

## Bayesian Belief Network untuk Menghasilkan Fuzzy Association Rules

Rolly Intan<sup>1</sup>, Oviliiani Yenty Yuliana<sup>2</sup>, Dwi Kristanto<sup>3</sup>

**Abstract:** Bayesian Belief Network (BBN), one of the data mining classification methods, is used in this research for mining and analyzing medical track record from a relational data table. In this paper, a mutual information concept is extended using fuzzy labels for determining the relation between two fuzzy nodes. The highest fuzzy information gain is used for mining fuzzy association rules in order to extend a BBN. Meaningful fuzzy labels can be defined for each domain data. For example, fuzzy labels of secondary disease and complication disease are defined for a disease classification. The implemented of the extended BBN in a application program gives a contribution for analyzing medical track record based on BBN graph and conditional probability tables.

**Keywords:** Medical record, mutual information, conditional probability, bayesian belief network, fuzzy association rules.

### Pendahuluan

*Bayesian Belief Network* (BBN) adalah sebuah model probabilistik yang dinyatakan dalam bentuk *Directed Acyclic Graph* (DAG) untuk menunjukkan hubungan probabilitas antar peristiwa (*event*) yang terjadi. Sebagaimana dinyatakan dalam DAG, maka *node* dalam *BBN* menunjukkan variabel pada suatu *domain*, dan anak panah menunjukkan hubungan ketergantungan antara variabel tersebut.

Terdapat beberapa penelitian terkait dengan BBN. Pengintegrasian *fuzzy data* ke dalam *bayesian network* dengan memanfaatkan model kondisi Gaussian untuk membuat prosedur *fuzzy* dilakukan oleh Lin *et al.* [9]. Peneliti yang mengintegrasikan *fuzzy logic* ke dalam *bayesian network* antara lain Lin *et al.* [9], Chiu *et al.* [2], Klir dan Yuan [6]. Pembuatan *bayesian network* menggunakan pendekatan teori informasi diperkenalkan oleh Cheng *et al.* [1]. Penelitian yang terakhir erat hubungannya dengan penelitian yang dilakukan dalam naskah ini, suatu konsep perhitungan *Mutual Information* (MI) dikembangkan menggunakan *fuzzy data*. Hasil MI digunakan untuk menentukan relasi antara dua *node fuzzy data*. Arah garis yang menghubungkan kedua *node* yang saling bergantung ditentukan berdasarkan hasil *conditional probability* dan bersifat *asymmetric*.

<sup>1,2,3</sup> Fakultas Teknologi Industri, Jurusan Teknik Informatika, Universitas Kristen Petra. Jl. Siwalankerto 121-131, Surabaya 60236. Email: rintan@petra.ac.id, ovi@petra.ac.id, dwi\_k87@hotmail.com

Naskah masuk 21 Desember 2009; revisi 7 Mei 2010; diterima untuk dipublikasikan 12 Mei 2010.

Naskah tersusun sebagai berikut. Metodologi Penelitian berisi konsep dasar dan notasi yang akan digunakan dalam pembahasan. Hasil dan pembahasan merupakan bagian inti kontribusi dari naskah ini yang disertai dengan contoh ilustrasi, serta pengujian kecepatan dan kehandalan terhadap aplikasi program BBN yang dibuat. Pada bagian akhir akan diberikan simpulan dari naskah ini.

### Metode Penelitian

#### Relasi

Skema dari sebuah *relational database* dinotasikan sebagai  $R$ . Suatu relasi  $R$  didefinisikan sebagai  $R \subseteq D_1 \times D_2 \times D_3 \times \dots \times D_n$ , dalam hal ini  $D_i$  adalah *domain* data ke- $i$ ,  $t_k \in R$  adalah merupakan *tuple* ke- $k$  dari suatu relasi  $R$  yang dinyatakan sebagai  $t_k = \langle d_{k1}, d_{k2}, d_{k3}, \dots, d_{km} \rangle$ , sedangkan  $d_{km}$  adalah nilai pada *tuple* ke- $k$  dan *domain* ke- $m$  (Codd, [3]).

#### Mutual Information

Dalam teori probabilitas dan teori informasi, MI dari 2 variabel acak merupakan nilai ukur yang menyatakan keterikatan/ketergantungan (*mutual dependence*) antara kedua variabel tersebut. Unit pengukur yang umum digunakan untuk menghitung MI adalah bit, sehingga menggunakan logaritma basis 2. Secara formal, MI antara 2 variabel  $A$  dan  $B$  dapat didefinisikan dengan persamaan (1) (Kulback dan Leibler, [8]; Rényi, [12]).

$$MI(A, B) = \sum_{b \in B} \sum_{a \in A} P(a, b) \log_2 \left( \frac{P(a, b)}{P(a) \times P(b)} \right) \quad (1)$$

$P(a,b)$  merupakan *joint probability distribution* dari  $A$  dan  $B$ ,  $P(a)$  dan  $P(b)$  merupakan *marginal probability distribution*  $a$  dan  $b$ . Nilai  $MI(A, B)=0$  jika dan hanya jika variabel  $A$  dan  $B$  merupakan variabel yang saling independent. Hal ini berhubungan dengan, jika  $a$  dan  $b$  independent, maka nilai  $P(a, b)=P(a) \times P(b)$ , sehingga:

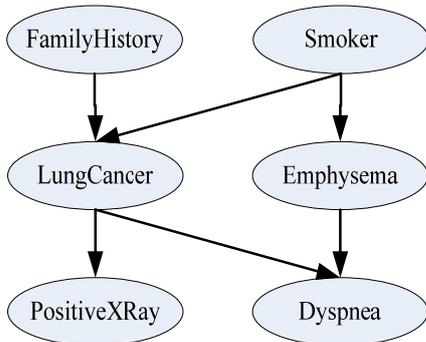
$$\log_2 \left( \frac{P(a,b)}{P(a) \times P(b)} \right) = \log 1 = 0$$

Selain itu, MI bersifat simetris, artinya nilai  $MI(A, B)=MI(B, A)$ .

**Bayesian Belief Network**

BBN adalah model probabilistik dalam bentuk DAG yang digunakan untuk menggambarkan hubungan probabilitas dan *probabilistic inference* antar variabel (Neapolitan [10]). BBN didefinisikan dengan 2 komponen oleh Han, Kamber, dan Pei [4]. Komponen pertama adalah DAG, berupa *node* dan anak panah. Setiap *node* menggambarkan suatu variabel acak (*random variable*) dan setiap anak panah menggambarkan suatu *probabilistic dependence*. Jika suatu anak panah digambarkan dari *node A* ke *node B*, maka  $A$  adalah *parent/immediate predecessor B* dan  $B$  merupakan *descendent A*. Setiap variabel *conditionally independent* dari *non-descendent* variabel, jika diketahui *node parent*-nya. Variabel yang terdapat pada *network* dapat bernilai diskrit atau kontinu. Contoh sederhana BBN dapat dilihat pada Gambar 1.

Komponen kedua adalah *Conditional Probability Table* (CPT) untuk setiap variabel yang terdapat pada *network*. CPT untuk variabel  $B$ , menspesifikasikan *conditional distribution*  $P(B|Parent(B))$ , dimana *parent(B)* adalah *node parent* dari  $B$ . Seperti yang tampak pada Tabel 1, *conditional probability* untuk setiap nilai *Lung Cancer* (LC) diberikan dari setiap kombinasi yang dapat diperoleh dari nilai *node parent*-nya (dalam hal ini yaitu *Family History* (FH) dan *Smoker* (S)).



Gambar 1. Bayesian belief network

Tabel 1. Conditional probability untuk variabel LC dengan node parent FH dan S

	FH,S	FH,~S	~FH,S	~FH,~S
LC	0,8	0,5	0,7	0,1
~LC	0,2	0,5	0,3	0,9

*Joint probability* setiap nilai variabel dihitung dengan persamaan sebagai berikut:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | Parents(X_i)) \tag{2}$$

dimana  $Parent(X_i)$  merupakan *node parent*  $X_i$  dan nilai  $P(x_i | Parent(X_i))$  bisa diperoleh dari CPT untuk  $X_i$ .

**Hasil dan Pembahasan**

Pada bagian ini akan dibahas tentang pengembangan MI untuk menghasilkan *fuzzy association rules* yang berguna untuk mengembangkan BBN.

**BBN dengan Pendekatan MI**

Dalam naskah ini, suatu pengembangan konsep MI diperkenalkan dengan persamaan (1) agar dapat digunakan untuk memroses *fuzzy data*. Suatu *fuzzy set A* terhadap *domain*  $D_i$  didefinisikan sebagai  $A: D_i \rightarrow [0,1]$  sehingga  $A \in F(D_i)$ , dimana  $F(D_i)$  adalah *fuzzy power set* dari *domain*  $D_i$ . Hal yang sama juga berlaku untuk sebuah *fuzzy set B* terhadap *domain*  $D_j$  didefinisikan sebagai  $B: D_j \rightarrow [0,1]$  sehingga  $B \in F(D_j)$ , dimana  $F(D_j)$  adalah *fuzzy power set* dari *domain*  $D_j$ . Nilai MI antara 2 *fuzzy sets*  $A \in F(D_i)$  dan  $B \in F(D_j)$  dapat didefinisikan dan dihitung dengan persamaan (3). Jika nilai  $MI(A, B)$  lebih besar dari 0, maka  $A$  dan  $B$  diasumsikan memiliki relasi.  $P(A,B)$ ,  $P(A)$ , dan  $P(B)$  masing-masing dihitung menggunakan persamaan (4), (5) dan (6); Intan and Mukaidono [5].

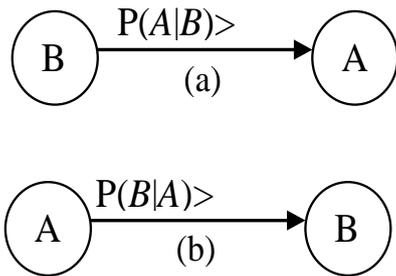
$$MI(A, B) = P(A, B)^2 \log \left( \frac{P(A, B)}{P(A) \times P(B)} \right) \tag{3}$$

$$P(A, B) = \frac{\sum_{k=1}^{|R|} \min(A(d_{ki}), B(d_{kj}))}{|R|}, d_{ki} \in D_i, d_{kj} \in D_j \tag{4}$$

$$P(A) = \frac{\sum_{k=1}^{|R|} A(d_{ki})}{|R|}, d_{ki} \in D_i \tag{5}$$

$$P(B) = \frac{\sum_{k=1}^{|R|} B(d_{kj})}{|R|}, d_{kj} \in D_j \tag{6}$$

Dalam hal ini  $A(d_{ki})$  dan  $B(d_{kj})$  adalah nilai keanggotaan  $d_{ki}$  dan  $d_{kj}$  terhadap *fuzzy set*  $A$  dan  $B$ .



**Gambar 2.** Conditional probability untuk menentukan arah relasi

Sebagaimana didefinisikan  $d_{ki}$  dan  $d_{kj}$  adalah data/ nilai pada *tuple* ke- $k$ , domain ke- $i$  dan ke- $j$ .  $|R|$  adalah jumlah *tuple* pada relasi  $R$ . Persamaan (4) didefinisikan sebagai *joint fuzzy probability relation*. Hal yang sama juga untuk persamaan (5) dan (6) didefinisikan sebagai *fuzzy probability relation* suatu *fuzzy set*  $A$  dan  $B$  pada relasi  $R$ . Untuk menentukan arah relasi antara dua *node* BBN yang akan dibangun dapat menggunakan hasil perhitungan *conditional probability fuzzy* data  $A$  jika diketahui  $B$  sebagaimana dinyatakan dalam persamaan (7).

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{\sum_{k=1}^{|R|} \min(A(d_{ki}), B(d_{kj}))}{\sum_{k=1}^{|R|} B(d_{kj})} \quad (7)$$

Persamaan (7) didefinisikan sebagai *fuzzy conditional probability relation* oleh Intan dan Mukaidono [5]. Jika  $P(A|B) > P(B|A)$ , maka relasi dari *node*  $B$  ke *node*  $A$  seperti tampak pada Gambar 2 (a). Sebaliknya jika  $P(A|B) < P(B|A)$ , maka relasi dari *node*  $A$  ke *node*  $B$  seperti tampak pada Gambar 2 (b). Andaikata yang terjadi adalah  $P(A|B) = P(B|A)$ , maka terdapat dua kemungkinan relasi yaitu dari *node*  $A$  ke *node*  $B$  dan *node*  $B$  ke *node*  $A$ .

### Algoritma

Secara garis besar kerangka proses membangun BBN untuk menghasilkan *fuzzy association rule* adalah mengkonversikan nilai setiap *domain* berdasarkan bobot yang ditetapkan dan membangun BBN. Untuk membangun BBN perlu menghitung MI, menentukan arah relasi, menggambarkan BBN, dan membuat tabel *conditional probability*. Algoritma untuk membangun BBN tampak pada Gambar 3. Langkah untuk menghitung total masing-masing domain tampak pada Gambar 3 baris ke-2 sampai dengan baris ke-11. Untuk mengetahui apakah dua *node* yang dianalisa memiliki relasi dapat dihitung menggunakan persamaan (4), tampak pada Gambar 3 baris ke-24. Untuk keperluan tersebut perlu dihitung  $P(A, B)$  menggunakan persamaan (4) tampak pada Gambar 3 baris ke-15 sampai dengan baris ke-23. Selain itu juga diperlukan perhitungan

$P(A)$  dan  $P(B)$  masing-masing menggunakan persamaan (5) dan (6), tampak pada Gambar 3 baris ke-13 dan ke-15. Jika MI lebih besar dari 0 baru ditentukan arah relasi dari kedua *node* tersebut menggunakan persamaan (7), tampak pada Gambar 3 baris ke-25 sampai dengan baris ke-32.

### Contoh Ilustrasi

Untuk memberi pemahaman yang lebih jelas tentang *mining fuzzy association rules* dengan BBN diberikan suatu *relational database* seperti yang tampak pada Tabel 2 (Kristanto [7]). Tabel tersebut memiliki domain Umur, Perokok, GolDar, Penyakit, dan Propinsi dengan jumlah *record* 10. Ketentuan *node* yang dianalisa adalah sebagai berikut:

*Field* Umur didefinisikan dengan persamaan berikut, *node* diberi *fuzzy* label “Muda”

$$Muda(x) = \begin{cases} 1, & x < 40 \\ \frac{60-x}{20}, & 40 \leq x < 60, \\ 0, & x \geq 60 \end{cases}$$

$x \in [0, 100]$  menyatakan umur.

*Field* Perokok dengan bobot nilai  $\left\{ \frac{1}{Ya}, \frac{0}{Tidak} \right\}$ , *node* diberi *fuzzy* label “Perokok”

*Field* GolDar dengan bobot nilai  $\left\{ \frac{1}{A}, \frac{1}{B}, \frac{0}{AB}, \frac{0}{O} \right\}$ , *node* diberi *fuzzy* label “Gol. Darah”

*Field* Penyakit dengan bobot nilai  $\left\{ \frac{1}{Bronkitis}, \frac{1}{Kanker}, \frac{0}{Penyakit\ lain} \right\}$ , *node* diberi *fuzzy* label “Peny. Paru”

*Field* Propinsi dengan bobot nilai  $\left\{ \frac{1}{Jawa\ Barat}, \frac{1}{Jawa\ Tengah}, \frac{1}{Jawa\ Timur}, \frac{0}{Propinsi\ Lain} \right\}$  *node* diberi *fuzzy* label “Prop. Jawa”

Nilai keanggotaan dari setiap *fuzzy set* di atas ditentukan oleh *users* sesuai dengan kebutuhan untuk mengklasifikasikan data ke dalam masing-masing *fuzzy set*. Angka 0 dan 1 pada contoh *fuzzy* label “Perokok”, “Gol.Darah”, “Peny. Paru” dan “Prop. Jawa” adalah merupakan nilai/derajat keanggotaan dari elemen *fuzzy sets* tersebut, sebagai contoh: elemen “Jawa Barat” memiliki derajat keanggotaan 1 terhadap *fuzzy set* “Prop. Jawa”. Hasil pembobotan Tabel 2 berdasarkan ketentuan *node* tersebut dapat dilihat pada Tabel 3.

```

1 {m: jumlah domain; |R| : jumlah tuple}
2 For i=1 to m
3   {memberi nilai awal pada jumlah bobot per
4   domain}
5   D[i]=0;
6 End For
7 For k=1 to |R|
8   {menghitung jumlah bobot per domain}
9   For i=1 to m
10    D[i]=D[i]+d[k,i]
11  End For
12 End For
13 For i=1 to m-1
14   P(A)=D[i]/|R| ;
15   For j=i+1 to m
16     P(B)=D[j]/|R| ; min=0;
17     For k=1 to |R|
18       If d[k,i] < d[k,j] Then
19         min=min+ d[k,i];
20       Else
21         min=min+ d[k,j];
22       End If
23     End For
24     P(A, B)=min/|R| ;
25     MI(A,B)=P(A,B)*log(P(A,B)/(P(A)*P(B),2);
26     If MI(A,B) > 0 Then
27       P(A|B)=P(A,B)/P(B); P(B|A)=P(A,B)/P(A);
28       If P(A|B) > P(B|A) Then
29         D[j]→D[i]
30       Else
31         D[i]→D[j]
32       End If
33     End If
34 End For
35 End For

```

Gambar 3. Algoritma untuk membangun BBN

Tabel 2. Contoh data rekam medik

Record	Umur	Perokok	GolDar	Penyakit	Propinsi
1	20	Ya	A	Bronkitis	Jabar
2	25	Ya	A	Bronkitis	Sumut
3	22	Ya	AB	Bronkitis	Jabar
4	27	Tidak	O	Diare	Jabar
5	35	Tidak	O	Diare	Sulsel
6	45	Ya	AB	Kanker	Jatim
7	40	Ya	O	Kanker	Jateng
8	50	Tidak	O	Diabetes	Jabar
9	60	Ya	B	Bronkitis	Sulsel
10	60	Ya	A	Kanker	Jatim

Tabel 3. Hasil pembobotan data rekam medik

Record	Muda	Perokok	Gol. Darah	Peny. Paru	Prop. Jawa
1	1,00	1	1	1	1
2	1,00	1	1	1	0
3	1,00	1	0	1	1
4	1,00	0	0	0	1
5	1,00	0	0	0	0
6	0,75	1	0	1	1
7	1,00	1	0	1	1
8	0,50	0	0	0	1
9	0,00	1	1	1	0
10	0,00	1	1	1	1
Σ	7,25	7	4	7	7

Tabel 4. Hasil perhitungan MI

Node 1	Node 2	P (Node 1)	P (Node 2)	P(Node 1, Node 2)	MI(Node 1, Node 2)
Muda	Perokok	0,725	0,700	0,475	-0,045
Muda	Gol. Darah	0,725	0,400	0,200	-0,107
Muda	Peny. Paru	0,725	0,700	0,475	-0,045
Muda	Prop. Jawa	0,725	0,700	0,525	0,026
Perokok	Gol. Darah	0,700	0,400	0,400	0,206
Perokok	Peny. Paru	0,700	0,700	0,700	0,360
Perokok	Prop. Jawa	0,700	0,700	0,500	0,015
Gol. Darah	Peny. Paru	0,400	0,700	0,400	0,206
Gol. Darah	Prop. Jawa	0,400	0,700	0,200	-0,097
Peny. Paru	Prop. Jawa	0,700	0,700	0,500	0,015

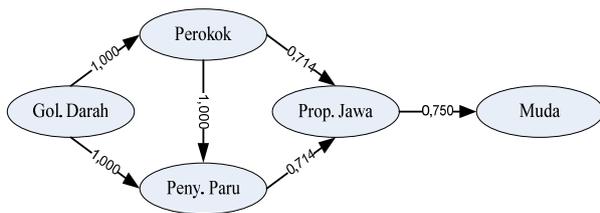
P(Muda, Perokok) dapat dihitung menggunakan persamaan (4). Untuk keperluan tersebut perlu dihitung terlebih dahulu P(Muda) dan P(Perokok) dengan persamaan (5) atau (6). Terakhir MI(Muda, Perokok) dihitung dengan persamaan (3). Perhitungan MI antara node dilakukan dengan cara yang sama. Hasil perhitungan keseluruhan MI tampak pada Tabel 4.

Langkah selanjutnya adalah menentukan node-node yang memiliki relasi, dengan prasyarat MI(Node 1, Node 2) > 0. Berdasarkan Tabel 4 tampak node-node yang memiliki relasi adalah sebagai berikut: Muda–Prop.Jawa; Perokok–Gol.Darah; Perokok–Peny. Paru; Perokok–Prop. Jawa; Gol.Darah–Peny. Paru; dan Peny.Paru–Prop. Jawa

Langkah berikutnya adalah menentukan arah node antar relasi menggunakan conditional probability dengan persamaan (7). Dari hasil perhitungan diperoleh P(Muda | Prop. Jawa) lebih besar dari pada P(Prop. Jawa | Muda). Untuk itu arah relasi yang terbentuk adalah Prop. Jawa menentukan Muda (Prop. Jawa → Muda). Dengan cara yang sama dapat dihitung conditional probability antar node yang berelasi dan ditentukan arah relasinya seperti yang tampak pada Tabel 5. Jika pada hasil perhitungan diperoleh P(Node 1 | Node 2) = P(Node 2 | Node 1), maka arah node dapat berupa Node 1 → Node 2 atau Node 2 → Node 1. Pada Tabel 5 terlihat ada 3 pasangan nodes yang memiliki hasil perhitungan conditional probability yang sama, sehingga terdapat 2<sup>3</sup>=8 macam kombinasi BBN yang dapat dihasilkan. Salah satu kombinasi BBN yang terbentuk dapat dilihat pada Gambar 4. Hasil perhitungan conditional probability P(Node1, Node 2)=1 memiliki arti Node 2 secara mutlak/sepurna menentukan Node 1. