

## Implementasi *Fuzzy Decision Tree* Untuk Prediksi Gagal Ginjal Kronis

Fitri Sofia Nur Khamidah<sup>1</sup>, Dian Hapsari<sup>2</sup>, Hendro Nugroho<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika ITATS, Jl. Arief Rahman Hakim No. 100 Surabaya

Email : fitrisofia96@gmail.com

**ABSTRACT.** *Fuzzy Decision Tree Implementation for Predicting Chronic Renal Failure.* Kidney is one of the important organs for the body. The main function of kidney is for filtering process. The gradual decreasing of kidney function will lead to kidney disease and if it is left unchecked, it will lead to chronic renal failure. Chronic renal failure is a type of disease that can cause death. Until now there is no antidote for the disease of chronic renal failure, therefore this disease cannot be cured but its development can be slowed or stopped. The early diagnosis of this disease will help to prevent the fatal consequences. To diagnose the disease requires some laboratory tests in which the results of the test will be calculated and summed up by a doctor or medical practitioner. The development of science and technology, especially in the field of computers will help the doctor's works in analyzing the results of laboratory test easier and faster. By some data as training data and implementing Fuzzy Decision Tree classification algorithm, it is expected to obtain high accuracy results that can be used as a reference for predicting chronic renal failure and avoid the occurrence of fatal consequences. The test was conducted by using some predetermined threshold and obtained the most optimal accuracy 98.28% with which indicated a fairly high level of accuracy. Thus the Fuzzy Decision Tree algorithm can be said to be able to predict the disease of chronic renal failure by the accuracy 98.28%.

**Keywords:** *Chronic Renal Failure, Classification, Fuzzy Decision Tree.*

**ABSTRAK.** *Ginjal adalah salah satu organ penting bagi tubuh. Fungsi utama dari ginjal adalah untuk proses penyaringan. Penurunan fungsi ginjal secara bertahap akan menyebabkan penyakit ginjal dan jika dibiarkan akan menjadi gagal ginjal kronis. Gagal ginjal kronis merupakan salah satu jenis penyakit yang dapat menyebabkan kematian. Sampai saat ini penyakit gagal ginjal kronis tidak ada obat penawarnya sehingga penyakit ini tidak dapat disembuhkan namun dapat diperlambat atau dihentikan perkembangannya. Diagnosa awal penyakit ini akan membantu untuk mencegah akibat fatal tersebut. Untuk mendiagnosa penyakit ini diperlukan beberapa tes laboratorium yang mana hasil dari tes tersebut akan dihitung dan disimpulkan hasilnya oleh dokter atau praktisi medis. Perkembangan ilmu pengetahuan dan teknologi terutama dalam bidang komputer akan membantu pekerjaan dokter untuk menganalisa hasil tes laboratorium menjadi semakin mudah dan cepat. Dengan beberapa data sebagai data training dan mengimplementasikan algoritma klasifikasi Fuzzy Decision Tree diharapkan dapat memperoleh hasil akurasi yang tinggi sehingga dapat dijadikan untuk acuan prediksi gagal ginjal kronis dan menghindari akibat fatal yang akan terjadi. Uji coba dilakukan dengan menggunakan beberapa threshold yang telah ditentukan dan didapatkan hasil akurasi paling optimal sebesar 98.28% yang menunjukkan tingkat akurasi yang cukup tinggi. Dengan demikian algoritma Fuzzy Decision Tree dapat dikatakan mampu untuk memprediksi penyakit gagal ginjal kronis dengan tingkat akurasi sebesar 98.28%.*

**Kata kunci:** *Gagal Ginjal Kronis, Klasifikasi, Fuzzy Decision Tree.*

## 1. Pendahuluan

### 1.1. Latar Belakang

Ginjal adalah salah satu organ penting bagi tubuh. Fungsi utama dari ginjal adalah untuk proses penyaringan. Proses penyaringan ini sangat berpengaruh pada keseimbangan dan kondisi tubuh karena saat ada suatu zat yang seharusnya tidak ikut tersaring, akan tetapi karena menurunnya fungsi ginjal zat tersebut jadi lolos tersaring. Hal ini mengakibatkan tubuh akan mengalami kerusakan sedikit demi sedikit. Peristiwa menurunnya fungsi ginjal ini akan menyebabkan terjadinya penyakit ginjal yang disebut gagal ginjal. Penyakit gagal ginjal jika tidak segera ditangani dan sudah lama berada di tubuh maka akan menjadi gagal ginjal kronis.

Gagal ginjal kronis adalah penyakit yang diderita oleh seseorang yang ditandai dengan menurunnya fungsi ginjal secara bertahap. Gagal ginjal kronis merupakan penyakit yang tidak ada obat yang dapat menyembuhkannya sehingga penyakit ini tidak dapat disembuhkan namun dapat diperlambat atau dihentikan perkembangannya dan dapat dicegah supaya tidak muncul kondisi serius lain yang dapat membahayakan pasien. Mayoritas penderita penyakit ini mengetahui bahwa dirinya terdiagnosis gagal ginjal setelah memasuki stadium akhir, sehingga menyebabkan susah untuk ditangani. Hal ini disebabkan karena penyakit gagal ginjal kronis tidak memiliki gejala yang nampak. Namun jika seseorang melakukan pengecekan urin dan darah secara teratur maka dapat diketahui lebih awal.

Untuk mengetahui seseorang menderita gagal ginjal atau tidak diperlukan beberapa tes laboratorium yang mana hasil dari tes tersebut akan dihitung oleh dokter dan akan disimpulkan hasilnya apakah positif gagal ginjal atau negatif gagal ginjal. Perhitungan manual yang dilakukan oleh seorang dokter atau praktisi medis tersebut membutuhkan ketelitian ekstra karena penyakit ini memiliki banyak atribut yang harus dihitung sehingga membutuhkan waktu yang cukup lama. Oleh karena itu, dokter dan praktisi medis sekarang ini beralih pada komputer untuk diagnosis suatu penyakit.

Perkembangan dari teknologi komputer membuat pekerjaan diagnosis atau pengambilan keputusan menjadi lebih mudah. Di dalam *machine learning* memungkinkan sebuah komputer mampu untuk memutuskan atau memberikan saran yang tepat, dimana terdapat banyak pendekatan dari teknik pengambilan keputusan, seperti *decision tree*, *artificial neural networks* dan *Bayesian learning* (Liang, 2005).

Pada penelitian ini teknik pengambilan keputusan yang akan digunakan adalah *decision tree* karena dapat mengklasifikasikan data baik kategoris maupun numerik. Selain itu konsep dari *decision tree* juga mudah untuk dipahami. Namun *decision tree* juga memiliki kelemahan yaitu ketidakstabilan sehubungan dengan gangguan kecil pada data pelatihan. Hal ini akan berpengaruh pada struktur dari *decision tree* yang jika terdapat perubahan sedikit dalam dataset, maka akan berubah pula struktur pohonnya. Untuk mengatasi masalah ini dikembangkanlah *Fuzzy Decision Tree* dengan memanfaatkan teori himpunan *fuzzy* untuk menggambarkan tingkat keterhubungan suatu atribut (Liang, 2005).

*Fuzzy ID3 (Iterative Dichotomiser 3)* merupakan perluasan dari algoritma ID3 dengan memanfaatkan *fuzzy entropy* untuk pemilihan atribut. Namun hasil dari *Fuzzy ID3* ini tidak memiliki akurasi yang tinggi. Oleh karena itu digunakan dua parameter penting yaitu *fuzziness control threshold* ( $\theta_r$ ) dan *leaf decision threshold* ( $\theta_n$ ). Contoh dari penggunaannya adalah pada penelitian yang dilakukan oleh Romansyah, Sitanggang, & Nurdianti pada tahun 2009 dengan tema prediksi diabetes menggunakan *Fuzzy Decision Tree*. Hasil akurasi yang diperoleh cukup tinggi mencapai 94,15% dengan *fuzziness control threshold* ( $\theta_r$ ) = 75% dan *leaf decision threshold* ( $\theta_n$ ) = 8% atau 10%.

### 1.2. Perumusan Masalah

Dari ulasan latar belakang di atas dapat diketahui bahwa permasalahan yang ada yaitu mengimplementasikan *Fuzzy Decision Tree* untuk memprediksi gagal ginjal kronis pada pasien.

### 1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membantu dokter atau praktisi medis dalam memprediksi penyakit gagal ginjal kronis pada pasien.

### 1.4. Manfaat Penelitian

Sedangkan manfaat dari penelitian ini adalah supaya dapat dilakukan tindakan pencegahan dalam bentuk pengobatan untuk menghambat atau menghentikan perkembangan dari penyakit karena ketika salah memprediksi dapat berakibat fatal yaitu kematian.

## 2. Tinjauan Pustaka

### 2.1. Fuzzy Decision Tree dengan Algoritma ID3

Algoritma ID3 pertama kali diperkenalkan oleh Quinlan, menggunakan teori informasi untuk menentukan atribut mana yang paling informatif, namun ID3 sangat tidak stabil dalam melakukan penggolongan berkenaan dengan gangguan kecil pada data latihan. Logika *fuzzy* dapat memberikan suatu peningkatan dalam melakukan penggolongan pada saat pelatihan (Romansyah, Sitanggang, & Nurdianti, 2009).

Algoritma *Fuzzy Decision Tree* merupakan pengembangan dari *decision tree*. *Fuzzy Decision Tree* merupakan penggabungan dari dua metode, dimana *fuzzy* digunakan untuk mengoptimalkan atribut-atribut pada *decision tree*, kemudian digunakan sebagai atribut untuk pengambilan keputusan oleh *decision tree*.

*Fuzzy Decision Tree* memungkinkan untuk menggunakan nilai-nilai *numeric-symbolic* selama konstruksi atau saat mengklasifikasikan kasus-kasus baru. Manfaat dari teori himpunan *fuzzy* dalam *decision tree* ialah meningkatkan kemampuan dalam memahami *decision tree* ketika menggunakan atribut-atribut kuantitatif. Bahkan, dengan menggunakan teknik *fuzzy* dapat meningkatkan ketahanan saat melakukan klasifikasi kasus-kasus baru (Romansyah, Sitanggang, & Nurdianti, 2009).

#### 2.1.1. Fuzzy Entropy dan Information Gain

Dengan menggunakan teknik *fuzzy*, maka akan dilakukan penyesuaian terhadap data. Pada himpunan data *fuzzy*, terdapat penyesuaian rumus untuk menghitung nilai *entropy* untuk atribut dan *information gain* karena adanya ekspresi data *fuzzy*. Untuk menentukan *fuzzy entropy* dan *information gain* dari suatu atribut digunakan persamaan sebagai berikut (Romansyah, Sitanggang, & Nurdianti, 2009):

$$H_f(S, A) = - \sum_{i=1}^C \frac{\sum_j \mu_{ij}}{s} \log_2 \frac{\sum_j \mu_{ij}}{s} \quad (1)$$

$$G_f(S) = H_f(S) - \sum_{v \in A} \frac{|S_v|}{|S|} * H_f(S_v, A) \quad (2)$$

dengan  $\mu_{ij}$  adalah nilai keanggotaan dari pola ke- $j$  untuk kelas ke- $i$ .  $H_f(S)$  menunjukkan *entropy* dari himpunan  $S$  dari data pelatihan pada *node*.  $|S_v|$  adalah ukuran dari subset  $S_v \subseteq S$  dari data pelatihan  $x_j$  dengan atribut  $v$ .  $|S|$  menunjukkan ukuran dari himpunan  $S$ .

#### 2.1.2. Threshold dalam Fuzzy Decision Tree

Kelemahan dalam *Fuzzy Decision Tree* adalah jika proses *learning* atau pembuatan pohon dihentikan sampai semua data contoh menjadi anggota sebuah kelas, maka akan dihasilkan akurasi yang rendah. Oleh karena itu untuk meningkatkan akurasinya proses *learning* tersebut harus dihentikan lebih awal atau melakukan pemotongan *tree* secara umum. Untuk itu diberikan 2 (dua)

*threshold* yang harus terpenuhi jika *tree* akan diekspansi, yaitu (Romansyah, Sitanggang, & Nurdiati, 2009). (1) *Fuzziness control threshold* (FCT) /  $\theta_r$ , jika proporsi dari himpunan data dari suatu kelas lebih besar atau sama dengan nilai *threshold*  $\theta_r$ , maka hentikan ekspansi *tree*. (2) *Leaf decision threshold* (LDT) /  $\theta_n$ , jika banyaknya anggota himpunan data pada suatu *node* lebih kecil dari *threshold*  $\theta_n$ , hentikan ekspansi *tree*.

Nilai *threshold* ini memiliki pengaruh besar pada hasil *tree* yang akan terbentuk. Oleh karena itu perlu dilakukan percobaan pada nilai berbeda-beda untuk menemukan nilai yang optimal (Liang, 2005).

## 2.2. Akurasi

Untuk mengetahui ketepatan dalam memprediksi digunakan perhitungan akurasi dengan metode *Confusion Matrix*. Rumus dari akurasi sebagai berikut:

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

Keterangan:

- TP (*True Positive*) : nilai prediksi sistem positif, nilai sebenarnya positif  
 TN (*True Negative*) : nilai prediksi sistem negatif, nilai sebenarnya negatif  
 FP (*False Positive*) : nilai prediksi sistem positif, nilai sebenarnya negatif  
 FN (*False Negative*) : nilai prediksi sistem negatif, nilai sebenarnya positif

## 3. Metode Penelitian

### 3.1. Rancangan Sistem

Prediksi gagal ginjal kronis merupakan suatu proses untuk mengetahui apakah seseorang menderita penyakit gagal ginjal kronis atau tidak. Keakuratan dalam prediksi ini sangat penting karena jika seseorang salah memprediksi maka akan berakibat fatal untuk pasien. Untuk memprediksi gagal ginjal kronis dibutuhkan beberapa atribut yang didapat dari hasil laboratorium seseorang. Data dari hasil laboratorium tersebut pada kenyataannya banyak terdapat nilai yang tidak lengkap dalam data. Kondisi ini akan mempengaruhi hasil dari klasifikasi. Oleh karena itu dibutuhkan penanganan untuk mengatasi kondisi tersebut.

Setelah melewati tahap *preprocessing*, yaitu pembersihan data yang tidak ada nilainya, selanjutnya data akan ditransformasi ke dalam bentuk data *fuzzy*. Data yang sudah ditransformasi akan dihitung menggunakan algoritma *fuzzy decision tree* dan akan diketahui hasil prediksinya.

Secara umum alur dari penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Alur penelitian

### 3.2. Identifikasi Atribut

Penelitian ini menggunakan data set *Chronic Kidney Disease (CKD)* yang diambil dari *Repository UCI Machine Learning*. Data dikumpulkan dari rumah sakit selama kurang lebih 2 bulan. Data set ini mencakup total 400 sampel dengan atribut numerik berjumlah 11 kolom dan nominal berjumlah 14 kolom. Sampel data disediakan sebanyak 400 baris dengan 250 sampel adalah kelompok ckd (positif gagal ginjal kronis) dan 150 sampel kelompok notckd (negatif gagal ginjal kronis). Namun setelah melalui tahap *preprocessing* data yang bisa dipakai berjumlah 158 baris dengan 43 sampel adalah kelompok ckd (positif gagal ginjal kronis) dan 115 sampel kelompok notckd (negatif gagal ginjal kronis).

### 3.3. Transformasi Data

Berikut adalah contoh 3 atribut yang telah ditransformasi ke dalam bentuk *fuzzy*:

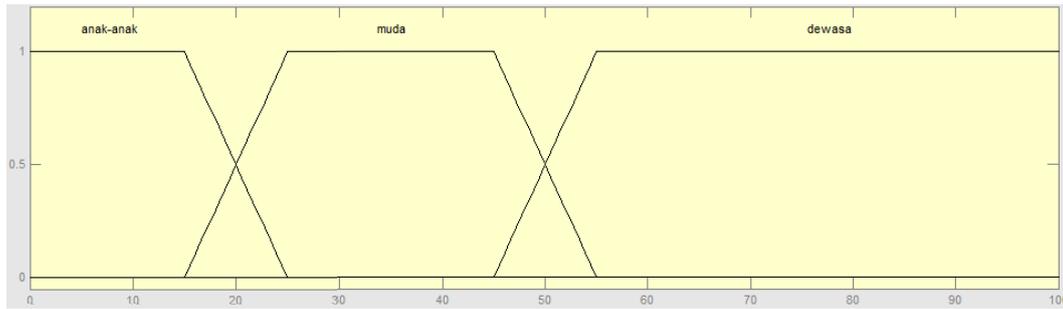
#### 3.3.1. Atribut age

Atribut age dibagi menjadi 3 kelompok yaitu anak ( $x < 20$ ), muda ( $20 \leq x < 50$ ), dewasa ( $x \geq 50$ ). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* anak, muda, dewasa untuk atribut age secara terpisah yaitu:

$$\mu_{anak-anak}(x) = \begin{cases} 1 & ; x \leq 15 \\ \frac{25-x}{10} & ; 15 < x < 25 \\ 0 & ; x \geq 25 \end{cases}$$

$$\mu_{dewasa}(x) = \begin{cases} 0 & ; x \leq 45 \\ \frac{x-45}{10} & ; 45 < x < 55 \\ 1 & ; x \geq 55 \end{cases}$$

$$\mu_{muda}(x) = \begin{cases} 0 & ; x \leq 15 \\ \frac{x-15}{10} & ; 15 < x < 25 \\ 1 & ; 25 \leq x \leq 45 \\ \frac{55-x}{10} & ; 45 < x \leq 55 \\ 0 & ; x \geq 55 \end{cases}$$



Gambar 2 Representasi grafik fungsi keanggotaan atribut age

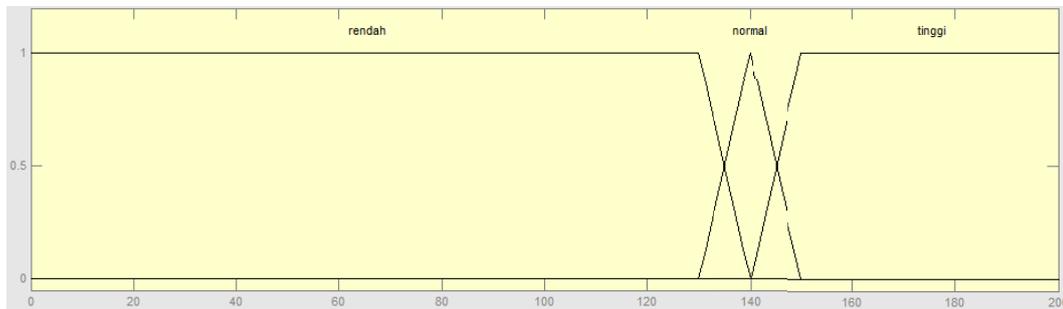
3.3.2. Atribut sod

Atribut sod dibagi menjadi 3 kelompok yaitu rendah ( $sod < 135$ ), normal ( $135 \leq sod < 145$ ), tinggi ( $sod \geq 145$ ). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* rendah, normal, tinggi untuk atribut sod secara terpisah yaitu:

$$\mu_{rendah}(x) = \begin{cases} 1 & ; x \leq 130 \\ \frac{140 - x}{10} & ; 130 < x < 140 \\ 0 & ; x \geq 140 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x \leq 140 \\ \frac{x - 140}{10} & ; 140 < x < 150 \\ 1 & ; x \geq 150 \end{cases}$$

$$\mu_{normal}(x) = \begin{cases} 0 & ; x \leq 130 \\ \frac{x - 130}{10} & ; 130 < x < 140 \\ 1 & ; 140 \leq x \leq 140 \\ \frac{150 - x}{10} & ; 140 < x \leq 150 \\ 0 & ; x \geq 150 \end{cases}$$



Gambar 3 Representasi grafik fungsi keanggotaan atribut sod

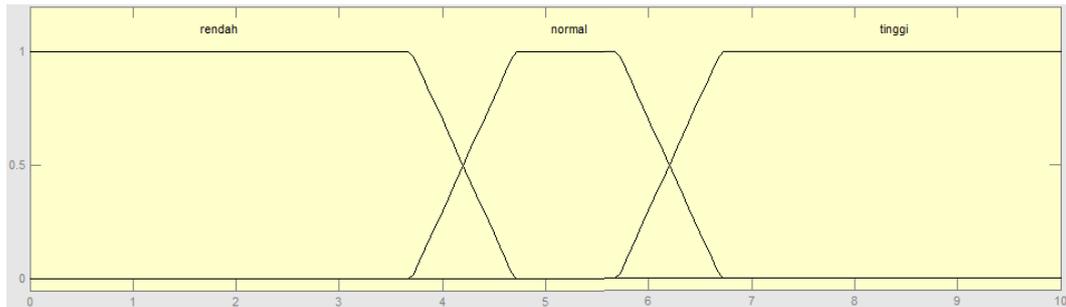
3.3.3. Atribut rbcc

Atribut rbcc dibagi menjadi 3 kelompok yaitu rendah ( $rbcc < 4.2$ ), normal ( $4.2 \leq rbcc < 6.2$ ), tinggi ( $rbcc \geq 6.2$ ). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* rendah, normal, tinggi untuk atribut rbcc secara terpisah yaitu:

$$\mu_{rendah}(x) = \begin{cases} 1 & ; x \leq 3.7 \\ \frac{4.7 - x}{1} & ; 3.7 < x < 4.7 \\ 0 & ; x \geq 4.7 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x \leq 5.7 \\ \frac{x - 5.7}{1} & ; 5.7 < x < 6.7 \\ 1 & ; x \geq 6.7 \end{cases}$$

$$\mu_{normal}(x) = \begin{cases} 0 & ; x \leq 3.7 \\ \frac{x - 3.7}{1} & ; 3.7 < x < 4.7 \\ 1 & ; 4.7 \leq x \leq 5.7 \\ \frac{6.7 - x}{1} & ; 5.7 < x \leq 6.7 \\ 0 & ; x \geq 6.7 \end{cases}$$



Gambar 4 Representasi grafik fungsi keanggotaan atribut rbcc

### 3.4. Penyusunan *Fuzzy Decision Tree*

Setelah data ditransformasi ke dalam bentuk data *fuzzy*, proses selanjutnya adalah penyusunan *fuzzy decision tree*. Langkah-langkahnya diuraikan sebagai berikut (Romansyah, Sitanggang, & Nurdianti, 2009): (1) Menentukan atribut yang akan digunakan. (2) Menentukan banyaknya *fuzzy set* untuk masing-masing atribut. (3) Mencari nilai *membership function* dari setiap data. (4) Menghitung *fuzzy entropy* dan *information gain*. (5) Memilih besarnya *threshold* yang akan digunakan. (6) Membangun *fuzzy decision tree* dengan algoritma *Fuzzy ID3*.

### 3.5. Skenario Pengujian

Dalam penelitian ini, akan dilakukan skenario pengujian untuk menghitung akurasi dalam hasil prediksi. Pada mulanya dataset dibagi menjadi 2 bagian, yang pertama adalah data *training* dan kedua adalah data *testing*. Data *training* diproses untuk menghasilkan pohon keputusan dan aturan-aturan keputusan. Setelah pohon berhasil terbentuk langkah selanjutnya adalah pemrosesan data *testing* dimana dari hasil prediksi tersebut akan dicocokkan dengan data aktualnya untuk mengetahui tingkat akurasi dari metode yang diusulkan.

## 4. Analisis dan Hasil Penelitian

### 4.1. Pengujian Sistem

Pengujian sistem pada klasifikasi menggunakan algoritma *Fuzzy Decision Tree* ini meliputi pengujian proses pembentukan himpunan *fuzzy*, pembentukan *tree*, dan perhitungan akurasi. Pengujian ini dilakukan dengan menggunakan data sampel sebanyak 158 data dengan proporsi data *training* sejumlah 100 data dan data *testing* sejumlah 58 data.

Dari data yang tersedia akan dilakukan pengujian untuk mengetahui pengaruh nilai *threshold* terhadap jumlah aturan yang dihasilkan dan besarnya akurasi. Ada dua *threshold* yang digunakan dalam pengujian ini yaitu *Fuzziness Control Threshold* (FCT) atau  $\theta_r$  dan *Leaf Decision Threshold* (LDT) atau  $\theta_n$ . Untuk nilai dari  $\theta_r$  yang digunakan sebanyak 14 yaitu mulai dari 85% sampai dengan 98%. Sedangkan untuk nilai  $\theta_n$  digunakan 6 yaitu 3%, 5%, 8%, 10%, 15%, dan 20%. Pemilihan nilai  $\theta_r$  dan  $\theta_n$  tersebut diambil berdasarkan acuan dari penelitian yang telah dilakukan oleh Romansyah, Sitanggang, & Nurdianti pada tahun 2009 yang kemudian disesuaikan dengan penelitian yang akan dilakukan. Tabel 1 menunjukkan hasil banyaknya jumlah aturan yang diperoleh dari setiap pengujian.

Tabel 1 Rata-rata jumlah aturan

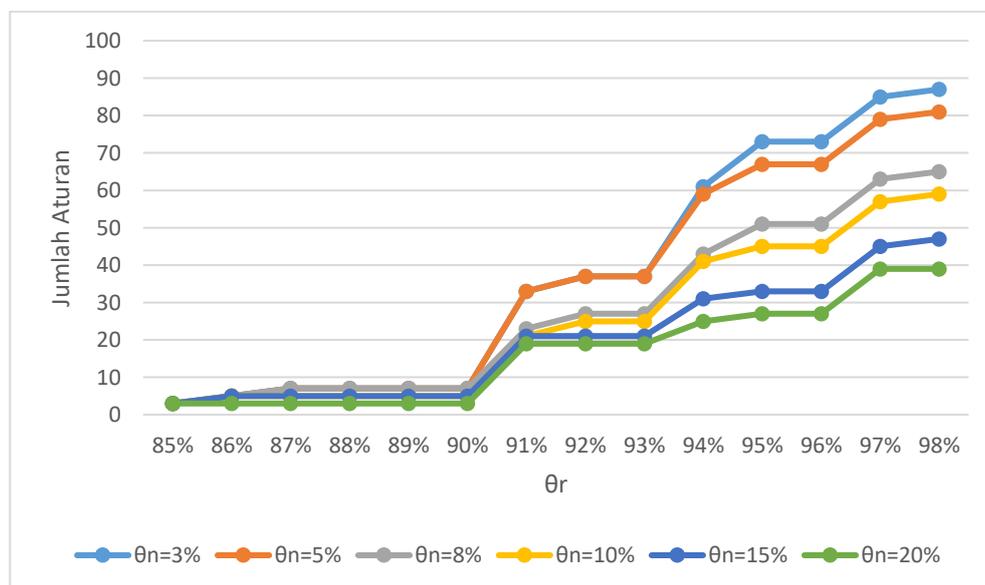
<i>Fuzziness Control Threshold</i> (FCT)	<i>Leaf Decision Threshold</i> (LDT)					
	3%	5%	8%	10%	15%	20%
85%	3	3	3	3	3	3
86%	5	5	5	5	5	3
87%	7	7	7	5	5	3
88%	7	7	7	5	5	3
89%	7	7	7	5	5	3
90%	7	7	7	5	5	3

91%	33	33	23	21	21	19
92%	37	37	27	25	21	19
93%	37	37	27	25	21	19
94%	61	59	43	41	31	25
95%	73	67	51	45	33	27
96%	73	67	51	45	33	27
97%	85	79	63	57	45	39
98%	87	81	65	59	47	39

Nilai-nilai  $\theta_r$  dan  $\theta_n$  yang digunakan pada pengujian ini dipilih berdasarkan hasil percobaan, karena dengan nilai-nilai tersebut terjadi perubahan yang cukup signifikan pada rata-rata jumlah aturan yang dihasilkan. Pengambilan nilai  $\theta_r$  dimulai dari 85% dikarenakan nilai-nilai di bawah 85% seperti 84%, 83%, 82%, dan seterusnya memiliki hasil yang tidak terjadi perubahan. Begitu juga dengan pemilihan nilai  $\theta_n$ .

Berdasarkan hasil pengujian pada Tabel 1 dapat diketahui bahwa semakin tinggi nilai  $\theta_r$  maka semakin banyak pula jumlah aturan yang dihasilkan. Hal ini terjadi karena sebelum suatu node didominasi oleh sebuah kelas dan proporsi dari kelas tersebut di atas atau sama dengan nilai  $\theta_r$ -nya maka *tree* akan terus diekspansi. Begitu juga dengan nilai  $\theta_n$  yang terlalu rendah juga akan menghasilkan aturan dengan jumlah yang banyak. Jadi untuk nilai  $\theta_r$  yang terlalu tinggi dan atau  $\theta_n$  yang terlalu rendah akan menyebabkan *Fuzzy Decision Tree* berperilaku seperti *Decision Tree* biasa yang tidak memerlukan adanya *threshold* sehingga menghasilkan *tree* dengan ukuran yang besar dan jumlah aturan yang banyak dikarenakan *tree* akan terus diekspansi sampai *leaf node* terdalam. Dapat dilihat pada Tabel 2 di atas jumlah aturan terbesar ditandai dengan warna hijau.

Sebaliknya, nilai  $\theta_r$  yang terlalu rendah dan atau  $\theta_n$  yang terlalu tinggi akan menghasilkan *tree* dengan ukuran yang kecil sehingga jumlah aturan yang dihasilkan juga sedikit. Hal ini terjadi karena *tree* yang sedang dibangun mengalami pemotongan karena sudah memenuhi *threshold* yang telah ditentukan. Dapat dilihat pada Tabel 2 di atas jumlah aturan yang memiliki nilai rendah ditandai dengan warna kuning. Dalam bentuk grafik ditunjukkan pada gambar 5.



Gambar 5 Perbandingan rata-rata jumlah aturan

Dari proses pengujian yang telah dilakukan dengan menggunakan beberapa *threshold* dapat diketahui bahwa jumlah aturan maksimal adalah 87 aturan dengan nilai  $\theta_r = 98\%$  dan  $\theta_n = 3\%$ . Sedangkan untuk jumlah aturan paling rendah adalah 3 aturan yaitu pada *range* nilai untuk  $\theta_r = 85\%$ , semua nilai  $\theta_n$  menghasilkan 3 aturan dan untuk nilai  $\theta_n = 20\%$ , nilai  $\theta_r$  pada *range* 85% – 90%. Berikut contoh potongan aturan yang dihasilkan.

Untuk jumlah aturan paling rendah yaitu 3 aturan :

IF RBCC rendah THEN positif ckd.

IF RBCC normal THEN negatif ckd.

IF RBCC tinggi THEN negatif ckd.

Dari beberapa percobaan yang telah dilakukan dengan menggunakan nilai dari  $\theta_r$  sebanyak 14 yaitu mulai dari 85% sampai dengan 98%. Sedangkan untuk nilai  $\theta_n$  digunakan 6 yaitu 3%, 5%, 8%, 10%, 15%, dan 20% dihasilkan aturan yang berbeda-beda. Pada setiap aturan yang terbentuk dapat diketahui bahwa atribut yang digunakan ada 6 atribut yaitu *age*, *bp*, *sg*, *al*, *sod*, dan *rc*. Hal ini memungkinkan bahwa keenam atribut tersebut merupakan atribut yang memiliki pengaruh besar dalam prediksi gagal ginjal kronis.

## 4.2. Perhitungan Akurasi

Mengacu pada persamaan 3, didapatkan hasil perhitungan akurasi untuk  $\theta_r = 85\%$  dan  $\theta_n = 3\%$  sebagai berikut:

TP (*True Positive*) : 9; TN (*True Negative*) : 48; FP (*False Positive*) : 0; FN (*False Negative*) : 1

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{9 + 48}{9 + 0 + 1 + 48} = \frac{57}{58} = 98.28\%$$

Hasil dari perhitungan akurasi setiap percobaan dengan menggunakan *threshold* yang berbeda-beda ditunjukkan pada tabel 2.

Tabel 2 Akurasi Fuzzy Decision Tree

Fuzziness Control Threshold (FCT)	Leaf Decision Threshold (LDT)					
	3%	5%	8%	10%	15%	20%
85%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
86%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
87%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
88%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
89%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
90%	98.28%	98.28%	98.28%	98.28%	98.28%	98.28%
91%	98.28%	98.28%	93.10%	93.10%	93.10%	93.10%
92%	98.28%	98.28%	93.10%	93.10%	93.10%	93.10%
93%	98.28%	98.28%	93.10%	93.10%	93.10%	93.10%
94%	96.55%	96.55%	91.38%	91.38%	89.66%	89.66%
95%	96.55%	96.55%	91.38%	91.38%	89.66%	89.66%
96%	96.55%	96.55%	91.38%	91.38%	89.66%	89.66%
97%	94.83%	94.83%	89.66%	89.66%	87.93%	87.93%
98%	94.83%	94.83%	89.66%	89.66%	87.93%	87.93%

Dari tabel 2 dapat dilihat bahwa kinerja dari algoritma *Fuzzy Decision Tree* mengalami penurunan jika nilai dari  $\theta_r$  semakin tinggi dan  $\theta_n$  semakin rendah. Nilai akurasi terbaik yang didapatkan dalam penelitian ini yaitu 98.28%.

## 5. Penutup

### 5.1. Kesimpulan

Pada penelitian yang telah dilakukan dengan menggunakan algoritma *Fuzzy Decision Tree* untuk memprediksi penyakit gagal ginjal kronis dapat disimpulkan sebagai berikut: (1) Data gagal ginjal kronis yang didapatkan dari *Repository UCI Machine Learning* pada mulanya berjumlah 24 atribut dan 1 kelas. Setelah dilakukan beberapa pengujian dapat diketahui bahwa atribut yang paling mempengaruhi dalam penelitian ini sebanyak 6 atribut yaitu *age*, *bp*, *sg*, *al*, *sod*, dan *rc*. (2) Pengaruh nilai *threshold* di dalam algoritma *Fuzzy Decision Tree* menjadi acuan terhadap banyaknya jumlah

aturan yang terbentuk dan besarnya akurasi yang didapatkan. Semakin besar nilai  $\theta_r$  dan semakin kecil nilai  $\theta_n$  akan mengakibatkan jumlah aturan yang terbentuk semakin banyak, sebaliknya semakin kecil nilai  $\theta_r$  dan semakin besar nilai  $\theta_n$  maka aturan yang dihasilkan sangatlah sedikit. Dengan melakukan beberapa uji coba pada nilai *threshold* yang berbeda-beda didapatkan akurasi paling optimal sebesar 98.28%.

## 5.2. Saran

Dari penelitian yang telah dilakukan ini saran untuk perkembangan penelitian selanjutnya adalah sebagai berikut: (1) Perlu dilakukan uji coba dengan menggunakan data primer karena data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari *Repository UCI Machine Learning*. (2) Perlu dilakukan uji coba dengan tahap *preprocessing* untuk menentukan atribut-atribut yang paling berpengaruh karena atribut yang digunakan dalam penelitian ini berjumlah 24 atribut dan 1 kelas. *Preprocessing* ini bisa dilakukan dengan metode seperti PCA (*Principal Component Analysis*). (3) Mengacu pada penelitian yang telah dilakukan oleh Romansyah pada tahun 2009 dengan menggunakan algoritma yang sama yaitu *Fuzzy Decision Tree* dan diimplementasikan pada data yang berbeda, hasil akurasi pada kedua penelitian ini dapat dikatakan cukup bagus. Sehingga perlu dilakukan penelitian lebih lanjut untuk mempertimbangkan algoritma *Fuzzy Decision Tree* dengan objek yang lainnya. (4) Perlu dilakukan uji coba dengan menggunakan fungsi *fuzzy* yang lain seperti fungsi *sigmoid*, fungsi segitiga, dan fungsi berbentuk *bell* agar dapat diketahui pengaruh fungsi keanggotaan *fuzzy* terhadap hasil akurasi.

## Referensi

- Aggarwal, C. C. (2015). *Data Mining*. New York: Springer.
- Chronic\_Kidney\_Disease Data Set*. (n.d.). Retrieved from UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Chronic\\_Kidney\\_Disease](https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease)
- Hamidah, I. (2012). *Aplikasi Data Mining untuk Memprediksi Masa Studi Mahasiswa menggunakan Algoritma C4.5*. Bandung: Universitas Komputer Indonesia.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Amerika: Elsevier Inc.
- Hanik, U. (2011). Fuzzy Decision Tree dengan Algoritma C4.5 pada Data Diabetes Indian Pima. *Tugas Akhir Periode Januari 2011*, 1-8.
- Jameson, J. L., & Loscalzo, J. (2010). *Harrison Nefrologi dan Gangguan Asam-Basa*. Jakarta: Buku Kedokteran EGC.
- Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- Liang, G. (2005). *A comparative study of three Decision Tree algorithms: ID3, Fuzzy ID3 and Probabilistic Fuzzy ID3*. Netherlands: Erasmus University.
- Purnomo, B. B. (2011). *Dasar-dasar Urologi*. Jakarta: CV Sagung Seto.
- Romansyah, F., Sitanggang, I. S., & Nurdiati, S. (2009). Fuzzy Decision Tree dengan Algoritma ID3 pada Data Diabetes. *Internetworking Indonesia Journal*, 45-52.
- Siswanto. (2010). *Kecerdasan Tiruan*. Yogyakarta: Graha Ilmu.
- Suyanto. (2008). *Soft Computing Membangun Mesin Ber-IQ Tinggi*. Bandung: Informatika.