi dan Waktu Komputasi.

Jenis Pengujian	Tingkat Akurasi		Waktu Komputasi
	Data Integer	Data Berlanjut	
Data Integer	8.85%	-7.86%	0.86 detik (31 %)
Data Berlanjut	6%	12.33%	1 detik (59 %)
Data Enumerasi	5.50%	--	1.16 detik (52 %)
Data Tidak Lengkap	4.96 %	--	0.2 detik (10.5 %)
Nilai rata-rata	0.8 detik (38.13 %)		

disimpan untuk peubah-peubah dalam pengamatan saat ini. Paket statistik modern telah membuat perjanjian dengan nilai-nilai lebih mudah hilang. Sering menggunakan perkiraan maksimum *likelihood* untuk ringkasan statistik, interval keyakinan, dan lain-lain [11]. Data-data yang tidak lengkap ini dapat menyebabkan bencana dan menyulitkan dalam memberikan kesimpulan yang handal dan akurat. Dengan demikian, peneliti selalu berusaha menghindari nilai-nilai yang hilang sebanyak mungkin [12]. Meskipun pada kenyataannya hal ini

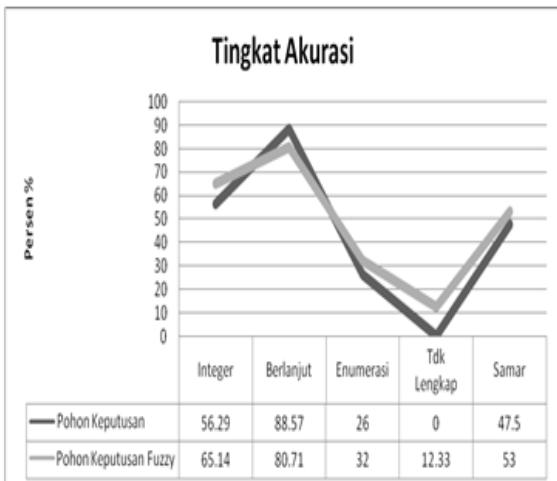
tidak dapat dihindari. Dalam percobaan pohon keputusan *fuzzy* ini, nilai-nilai yang hilang ditandai dengan nilai -1.

5. Pengujian Data Samar (*Fuzzy*).

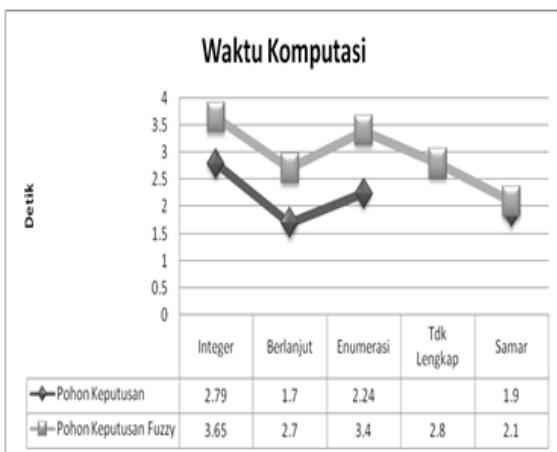
Logika *fuzzy* memungkinkan nilai keanggotaan antara 0 dan 1, tingkat keabuan dan juga hitam dan putih, dan dalam bentuk *linguistic*, konsep tidak pasti seperti sedikit, lumayan, dan sangat. Dia berhubungan dengan *setfuzzy* dan teori kemungkinan [13]. Data samar (*fuzzy*) dalam data geografi digunakan untuk mengelompokkan suatu tanah termasuk klasifikasi sangat baik (*best*), baik (*good*), sedang (*average*) ataupun tidak baik (*low*) untuk dibudidayakan suatu tanaman tertentu. Tidak adanya batasan yang jelas dalam pengelompokkan tersebut membuat kesulitan dalam memberikan hasil analisa yang akurat dan handal.

HASIL DAN PEMBAHASAN

Berdasarkan hasil ujicoba terhadap seluruh pengujian didapat bahwa (seperti ditunjukkan pada Tabel 1) untuk lima jenis pengujian sebesar 4,96 %, sedangkan waktu komputasi metode keputusan *fuzzy* lebih lambat dari metode keputusan untuk lima jenis pengujian sebesar 0,8 detik (38,13 %).



Gambar 6. Grafik Tingkat Akurasi.



Gambar 7. Grafik Waktu Komputasi.

Aspek pertama yang dianalisis adalah tingkat akurasi metode pohon keputusan *fuzzy* yang lebih tinggi dibandingkan metode pohon keputusan. Hal ini dapat dijelaskan antara lain karena cabang *fuzzy* yang terbentuk (metode pohon keputusan *fuzzy*) pada aturan hirarki pohon keputusan memberikan kesempatan kepada *dataset* untuk dapat masuk ke tiap ranting (*node*) dengan prosentase tertentu. Selanjutnya tiap ranting tersebut akan membentuk sub-ranting (*node*) atau anak cabang kembali sehingga *dataset* tersebut dapat memiliki kesempatan kembali untuk masuk kedalamnya dengan prosentase tertentu. Prosentase ini dapat semakin membesar atau makin mengecil sehingga dapat diabaikan (dihilangkan).

Hal ini membuat pembentukan aturan hirarki pohon keputusan menjadi semakin dalam. Dengan semakin dalamnya aturan

hirarki yang terbentuk membuat data yang diujikan dapat semakin akurat karena data tersebut dapat masuk ke beberapa kelas tetapi dengan nilai prosentase yang berbeda. Nilai prosentase yang tertinggi merupakan hasil klasifikasi dari metode pohon keputusan *fuzzy*.

Pada pengujian data berlanjut, mengalami perbedaan yang tajam, yaitu tingkat akurasi metode pohon keputusan *fuzzy* lebih rendah 7,86 % dari metode pohon keputusan (Gambar 6). Hal ini disebabkan proses cara kerja masing-masing metode yang berbeda. Pada metode pohon keputusan *fuzzy*, data masukan yang diberikan akan dikelompokkan menjadi dua bagian yaitu data *discrete* dan data *fuzzy* (dengan parameter yang ditentukan). Pengelompokan ini ditentukan dengan mendefinisikan terlebih dahulu data yang akan diolah. Pada kasus ini, data *integer*, data berlanjut, dan data *fuzzy* dikelompokkan dalam data *fuzzy* dengan parameter yang ditentukan agar dapat hasil yang optimal. Data berlanjut (pecahan) yang diolah sebagai data *fuzzy*, ternyata hanya mampu mengenali sampai tiga digit di belakang koma, sedangkan data berlanjut yang digunakan memiliki sembilan digit di belakang koma. Hal ini sangat mengurangi ketelitian dalam mengenali data masukan, mengingat rentang data dalam satu atribut sangat pendek yaitu antara 0 sampai 0,26.

Sedangkan pada metode pohon keputusan, data masukan yang diterima akan dikelompokkan menjadi dua bagian, yaitu data *discrete* dan data berlanjut. Pengelompokan ini dikelompokkan secara otomatis dan didefinisikan secara otomatis. Data *integer* dan data *continue* akan dimasukkan dalam kelompok data berlanjut, sedangkan data enumerasi dan data *fuzzy* akan dikelompokkan ke dalam data *discrete*. Data berlanjut yang diolah ternyata mampu mengenali data masukan hingga lebih dari sembilan digit di belakang koma, dengan kata lain metode ini sangat akurat dalam mengenali data berlanjut karena memiliki ketelitian hingga sembilan digit di belakang koma.

Pada pengukuran data tidak lengkap, metode pohon keputusan mengalami kegagalan (*hang/error*) dalam mengenali data tidak lengkap (nilai yang hilang). Hal ini disebabkan data masukan (nilai yang hilang) yang diterima ketika masuk kedalam suatu titik percabangan tidak dapat masuk ke salah satu anak cabang

karena tidak dapat memenuhi syarat dari salah satu dari fungsi logika, yang mengakibatkan program berhenti seketika (*hang/error*). Sedangkan pada metode pohon keputusan *fuzzy*, ketika nilai yang hilang memasuki suatu titik percabangan, maka ia dapat memasuki ke semua ranting dengan prosentase yang sama. Selanjutnya tingkat prosentasi tertinggi pada tiap anak cabang akan diukur sehingga dapat ditentukan kelas klasifikasinya.

Aspek selanjutnya yang dianalisis yaitu waktu komputasi metode pohon keputusan *fuzzy* yang lebih lambat dibandingkan metode pohon keputusan untuk empat jenis pengujian di atas. Hal ini dapat dijelaskan antara lain seperti pada Gambar 7. Dengan semakin dalamnya aturan hirarki yang terbentuk pada pohon keputusan *fuzzy* membuat hasil klasifikasi semakin akurat, tetapi efek yang terjadi pada pembentukan aturan hirarki yang semakin dalam adalah terjadinya pembentukan anak cabang (*leaf*) yang jauh lebih banyak sehingga dapat mengurangi keefektifan yang mengakibatkan bertambahnya waktu komputasi.

