

Internetworking Indonesia Journal

The Indonesian Journal of ICT and Internet Development

Guest Editors' Introduction: Special Issue on Data Mining <i>by Anto Satriyo Nugroho & Moch Arif Bijaksana</i>	1
Enhanced SMART-TV: A Classifier with Vertical Data Structure and Dimensional Projections Pruning <i>by Taufik Fuadi Abidin & William Perrizo</i>	3
Using Business Intelligence Solutions for Achieving Organization's Strategy: Arab International University Case Study <i>by Mouhib Alnoukari</i>	11
Feature Selection for Large Scale Data by Combining Class Association Rule Mining and Information Gain: a Hybrid Approach <i>by Appavu alias Balamurugan, Pramala, Rajalakshmi & Rajaram</i>	17
Detecting the Originality of Extended Research Articles Using Similarity Techniques - A Proposal <i>by Shanmugasundaram Hariharan</i>	25
Prediksi Masa Studi Sarjana dengan Artificial Neural Network <i>by Muhamad Hanief Meinanda, Metri Annisa, Narendi Muhandri & Kadarsyah Suryadi</i>	31
Adaptive Content-based Navigation Generating System: Data Mining on Unorganized Multi Web Resources <i>by Diana Purwitasari, Yasuhisa Okazaki & Kenzi Watanabe</i>	37
Fuzzy Decision Tree dengan Algoritme ID3 pada Data Diabetes <i>by F. Romansyah, I. S. Sitanggang & S. Nurdiati</i>	45

Volume 1
Number 2
Fall 2009

ISSN: 1942-9703
IJ © 2009
www.InternetworkingIndonesia.org

Internetworking Indonesia Journal

The Indonesian Journal of ICT and Internet Development

ISSN: 1942-9703

Internetworking Indonesia is a semi-annual electronic journal devoted to the timely study of the Information and Communication Technology (ICT) and Internet development in Indonesia. The journal seeks high-quality manuscripts on the challenges and opportunities presented by information technology and the Internet in Indonesia.

Journal mailing address: Internetworking Indonesia Journal, PO Box 397110 MIT Station, Cambridge, MA 02139, USA.

Co-Editors

Thomas Hardjono, PhD
(MIT Kerberos Consortium, MIT, USA)

Budi Rahardjo, PhD
(ITB, Indonesia)

Kuncoro Wastuwibowo, MSc
(PT. Telkom, Indonesia)

Editorial Advisory Board

Prof. Edy Tri Baskoro, PhD (ITB, Indonesia)
Mark Baugher, MA (Cisco Systems, USA)
Lakshminath Dondeti, PhD (Qualcomm, USA)
Prof. Svein Knapskog, PhD (NTNU, Norway)
Prof. Merlyna Lim, PhD (Arizona State University, USA)
Prof. Bambang Parmanto, PhD (University of Pittsburgh, USA)

Prof. Wishnu Prasetya, PhD (Utrecht University, The Netherlands)
Prof. Jennifer Seberry, PhD (University of Wollongong, Australia)
Prof. Willy Susilo, PhD (University of Wollongong, Australia)
Prof. David Taniar, PhD (Monash University, Australia)

Focus & Scope: The Internetworking Indonesia Journal aims to become the foremost publication for practitioners, teachers, researchers and policy makers to share their knowledge and experience in the design, development, implementation, and the management of ICT and the Internet in Indonesia.

Topics of Interest: The journal welcomes and strongly encourages submissions based on interdisciplinary approaches focusing on ICT & Internet development and its related aspects in the Indonesian context. These include (but not limited to) information technology, communications technology, computer sciences, electrical engineering, and the broader social studies regarding ICT and Internet development in Indonesia. A list of topics can be found on the journal website at www.InternetworkingIndonesia.org.

Open Access Publication Policy: The Internetworking Indonesia Journal provides open access to all of its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. This follows the philosophy of the Open Journal Systems (see the Public Knowledge Project at pkp.sfu.ca). The journal will be published electronically and there are no subscription fees. Such access is associated with increased readership and increased citation of an author's work.

Manuscript Language: The Internetworking Indonesia Journal accepts and publishes papers in Bahasa Indonesia and English.

Manuscript Submission: Manuscripts should be submitted according to the IEEE Guide for authors, and will be refereed in the standard way. Manuscript pages should not exceed 7 pages of the IEEE 2-column format in a Microsoft-Word file format. Links to the IEEE style files can be found at www.InternetworkingIndonesia.org. Manuscripts submitted to Internetworking Indonesia must not have been previously published or committed to another publisher under a copyright transfer agreement, and must not be under consideration by another journal. Authors of accepted papers are responsible for the Camera Ready Copy (CRC) in the IEEE 2-column format (delivered to the Editor in a Microsoft-Word file). Authors are advised that no revisions of the manuscript can be made after acceptance by the Editor for publication. The benefits of this procedure are many, with speed and accuracy being the most obvious. We look forward to working with your electronic submission which will allow us to serve you more efficiently. Please email manuscripts or inquiries to the editors at the following address: editor@InternetworkingIndonesia.org.

Submission Guidelines: Internetworking Indonesia accepts a variety of manuscripts in either English or Bahasa Indonesia. Please review the descriptions below and identify the submission type best suited to your intended submission:

- **Research Papers:** Research papers report on results emanating from research projects, both theoretical and practical in nature.
- **Short papers:** Short research papers provide an introduction to new developments or advances regarding on-going work.
- **Policy Viewpoints:** Policy Viewpoints explore competing perspectives in the Indonesian policy debate that are informed by academic research.
- **Teaching Innovation:** Teaching Innovation papers explore creative uses of information technology tools and the Internet to improve learning and education in Indonesia.
- **Book Reviews:** A review of a book, or other book-length document, such as a government report or foundation report.

Fuzzy Decision Tree dengan Algoritme ID3 pada Data Diabetes

F. Romansyah, I. S. Sitanggang, S. Nurdiani

Abstract— Decision tree is one of widely used methods in developing classification models. In order to handle uncertainty, fuzzy approach is used. This work applied a classification technique using fuzzy decision tree method on diabetes dataset to obtain classification rules for predicting new data. Fuzzy ID3 (fuzzy Iterative Dichotomiser 3) was used to develop fuzzy decision tree with high accuracy. The result is a fuzzy decision tree with the highest accuracy 94,15% for fuzziness control threshold (θ_c) = 75% and leaf decision threshold (θ_l) = 8 % or 10%.

Index Terms— Fuzzy decision tree, Fuzzy ID3

I. PENDAHULUAN

Data mining merupakan proses ekstraksi informasi atau pola penting dalam basis data berukuran besar [3]. Pada penelitian ini akan diterapkan salah satu teknik dalam data mining, yaitu klasifikasi terhadap data diabetes. Data diabetes yang digunakan adalah data hasil pemeriksaan lab pasien dari sebuah rumah sakit yang meliputi pemeriksaan GLUN (Glukosa Darah Pusa), GPOST (Glukosa Darah 2 Jam PP), Tg (Trigliserida), HDL (Kolesterol HDL), serta diagnosa pasien berdasarkan nilai GLUN, GPOST, HDL, dan TG.

Klasifikasi merupakan salah satu metode dalam data mining untuk memprediksi label kelas dari suatu record dalam data. Metode yang digunakan dalam penelitian ini yaitu metode klasifikasi dengan fuzzy decision tree. Penggunaan teknik fuzzy memungkinkan melakukan prediksi suatu objek yang dimiliki oleh lebih dari satu kelas. Dengan menerapkan teknik data mining pada data diabetes diharapkan dapat ditemukan aturan klasifikasi yang dapat digunakan untuk memprediksi potensi seseorang terserang penyakit diabetes.

Naskah diterima pada 2 Desember 2009.

F. Romansyah adalah Associate Business Consultant (IT Consultant) di PT AGIT (Astra Graphia Information Technology) 22/F ANZ Tower Jl. Jend Sudirman Kavling 33A Jakarta 10220 (e-mail: fyro_mans@yahoo.com).

I. S. Sitanggang adalah staf pengajar di Departemen Ilmu Komputer FMIPA Institut Pertanian Bogor, Jl. Meranti, Wing 20 Level V, Kampus IPB Darmaga, Bogor 16680 - Indonesia (e-mail: imas.sitanggang@ipb.ac.id).

S. Nurdiani adalah staf pengajar di Departemen Ilmu Komputer FMIPA Institut Pertanian Bogor, Jl. Meranti, Wing 20 Level V, Kampus IPB Darmaga, Bogor 16680 - Indonesia (e-mail: nurdiati@ipb.ac.id)

Penelitian ini bertujuan untuk: 1) Menerapkan salah satu teknik klasifikasi yaitu Fuzzy ID3 (Iterative Dichotomiser 3) Decision Tree pada data hasil pemeriksaan lab pasien; 2) Menemukan aturan klasifikasi pada data diabetes yang menjelaskan dan membedakan kelas-kelas atau konsep sehingga dapat digunakan untuk memprediksi penyakit diabetes berdasarkan nilai dari atribut lain yang diketahui. Model yang dihasilkan pada penelitian ini diharapkan dapat digunakan oleh pihak yang berkepentingan untuk memprediksi potensi seseorang atau pasien terserang penyakit diabetes, sehingga terjadinya penyakit ini pada seseorang dapat diprediksi sedini mungkin dan dapat dilakukan tindakan antisipasi.

II. FUZZY DECISION TREE

A. Peubah Linguistik dan Linguistic Term

Peubah linguistik merupakan peubah verbal yang dapat digunakan untuk memodelkan pemikiran manusia yang diekspresikan dalam bentuk himpunan fuzzy. Peubah linguistik dikarakterisasi oleh quintuple $(x, T(x), X, G, M)$ dengan x adalah nama peubah, $T(x)$ adalah kumpulan dari linguistic term, G adalah aturan sintaks, M adalah aturan semantik yang bersesuaian dengan setiap nilai peubah linguistik. Sebagai contoh, jika umur diinterpretasikan sebagai peubah linguistik, maka himpunan dari linguistic term $T(\text{umur})$ menjadi :

$$T(\text{umur}) = \{\text{sangat muda, muda, tua}\}$$

Setiap term dalam $T(\text{umur})$ dikarakterisasi oleh himpunan fuzzy, X menunjukkan nilai interval x . Aturan semantik menunjukkan fungsi keanggotaan dari setiap nilai pada himpunan linguistic term [1].

Linguistic term didefinisikan sebagai kumpulan himpunan fuzzy yang didasarkan pada fungsi keanggotaan yang bersesuaian dengan peubah linguistik. Misalkan \mathcal{D} adalah kumpulan dari record yang terdiri dari himpunan atribut $I = \{I_1, \dots, I_n\}$, dengan $I_v, v = 1, \dots, n$. Atribut I dapat berupa atribut numerik atau kategorikal. Untuk setiap record d elemen \mathcal{D} , $d[I_v]$ menotasikan nilai i dalam record d untuk atribut I_v . Kumpulan linguistic term dapat didefinisikan pada seluruh domain dari atribut kuantitatif. $L_{vr}, r = 1, \dots, s_v$ menotasikan linguistic term yang berasosiasi dengan atribut I_v , sehingga himpunan fuzzy dapat didefinisikan untuk setiap L_{vr} . Himpunan fuzzy, $L_{vr}, r = 1, \dots, s_v$ didefinisikan sebagai:

$$L_{vr} = \begin{cases} \sum_{i_v \in \text{dom}(I_v)} \frac{\mu_{L_{vr}}(i_v)}{i_v} & \text{jika } I_v \text{ diskret} \\ \int_{\text{dom}(I_v)} \frac{\mu_{L_{vr}}(i_v)}{i_v} & \text{jika } I_v \text{ kontinu} \end{cases}$$

untuk semua $i_v \in \text{dom}(I_v)$, dengan $\text{dom}(I_v) = \{i_{v1}, \dots, i_{vn}\}$. Derajat keanggotaan dari nilai $i_v \in \text{dom}(I_v)$ dengan beberapa *linguistic term* L_{vr} dinotasikan oleh $\mu_{L_{vr}}$.

B. Fuzzy Decision Tree (FDT)

Decision tree merupakan suatu pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan permasalahan klasifikasi. Metode ini digunakan untuk memperkirakan nilai diskret dari fungsi target, yang mana fungsi pembelajaran direpresentasikan oleh sebuah *decision tree* [5]. *Decision tree* merupakan himpunan aturan IF...THEN. Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, di mana premis terdiri atas sekumpulan *node-node* yang ditemui, dan kesimpulan dari aturan terdiri atas kelas yang terhubung dengan *leaf* dari *path* [6].