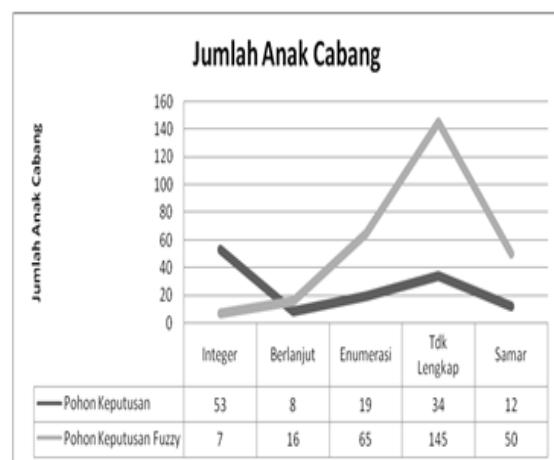
Untuk pengujian data tidak lengkap pada metode pohon keputusan, karena mengalami kegagalan seperti telah dijelaskan sebelumnya, maka tidak dapat dilakukan pengukuran waktu komputasi, ditunjukkan pada Gambar 8.

Jumlah anak cabang yang terbentuk mengalami puncaknya pada pengujian data tidak lengkap pada metode pohon keputusan *fuzzy*. Hal ini menandakan semakin sulitnya dan rumitnya aturan hirarki yang dibangun dalam menentukan kelas klasifikasi. Sedangkan pada pengujian data *integer*, jumlah anak cabang pada metode pohon keputusan *fuzzy* lebih sedikit dari metode pohon keputusan tetapi dengan tingkat akurasi yang lebih tinggi dan waktu komputasi yang lebih lama. Hal ini menandakan untuk data-data yang biasa, dalam hal ini adalah data *integer* dan data berlanjut dengan batasan tertentu (dengan maksimal tiga digit di belakang koma), sangat efektif digunakan metode pohon keputusan *fuzzy* dengan pembentukan anak cabang yang lebih sedikit tetapi efisien menghasilkan tingkat akurasi yang tinggi walaupun waktu komputasinya lebih lama.

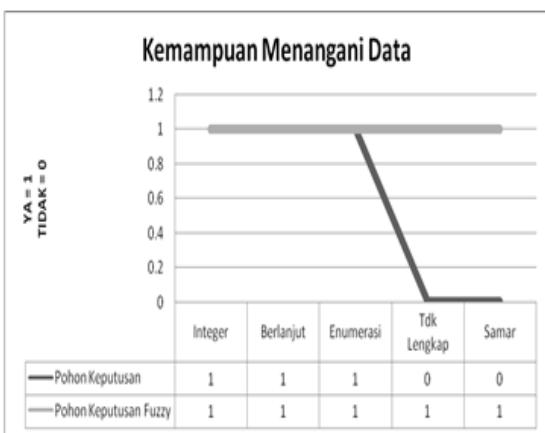
Algoritma metode pohon keputusan *fuzzy* menggunakan algoritma yang sama dengan metode pohon keputusan yaitu *ID3*. Tetapi dalam proses penyederhanaan pohon (*pruning*)

dalam pohon keputusan, metode pohon keputusan *fuzzy* memangkas banyak anak cabang terutama pada tingkat yang di bawah dikarenakan hasil perhitungan anak cabang yang berada di tingkat bawah sangat kecil prosentasenya dan sudah diwakili oleh anak cabang pada tingkat diatasnya dengan kelas yang sama seperti yang ditunjukkan pada Gambar 9.

Bila melihat dari proses pengolahan data pada masing-masing metode, akan dapat terlihat bahwa walaupun sebuah metode tersebut mampu memberikan hasil keluaran tetapi sebenarnya metode tersebut melakukan proses pengolahan yang salah. Kemampuan menangani data diberi nilai 1, dan bila tidak mampu diberi nilai 0. Dari Gambar 9, menunjukkan bahwa metode pohon keputusan tidak dapat menangani data tidak lengkap dan data samar. Pada pengujian data tidak lengkap, metode pohon keputusan mengalami kegagalan (*hang / error*). Sedangkan pada pengujian data samar, metode tersebut mampu memberikan hasil keluaran tetapi dengan pendekatan yang tidak diinginkan, yaitu dengan menganggap data samar sebagai data enumerasi yang dalam proses internalnya dimasukkan ke dalam kelompok data *discrete*. Sehingga keluaran yang dihasilkan dinilai tidak valid, karena tidak memberikan pendekatan yang diinginkan yaitu sebagai data samar.



Gambar 8. Grafik Jumlah Anak Cabang.



Gambar 9. Grafik Kemampuan Menangani Data.

SIMPULAN DAN SARAN

Berdasarkan hasil ujicoba dan analisisnya dapat diambil empat simpulan hasil penelitian. Satu, metode pohon keputusan *fuzzy* lebih baik

dibandingkan dengan metode pohon keputusan pada lima jenis pengujian, yaitu pengujian data integer, data berlanjut, data enumerasi, data tidak lengkap, dan data samar. Dua, penerapan konsep *fuzzy* pada metode pohon keputusan *fuzzy* membuat pembentukan aturan hirarki yang semakin dalam, dan meningkatkan tingkat akurasi. Namun konsep ini menyebabkan terjadi penambahan jumlah anak cabang yang tajam dan waktu komputasi yang lebih lama. Tiga, metode pohon keputusan tidak dapat mengenali data tidak lengkap. Sedangkan ketika mengenali data samar, menganggap data samar sebagai data enumerasi. Empat, berdasarkan hasil ujicoba terhadap seluruh pengujian dapat disimpulkan bahwa tingkat akurasi metode pohon keputusan *fuzzy* lebih baik 4,96 % dari metode pohon keputusan pada klasifikasi penutup lahan, tetapi memiliki waktu komputasi lebih lama, 38,13 %. Oleh karena itu diperlukan penelitian lebih lanjut untuk mengurangi waktu komputasi yang dibutuhkan pada metode pohon keputusan *fuzzy*.

DAFTAR PUSTAKA

- [1] Huang X, and Jensen J, A Machine Learning Approach to Automated Knowledge-base Building Remote Sensing Image Analysis With Gis Data, *Photogrammetric engineering and remote sensing Journal*, 63(10) : 1185-1194, 1998.
- [2] Wei W, Zhang X, Chen X, Tang J, and Jiang M, Wetland Mapping Using Subpixel Analysis And Decision Tree Classification In The Yellow River Delta Area, *The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Science*, 37: Part B7, Beijing, pp. 667-670, 2008.
- [3] Kotsiantis S, Zaharakis I, and Pintelas P, Machine learning: a review of classification and combining technique, Springer Netherlands, *Informatica Journal*, 31 : 249-268, 2007.
- [4] Basuki A, dan Syarif I, *Decision Tree*, 2003, URL: <http://lecturer.eepis.its.edu/~basuki/lecture/DecisionTree.pdf>, diakses 16 Maret 2008.
- [5] Quinlan JR, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993, or via links at <http://www.rulequest.com/Personal/>, diakses 16 Maret 2008.
- [6] Peng Y, and Flach P, Soft Discretization to Enhance the Continuous Decision Tree Induction. *Integrating Aspects of Data Mining, Decision Support and Meta-Learning*. Christophe Giraud-Carrier, Nada Lavrac, Steve Moyle, (eds.), pp. 109–118, September 2001.
- [7] Techterms.com, *Integer*, URL: <http://www.techterms.com/definition/integer>, diakses 10 Februari 2010.
- [8] Godam64, *Klasifikasi jenis dan macam data, pembagian data dalam ilmu eksak sains statistik statitiska*, URL:http://organisasi.org/klasifikasi_jenis_dan_macam_data_pembagian_data_dalam_ilmu_eksak_sains_statistik_statitiska, diakses 12 Februari 2010.
- [9] Permana I, *Mempelajari matematika diskrit*, URL: <http://kipsaint.com/isi/>

- mempelajari-matematika-diskrit.html, diakses 12 Februari 2010.
- [10] The enum type, *MySQL 5 Reference Manual*, 1997, URL: <http://dev.mysql.com/doc/refman/5.0/en/enum.html>, diakses 12 Februari 2010.
- [11] Unidata Program Center, *Writing NetCDF: Best Practise, Missing data values*, URL: <http://www.unidata.ucar.edu/software/netcdf/docs/BestPractices.html>, diakses 12 Februari 2010.
- [12] AdèrHJ, and Mellenbergh GJ, *Advising on Research Methods: A consultant's companion*, Huizen, The Netherlands: Johannes van Kessel Publishing,pp. 305-332, 2008.
- [13] Schneider, M., *Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases* Logika, 2008, URL: <http://www.cise.ufl.edu/~mschneid/Research/papers/Sch08BoCh.pdf>, diakses 12 Februari 2010.