Dalam pohon keputusan, *leaf node* diberikan sebuah label kelas. *Non-terminal node*, yang terdiri atas *root* dan *internal node* lainnya, mengandung kondisi-kondisi uji atribut untuk memisahkan *record* yang memiliki karakteristik yang berbeda. *Edge-edge* dapat dilabelkan dengan nilai-nilai *numeric-symbolic*. Sebuah atribut *numeric-symbolic* adalah sebuah atribut yang dapat bernilai *numeric* ataupun *symbolic* yang dihubungkan dengan sebuah variabel kuantitatif. Sebagai contoh, ukuran seseorang dapat dituliskan sebagai atribut *numeric-symbolic*: dengan nilai kuantitatif, dituliskan dengan "1,72 meter", ataupun sebagai nilai *numeric-symbolic* seperti "tinggi" yang berkaitan dengan suatu ukuran (*size*). Nilai-nilai seperti inilah yang menyebabkan perluasan dari *decision tree* menjadi *fuzzy decision tree* [8]. Penggunaan teknik *fuzzy* memungkinkan melakukan prediksi suatu objek yang dimiliki oleh lebih dari satu kelas.

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 [6].

C. Fuzzy ID3 Decision Tree

ID3 (*Iterative Dichotomiser 3*) merupakan salah satu algoritme yang banyak digunakan untuk membuat suatu *decision tree*. Algoritme ini 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 pelatihan. Logika *fuzzy* dapat memberikan suatu peningkatan untuk dalam melakukan penggolongan pada saat pelatihan [5].

Algoritme *fuzzy ID3* merupakan algoritme yang efisien untuk membuat suatu *fuzzy decision tree*. Algoritme *fuzzy ID3* adalah sebagai berikut [5]:

1. Create a Root node that has a set of fuzzy data with membership value 1.
2. If a node t with a fuzzy set of data D satisfies the following conditions, then it is a leaf node and assigned by the class name.
 - The proportion of class C_k is greater than or equal to θ_r ,

$$\frac{|D^{C_k}|}{|D|} \geq \theta_r$$
 - the number of a data set is less than θ_n
 - there are no attributes for more classifications
3. If a node D does not satisfy the above conditions, then it is not a leaf-node. And a new sub-node is generated as follow:
 - For A_i 's ($i=1, \dots, L$) calculate the information gain, and select the test attribute A_{max} that maximizes them.
 - Devide D into fuzzy subset D_1, \dots, D_m according to A_{max} , where the membership value of the data in D_j is the product of the membership value in D and the value of $F_{max,j}$ of the value of A_{max} in D .
 - Generate new node t_1, \dots, t_m for fuzzy subsets D_1, \dots, D_m and label the fuzzy sets $F_{max,j}$ to edges that connect between the nodes t_j and t .
 - Replace D by D_j ($j=1, 2, \dots, m$) and repeat from 2 recursively.

D. Fuzzy Entropy dan Information Gain

Information gain adalah suatu nilai statistik yang digunakan untuk memilih atribut yang akan mengekspansi *tree* dan menghasilkan *node* baru pada algoritme ID3. Suatu *entropy* dipergunakan untuk mendefinisikan nilai *information gain*. *Entropy* dirumuskan sebagai berikut [5]:

$$H_s(S) = \sum_{i=1}^N -P_i \cdot \log_2(P_i) \tag{1}$$

dengan P_i adalah rasio dari kelas C_i pada himpunan contoh $S = \{x_1, x_2, \dots, x_k\}$.

$$P_i = \frac{\sum_{j=1}^k x_j \in C_i}{S} \tag{2}$$

Information gain digunakan sebagai ukuran seleksi atribut, yang merupakan hasil pengurangan *entropy* dari himpunan contoh setelah membagi ukuran himpunan contoh dengan jumlah atributnya. *Information gain* didefinisikan sebagai berikut [5]:

$$G(S, A) = H(S) - \sum_{v \in \text{value}(A)} \frac{|S_v|}{|S|} H(S_v) \tag{3}$$

dengan bobot $W_i = \frac{|S_v|}{|S|}$ adalah rasio dari data dengan atribut v pada himpunan contoh.

Pada himpunan data *fuzzy*, terdapat penyesuaian rumus untuk menghitung nilai *entropy* untuk atribut dan *information*

gain karena adanya ekspresi data fuzzy. Berikut adalah persamaan untuk mencari nilai fuzzy entropy dari keseluruhan data [5]:

$$H_f(S) = H_r(S) = \sum_{i=1}^N -P_i \cdot \log_2(P_i) \quad (4)$$

Untuk menentukan fuzzy entropy dan information gain dari suatu atribut pada algoritme fuzzy ID3 (FID3) digunakan persamaan sebagai berikut [5]:

$$H_f(S, A) = -\sum_{i=1}^c \frac{\sum_j \mu_{ij}}{S} \log_2 \frac{\sum_j \mu_{ij}}{S} \quad (5)$$

$$G_f(S) = H_f(S) - \sum_{v \in A} \frac{|S_v|}{|S|} \cdot H_f(S_v, A) \quad (6)$$

dengan μ_j 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 .

E. Threshold dalam Fuzzy Decision Tree

Jika pada proses learning dari FDT dihentikan sampai semua data contoh pada masing-masing leaf-node menjadi anggota sebuah kelas, akan dihasilkan akurasi yang rendah. Oleh karena itu untuk meningkatkan akurasi, proses learning harus dihentikan lebih awal atau melakukan pemotongan tree secara umum. Untuk itu diberikan 2 (dua) threshold yang harus terpenuhi jika tree akan diekspansi, yaitu [5]:

- Fuzziness control threshold (FCT) / θ_f .

Jika proporsi dari himpunan data dari kelas C_k lebih besar atau sama dengan nilai threshold θ_n , maka hentikan ekspansi tree. Sebagai contoh: jika pada sebuah sub-dataset rasio dari kelas 1 adalah 90%, kelas 2 adalah 10% dan θ_n adalah 85%, maka hentikan ekspansi tree.

- Leaf decision threshold (LDT) / θ_n .

Jika banyaknya anggota himpunan data pada suatu node lebih kecil dari threshold θ_n , hentikan ekspansi tree. Sebagai contoh: sebuah himpunan data memiliki 600 contoh dengan θ_n adalah 2%. Jika jumlah data contoh pada sebuah node lebih kecil dari 12 (2% dari 600), maka hentikan ekspansi tree.

III. TAHAPAN PENELITIAN

Proses dasar sistem mengacu pada proses dalam Knowledge Discovery in Database (KDD). Proses tersebut dapat diuraikan sebagai berikut :

- Pembersihan data, membuang data dengan nilai yang hilang dan data yang duplikat.
- Transformasi data ke dalam bentuk data fuzzy.
- Data dibagi menjadi training set dan test set dengan menggunakan metode k-fold cross validation [2].
- Aplikasi teknik data mining, merupakan tahap yang penting karena pada tahap ini teknik data mining diaplikasikan terhadap data. Untuk menemukan aturan klasifikasi digunakan metode fuzzy decision tree. Langkah-langkah pada metode tersebut yaitu:
 - Menentukan atribut yang akan digunakan.

- Menentukan banyaknya fuzzy set untuk masing-masing atribut.
- Menentukan banyaknya training set yang akan digunakan.
- Menghitung membership value.
- Memilih besarnya threshold yang akan digunakan.
- Membangun fuzzy decision tree dengan algoritme Fuzzy ID3.

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut :

Perangkat keras berupa komputer personal dengan spesifikasi: Prosesor AMD Athlon 64 2800+, Memori DDR 512 MB, dan Harddisk 80 GB. Perangkat lunak yang digunakan adalah sistem operasi Windows XP Profesional, MATLAB 7.0 sebagai bahasa pemrograman, dan Microsoft Excel 2003.

A. Transformasi Data

Dari 5 atribut yang digunakan, 4 di antaranya merupakan atribut yang kontinu, yaitu GLUN, GPOST, HDL, dan TG, sedangkan atribut Diagnosa adalah atribut kategorik. Berdasarkan referensi hasil laboratorium, range normal untuk atribut GLUN, GPOST, HDL, dan TG diperlihatkan pada Tabel 1.

TABLE 1
NILAI REFERENSI HASIL LABORATORIUM

Kode Pemeriksaan	Satuan	Nilai Normal
GLUN	Mg/DL	70 ~ 100
GPOST	Mg/DL	100 ~ 140
HDL	Mg/DL	40 ~ 60
TG	Mg/DL	50 ~ 150

Atribut yang telah ditransformasi ke dalam bentuk fuzzy antara lain:

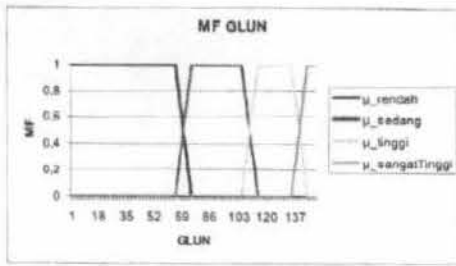
- Atribut GLUN

Atribut GLUN dibagi menjadi 4 kelompok atau linguistic term, yaitu rendah (GLUN < 70 mg/DL), sedang (70 mg/DL <= GLUN < 110 mg/DL), tinggi (110 mg/DL <= GLUN < 140 mg/DL), dan sangat tinggi (GLUN >= 140 mg/DL) [4]. Dari pembagian itu dapat ditentukan membership function dari himpunan fuzzy rendah, sedang, tinggi, dan sangat tinggi untuk atribut GLUN secara terpisah yaitu:

$$\mu_{rendah}(x) = \begin{cases} 1 & ; x < 65 \\ \frac{x-75}{-10} & ; 65 \leq x < 75 \\ 0 & ; x \geq 75 \end{cases} \quad \mu_{sedang}(x) = \begin{cases} 0 & ; x < 65 \\ \frac{x-65}{10} & ; 65 \leq x < 75 \\ 1 & ; 75 \leq x < 105 \\ \frac{x-115}{-10} & ; 105 \leq x < 115 \\ 0 & ; x \geq 115 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x < 105 \\ \frac{x-105}{10} & ; 105 \leq x < 115 \\ 1 & ; 115 \leq x < 135 \\ \frac{x-145}{-10} & ; 135 \leq x < 145 \\ 0 & ; x \geq 145 \end{cases} \quad \mu_{sangatTinggi}(x) = \begin{cases} 0 & ; x < 135 \\ \frac{x-135}{10} & ; 135 \leq x < 145 \\ 1 & ; x \geq 145 \end{cases}$$

Himpunan fuzzy untuk setiap linguistic term menggunakan kurva dengan bentuk trapesium seperti pada Gambar 1.



Gambar 1 Himpunan fuzzy atribut GLUN

• Atribut GPOST

Atribut GPOST dibagi menjadi 4 kelompok atau *linguistic term*, yaitu rendah (GPOST < 100 mg/DL), sedang (100 mg/DL <= GPOST < 140 mg/DL), tinggi (140 mg/DL <= GPOST < 200 mg/DL), dan sangat tinggi (GPOST >= 200 mg/DL) [4]. Dari pembagian itu dapat ditentukan *membership function* dari himpunan fuzzy rendah, sedang, tinggi, dan sangat tinggi untuk atribut GPOST secara terpisah yaitu:

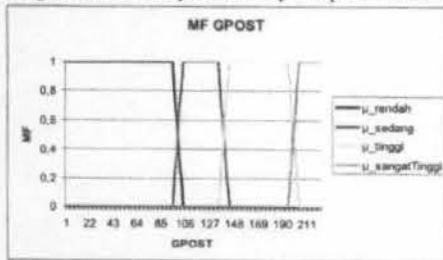
$$\mu_{rendah}(x) = \begin{cases} 1 & ; x < 95 \\ \frac{x-105}{-10} & ; 95 \leq x < 105 \\ 0 & ; x \geq 105 \end{cases}$$

$$\mu_{sedang}(x) = \begin{cases} 0 & ; x < 95 \\ \frac{x-95}{10} & ; 95 \leq x < 105 \\ 1 & ; 105 \leq x < 135 \\ \frac{x-145}{-10} & ; 135 \leq x < 145 \\ 0 & ; x \geq 145 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x < 135 \\ \frac{x-135}{10} & ; 135 \leq x < 145 \\ 1 & ; 145 \leq x < 195 \\ \frac{x-205}{-10} & ; 195 \leq x < 205 \\ 0 & ; x \geq 205 \end{cases}$$

$$\mu_{sangatTinggi}(x) = \begin{cases} 0 & ; x < 195 \\ \frac{x-195}{10} & ; 195 \leq x < 205 \\ 1 & ; x \geq 205 \end{cases}$$

Himpunan fuzzy untuk setiap *linguistic term* menggunakan kurva dengan bentuk trapesium seperti pada Gambar 2.



Gambar 2 Himpunan fuzzy atribut GPOST

• Atribut HDL

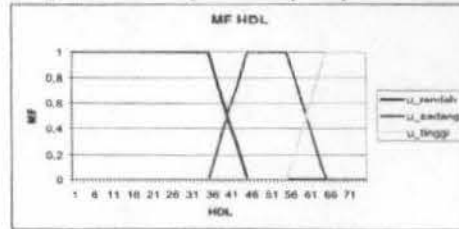
Atribut HDL dibagi menjadi 3 kelompok atau *linguistic term*, yaitu rendah (HDL < 40 mg/DL), sedang (40 mg/DL <= HDL < 60 mg/DL), dan tinggi (HDL >= 60 mg/DL) [4]. Dari pembagian itu dapat ditentukan *membership function* dari himpunan fuzzy rendah, sedang, tinggi, dan sangat tinggi untuk atribut HDL secara terpisah yaitu:

$$\mu_{rendah}(x) = \begin{cases} 1 & ; x < 35 \\ \frac{x-45}{-10} & ; 35 \leq x < 45 \\ 0 & ; x \geq 45 \end{cases}$$

$$\mu_{sedang}(x) = \begin{cases} 0 & ; x < 35 \\ \frac{x-35}{10} & ; 35 \leq x < 45 \\ 1 & ; 45 \leq x < 55 \\ \frac{x-65}{-10} & ; 55 \leq x < 65 \\ 0 & ; x \geq 65 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x < 55 \\ \frac{x-55}{10} & ; 55 \leq x < 65 \\ 1 & ; x \geq 65 \end{cases}$$

Himpunan fuzzy untuk setiap *linguistic term* menggunakan kurva dengan bentuk trapesium seperti pada Gambar 3.



Gambar 3 Himpunan fuzzy atribut HDL

• Atribut TG

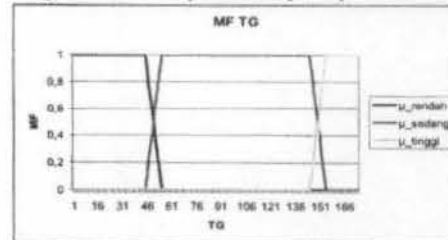
Atribut TG dibagi menjadi 3 kelompok atau *linguistic term*, yaitu rendah (TG < 50 mg/DL), sedang (50 mg/DL <= TG < 150 mg/DL), dan tinggi (TG >= 150 mg/DL) [4]. Dari pembagian itu dapat ditentukan *membership function* dari himpunan fuzzy rendah, sedang, tinggi, dan sangat tinggi untuk atribut TG secara terpisah yaitu:

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

$$\mu_{sedang}(x) = \begin{cases} 0 & ; x < 45 \\ \frac{x-45}{10} & ; 45 \leq x < 55 \\ 1 & ; 55 \leq x < 145 \\ \frac{x-155}{-10} & ; 145 \leq x < 155 \\ 0 & ; x \geq 155 \end{cases}$$

$$\mu_{tinggi}(x) = \begin{cases} 0 & ; x < 145 \\ \frac{x-145}{10} & ; 145 \leq x < 155 \\ 1 & ; x \geq 155 \end{cases}$$

Himpunan fuzzy untuk setiap *linguistic term* menggunakan kurva dengan bentuk trapesium seperti pada Gambar 4.



Gambar 4 Himpunan fuzzy atribut TG

Data dari atribut GLUN, GPOST, HDL, dan TG kemudian akan ditransformasi ke dalam bentuk fuzzy dengan menghitung derajat keanggotaan fuzzy pada tiap-tiap himpunan dari domain setiap atribut linguistik.

• Atribut Dignosa

Atribut Diagnosa selanjutnya akan disebut sebagai CLASS, direpresentasikan oleh dua buah peubah linguistik yaitu "negatif diabetes" dan "positif diabetes". Kedua *linguistic term*nya tersebut didefinisikan sebagai berikut:

"negatif diabetes" = 1
 "positif diabetes" = 2

Nilai atribut CLASS yang akan dikategorikan sebagai positif diabetes adalah diagnosa dengan label E10 (*Insulin-dependent diabetes mellitus*), E11 (*Non-insulin-dependet diabetes mellitus*), E12 (*Malnutrition-related diabetes mellitus*), dan E13 (*Unspecified diabetes mellitus*), selainnya dikategorikan sebagai negatif diabetes.

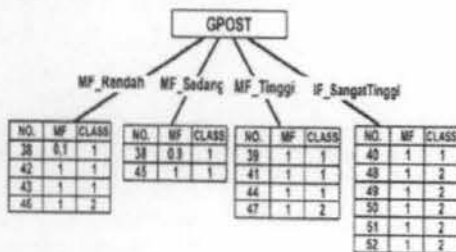
B. Data Mining

Pada tahap ini dilakukan teknik *data mining* menggunakan algoritme FID3 untuk membangun *fuzzy decision tree* (FDT). Data yang telah ditransformasi akan dibagi menjadi *training set* dan *test set*. Pembagian data ini menggunakan metode *10-fold cross validation*. Data akan dibagi menjadi 10 *subset* (S_1, \dots, S_{10}) yang berbeda dengan jumlah yang sama besar. Setiap kali sebuah *subset* digunakan sebagai *test set* maka 9 buah partisi lainnya akan dijadikan sebagai *training set*.

Fase *training* dilakukan untuk membangun FDT dengan menggunakan algoritme FID3. Proses *training* harus melalui berbagai tahapan, sebagai contoh akan dijelaskan pembentukan FDT dengan menggunakan contoh *training set* yang terdiri dari 15 data atau *record*.

Pada contoh *training set* tersebut akan diterapkan algoritme fuzzy ID3 untuk menemukan model atau aturan klasifikasi. Langkah-langkah pembentukan aturan klasifikasi dengan algoritme FID3, yaitu:

1. Membuat *root node* dari semua data *training* yang ada dengan memberi nilai derajat keanggotaan untuk semua *record* sama dengan 1.
2. Menghitung *fuzzy entropy* dari *training set* yang ada. Dari hasil perhitungan diperoleh nilai *fuzzy entropy* sebesar 0.9968. Nilai ini akan digunakan untuk menghitung nilai *information gain* dari masing-masing atribut.
3. Menghitung *information gain* dari atribut GLUN, GPOST, HDL, dan TG, masing-masing diperoleh nilai 0.2064, 0.3330, 0.0304, dan 0.0050. Dari hasil tersebut dipilih atribut dengan nilai *information gain* terbesar yaitu GPOST. Atribut GPOST selanjutnya digunakan untuk mengekspansi *tree* atau menjadi *root-node*, tapi pada *sub-node* selanjutnya atribut GPOST tidak dapat digunakan untuk mengekspansi *tree*.
4. *Data training* diekspansi berdasarkan atribut GPOST sehingga diperoleh hasil seperti pada Gambar 5.



Gambar 5 Hasil ekspansi *training set* berdasarkan atribut GPOST

Nilai derajat keanggotaan yang baru masing-masing *record* pada *sub-node* diperoleh dari hasil perkalian antara derajat keanggotaan pada *root node* dan derajat keanggotaan atribut yang digunakan untuk mengekspansi *tree*. Misalkan, untuk *sub-node* dengan nilai atribut rendah, nilai derajat keanggotaan dari data no.38 $\mu_i = 0.1$ dan derajat keanggotaan dari data no.38 pada *root node* $\mu_{old} = 1$, maka nilai derajat keanggotaannya pada *sub-node* μ_{new} adalah:

$$\mu_{new} = \mu_i \otimes \mu_{old} = 1 \otimes 0.1 = 0.1$$

5. Menghitung proporsi dari tiap kelas yang ada pada tiap-tiap *node*. Misalkan, untuk *sub-node* dengan nilai keanggotaan atribut rendah, proporsi kelasnya adalah:

$$C_1 = 0.1 + 1 + 1 = 2.1$$

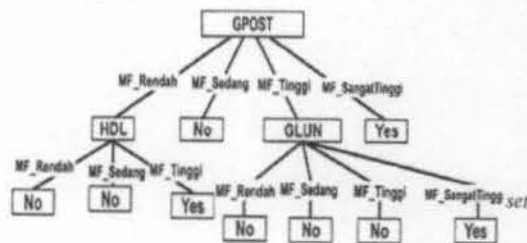
$$C_2 = 1$$

$$\text{Proporsi kelas 1 (negatif diabetes)} = \frac{C_1}{C_1 + C_2} * 100\% = 67.74\%$$

Proporsi kelas 2 (positif diabetes)

$$= \frac{C_2}{C_1 + C_2} * 100\% = 32.26\%$$

6. Pada contoh ini digunakan *fuzziness control threshold* (θ_f) sebesar 80% dan *leaf decision threshold* (θ_n) sebesar $20\% * 15 = 3$. Kedua *threshold* tersebut akan menentukan apakah *sub-node* akan terus diekspansi atau tidak. Misalkan *sub-node* dengan nilai atribut rendah yang memiliki data sebanyak 4, berdasarkan nilai proporsi kelas 1 (67.74%) dan kelas 2 (32.26%) yang lebih kecil dari θ_f (80%) dan banyaknya data atau *record* pada *sub-node* tersebut lebih besar dari θ_n , maka *sub-node* tersebut akan terus diekspansi. Lain halnya jika θ_f yang digunakan adalah 65%, maka *sub-node* tersebut tidak akan diekspansi.
7. Lakukan terus ekspansi dari *sub-node* yang ada sampai tidak ada lagi data yang dapat diekspansi atau tidak ada lagi atribut yang dapat digunakan untuk mengekspansi *tree* yaitu ketika *tree* yang terbentuk sudah mencapai kedalaman maksimum atau *sub-node* tidak memenuhi syarat dari *threshold* yang diberikan. Jika *sub-node* sudah tidak dapat diekspansi maka nilai proporsi kelas terbesar merupakan kesimpulan dari sekumpulan aturan yang diperoleh dengan menghubungkan setiap *node* yang dilewati dari *root node* hingga *leaf node*. Gambar 6 menunjukkan *fuzzy decision tree* yang diperoleh dari *training set*.



Gambar 6 Fuzzy decision tree untuk *training set*

IV. HASIL DAN PEMBAHASAN

Berdasarkan langkah-langkah algoritme FID3 dalam Bab III, diperoleh sebuah model yang terdiri atas 9 aturan dengan menggunakan *training set*. Model atau aturan klasifikasi yang diperoleh:

1. IF GPOST rendah AND HDL rendah THEN Negatif Diabetes
2. IF GPOST rendah AND HDL sedang THEN Negatif Diabetes
3. IF GPOST rendah AND HDL tinggi THEN Positif Diabetes
4. IF GPOST sedang THEN Negatif Diabetes

5. IF GPOST tinggi AND GLUN rendah THEN Negatif Diabetes
6. IF GPOST tinggi AND GLUN sedang THEN Negatif Diabetes
7. IF GPOST tinggi AND GLUN tinggi THEN Negatif Diabetes
8. IF GPOST tinggi AND GLUN sangat tinggi THEN Positif Diabetes
9. IF GPOST sangat tinggi THEN Positif Diabetes

Berdasar metode uji *10-fold cross validation*, proses *training* dilakukan sebanyak 240 kali. Untuk setiap *training set*, proses *training* dilakukan sebanyak 24 kali, dengan mengubah nilai θ , sebanyak 6 kali yaitu 75%, 80%, 85%, 90%, 95%, dan 98%, dan untuk masing-masing nilai θ , yang sama diberikan nilai θ_n yang berbeda-beda yaitu 3%, 5%, 8%, dan 10%. Rata-rata jumlah aturan yang dihasilkan pada proses *training* dan waktu eksekusi yang dibutuhkan dapat dilihat pada Tabel 2 dan Tabel 3.

TABLE 2
RATA-RATA JUMLAH ATURAN

θ	θ_n			
	3%	5%	8%	10%
75%	4	4	4	4
80%	7	7	7	6
85%	11	10	10	8
90%	12	11	10	8
95%	20	18	15	11
98%	27	24	20	16

Nilai-nilai θ dan θ_n yang digunakan dipilih berdasarkan hasil percobaan, karena dengan nilai-nilai tersebut jumlah model atau aturan yang dihasilkan mengalami perubahan yang cukup signifikan. Dengan nilai θ , 70%, aturan yang diperoleh tidak berbeda dengan nilai θ , 75% sehingga nilai θ , hanya dipilih sampai 75%.

TABLE 3
RATA-RATA WAKTU EKSEKUSI DALAM DETIK

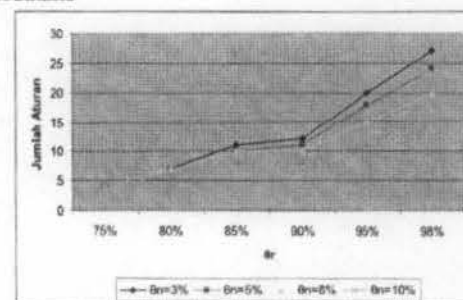
θ	θ_n			
	3%	5%	8%	10%
75%	0,125	0,128	0,120	0,127
80%	0,239	0,227	0,214	0,183
85%	0,344	0,334	0,308	0,244
90%	0,404	0,361	0,328	0,264
95%	0,717	0,594	0,492	0,356
98%	0,955	0,842	0,712	0,544

Berdasarkan hasil percobaan, semakin tinggi nilai θ , semakin banyak pula aturan yang dihasilkan, hal ini terjadi karena sebelum suatu *node* didominasi oleh sebuah kelas dan proporsi untuk kelas tersebut di atas atau sama dengan nilai θ , maka *tree* akan terus diekspansi. Pada Tabel 2 terlihat bahwa,

peningkatan yang paling signifikan terjadi pada saat nilai θ , dinaikkan dari 90% menjadi 95% dan dari 95% menjadi 98%. Kondisi semacam ini disebabkan karena pada saat ekspansi *training set* yang pertama kali dilakukan banyak *sub-node* yang proporsi salah satu kelasnya sudah mencapai nilai di atas 90%, sehingga *sub-node* tersebut tidak perlu diekspansi lagi. Ketika nilai θ , ditingkatkan sampai 95%, baru terjadi ekspansi pada beberapa *sub-node* yang lain dan hasil proporsi kelas pada *node* di bawahnya pun ternyata banyak yang lebih rendah dari nilai θ , yang mengakibatkan *training set* akan terus diekspansi sampai seluruh atribut terpakai.

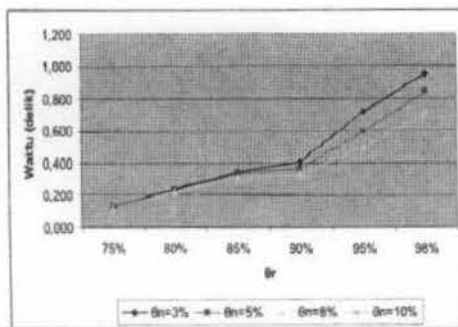
Jika diamati dengan seksama pada Tabel 2, walaupun nilai θ_n ditingkatkan, jumlah aturan yang dihasilkan tidak mengalami penurunan secara signifikan. Berdasarkan pengamatan yang dilakukan, ternyata karakteristik data pada *training set* yang digunakan tidak terlalu berbeda, pada saat terjadi ekspansi *tree* data tidak akan terlalu menyebar, karenanya jumlah himpunan data yang ada pada *sub-node* tidak berbeda jauh dengan jumlah himpunan data yang ada pada *root-node*. Dengan adanya situasi yang demikian, syarat untuk menghentikan ekspansi *tree* yaitu jumlah data atau *record* pada *sub-node* harus lebih kecil dari nilai θ_n sulit untuk tercapai.

Nilai θ , yang terlalu rendah dan atau θ_n yang terlalu tinggi akan menghasilkan *tree* dengan ukuran yang kecil sehingga jumlah aturan yang dihasilkan juga sangat sedikit, hal ini terjadi karena *tree* yang sedang dibangun mengalami pemotongan (*pruning*) pada saat model masih mempelajari struktur dari *training set*. Sebaliknya, nilai θ , yang terlalu tinggi dan atau θ_n yang terlalu rendah kadang kala akan menyebabkan *fuzzy decision tree* berperilaku seperti *decision tree* biasa yang tidak memerlukan adanya *threshold* sehingga menghasilkan *tree* dengan ukuran sangat besar dan jumlah aturan yang juga sangat banyak, karena *tree* akan terus diekspansi sampai *leaf-node* terdalam.



Gambar 7 Perbandingan rata-rata jumlah aturan

Gambar 7 menunjukkan perbandingan rata-rata jumlah aturan yang dihasilkan pada proses *training*. Dapat terlihat bahwa semakin tinggi nilai θ , akan menyebabkan jumlah aturan yang dihasilkan juga meningkat, hal sebaliknya terjadi jika nilai θ_n semakin tinggi maka aturan yang dihasilkan cenderung berkurang.

Gambar 8 Perbandingan rata-rata waktu eksekusi proses *training*

Dengan melihat Gambar 7 dan Gambar 8, dapat disimpulkan bahwa, semakin tinggi nilai θ_r yang digunakan akan menghasilkan jumlah aturan yang semakin banyak sehingga waktu yang dibutuhkan untuk menghasilkan aturan-aturan tersebut juga meningkat, hal ini terjadi karena proses yang harus dilakukan untuk membangun *tree* semakin banyak.

Untuk mengukur akurasi dari model yang dihasilkan pada fase *training*, proses *testing* dilakukan sebanyak 240 kali. Proses *testing* dilakukan dengan cara memasukkan aturan yang diperoleh dari proses *training* ke dalam sebuah FIS Mamdani [7] untuk menentukan kelas dari masing-masing *record* pada *test set*. Untuk satu kali proses *training* dilakukan satu kali proses *testing*.

Dengan melihat hasil *testing*, mengubah nilai θ_r dan θ_n tidak menyebabkan adanya perubahan pada nilai akurasi dari model yang ada, walaupun jumlah aturan yang dihasilkan proses *training* mengalami peningkatan. Hal ini terjadi karena secara kebetulan semua aturan pada kedua model tersebut memiliki kelas target yang sama yaitu negatif diabetes dan *test set* yang digunakan juga berasal dari kelas target yang sama sehingga hasil *testing* dari kedua buah model dengan *test set* yang sama tidak mengalami perubahan. Namun model dengan aturan yang seragam seperti ini tidak dapat dipakai untuk melakukan prediksi karena berapapun nilai atribut yang diberikan hasilnya akan selalu negatif diabetes. Hasil *testing* akan berbeda jika aturan-aturan pada model yang dihasilkan memiliki kelas target yang berbeda. Model untuk *training set* pertama dengan θ_r (75%) dan θ_n (3%):

```
IF GPOST rendah THEN Negatif Diabetes
IF GPOST sedang THEN Negatif Diabetes
IF GPOST tinggi THEN Negatif Diabetes
IF GPOST sangat tinggi THEN Negatif Diabetes
```

Pada *test set* ketujuh dan kedelapan, nilai akurasi mengalami penurunan saat θ_r ditingkatkan menjadi 80%. Berdasarkan pengamatan yang dilakukan, keadaan seperti ini dinamakan *overfitting* karena terlalu tingginya θ_r untuk *training set* tersebut, sehingga *tree* akan terus diekspansi sampai betul-betul sesuai dengan *training set*. Akibatnya *tree* memiliki node-node yang mengandung data yang mengalami kesalahan klasifikasi.

V. EVALUASI KINERJA FID3

Evaluasi kinerja algoritme FID3 dapat diketahui dengan cara menghitung rata-rata akurasi dari seluruh proses *testing* pada 10 *test set* yang berbeda. Evaluasi kinerja dari algoritme FID3 pada nilai θ_r dan θ_n yang berbeda dapat dilihat pada Tabel 4.

TABLE 4

AKURASI FUZZY DECISION TREE

θ_r	θ_n			
	3%	5%	8%	10%
75%	94,14	94,14	94,15	94,15
80%	92,07	92,07	93,45	93,45
85%	92,07	92,07	93,45	93,45
90%	92,07	92,07	93,45	93,45
95%	90,69	91,73	93,10	93,45
98%	90,69	91,73	93,10	93,45

Dari Tabel 4 dapat dilihat bahwa kinerja algoritme FID3 mengalami penurunan jika nilai θ_r semakin besar dan atau nilai θ_n semakin kecil, walaupun penurunan yang terjadi tidaklah signifikan sehingga masih dapat ditoleransi. Seperti telah dijelaskan sebelumnya kondisi ini disebabkan karena terjadinya fenomena *overfitting*. Nilai akurasi terbaik yaitu 94,15% diperoleh pada $\theta_r = 75%$ dan $\theta_n = 8%$ atau 10%.

VI. KESIMPULAN

Dari berbagai percobaan yang dilakukan terhadap data Diabetes didapat kesimpulan bahwa algoritme FID3 memiliki kinerja yang baik dalam membentuk *fuzzy decision tree* untuk data Diabetes yang ada. Nilai akurasi terbaik dari model yaitu 94,15% diperoleh pada *fuzziness control threshold* (θ_r) = 75% dan *leaf decision threshold* (θ_n) = 8% atau 10%. Nilai θ_r dan θ_n sangat berpengaruh terhadap jumlah aturan yang dihasilkan, nilai θ_r yang terlalu tinggi akan menyebabkan turunnya nilai akurasi. Di lain pihak, nilai θ_n yang terlalu rendah juga dapat menyebabkan akurasi menurun.

REFERENCES

- [1] E. Cox. *Fuzzy Modeling and Algorithms for Data Mining and Exploration*. USA: Academic Press, 2005.
- [2] L. Fu. *Neural Network In Computer Intelligence*. Singapura: McGraw Hill, 1994.
- [3] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Simon Fraser University, USA: Morgan Kaufman, 2006.
- [4] Herwanto. *Pengembangan Sistem Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan Algoritme Classification Based Association* [Tesis]. Bogor. Departemen Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor, 2006.
- [5] G. Liang. *A Comparative Study of Three Decision Tree algorithms: ID3, Fuzzy ID3 and Probabilistic Fuzzy ID3*. Informatics & Economics Erasmus University Rotterdam Rotterdam, the Netherlands, 2005.
- [6] C. Marsala. *Application of Fuzzy Rule Induction to Data Mining*. University Pierre et Marie Curie, 1998.
- [7] Ormos L. *Soft Computing Method On Thom's Catastrophe Theory For Controlling Of Large-Scale Systems*. University of Miskolc, Department of Automation, 2004.
- [8] Y. Yuan dan Shaw M J. *Induction of fuzzy decision trees*. *Fuzzy Sets and Systems* Vol. 69, 1995.

Firat Romansyah dilahirkan di Indonesia pada tahun 1985. Dia memperoleh gelar Sarjana Ilmu Komputer dari Departemen Ilmu Komputer, Institut Pertanian Bogor, Indonesia pada tahun 2007. Saat ini dia bekerja sebagai *Associate Business Consultant (IT Consultant)* di PT AGIT (Astra Graphia Information Technology).

Imas Sukaesih Sitanggang dilahirkan di Indonesia pada tahun 1975. Dia memperoleh gelar Sarjana Matematika dari Departemen Matematika, Institut Pertanian Bogor, Indonesia pada tahun 1997 dan mendapat gelar Master Ilmu Komputer dari Universitas Gadjah Mada, Indonesia pada tahun 2002. Saat ini dia bekerja sebagai dosen di Departemen Ilmu Komputer, Institut Pertanian Bogor, Indonesia. Topik utama penelitiannya adalah dalam *spatial data mining*.

Sri Nurdianti dilahirkan di Indonesia pada tahun 1960. Dia memperoleh gelar Sarjana Statistika dari Institut Pertanian Bogor, Indonesia pada tahun 1984 dan mendapat gelar Master Ilmu Komputer dari *The University of Western Ontario* di Canada pada tahun 1991. Pada tahun 2005 dia memperoleh gelar Doktor Matematika Terapan dari *Twente University* di Belanda. Saat ini dia bekerja sebagai dosen di Departemen Ilmu Komputer, Institut Pertanian Bogor, Indonesia. Topik utama penelitiannya adalah *scientific computing*.

Internetworking Indonesia Journal

The Indonesian Journal of ICT and Internet Development

ISSN: 1942-9703

About the Internetworking Indonesia Journal

The Internetworking Indonesia Journal (IJ) was established out of the need to address the lack of an Indonesia-wide independent academic and professional journal covering the broad area of Information and Communication Technology (ICT) and Internet development in Indonesia.

The broad aims of the Internetworking Indonesia Journal (IJ) are therefore as follows:

- **Provide an Indonesia-wide independent journal on ICT:** The IJ seeks to be an Indonesia-wide journal independent from any specific institutions in Indonesia (such as universities and government bodies). Currently in Indonesia there are numerous university-issued journals that publish papers only from the respective universities. Often these university journals experience difficulty in maintaining sustainability due to the limited number of internally authored papers. Additionally, most of these university-issued journals do not have an independent review and advisory board, and most do not have referees and reviewers from the international community.
- **Provide a publishing venue for graduate students:** The IJ seeks also to be a publishing venue for graduate students (such as Masters/S2 and PhD/S3 students) as well as working academics in the broad field of ICT. This includes graduate students from Indonesian universities and those studying abroad. The IJ provides an avenue for these students to publish their papers to a journal which is international in its reviewer scope and in its advisory board.
- **Improve the quality of research & publications on ICT in Indonesia:** One of the long term goals of the IJ is to promote research on ICT and Internet development in Indonesia, and over time to improve the quality of academic and technical publications from the Indonesian ICT community. Additionally, the IJ seeks to be the main publication venue for various authors worldwide whose interest include Indonesia and its growing area of information and communication technology.
- **Provide access to academics and professionals overseas:** The Editorial Advisory Board (EAB) of the IJ is intentionally composed of international academics and professionals, as well as those from Indonesia. The aim here is to provide Indonesian authors with access to international academics and professionals who are aware of and sensitive to the issues facing a developing nation. Similarly, the IJ seeks to provide readers worldwide with easy access to information regarding ICT and Internet development in Indonesia.
- **Promote the culture of writing and authorship:** The IJ seeks to promote the culture of writing and of excellent authorship in Indonesia within the broad area of ICT. It is for this reason that the IJ is bilingual in that it accepts and publishes papers in either English or Bahasa Indonesia. Furthermore, the availability of an Indonesia-wide journal with an international advisory board may provide an incentive for young academics, students and professionals in Indonesia to develop writing skills appropriate for a journal. It is hoped that this in-turn may encourage them to subsequently publish in other international journals in the future.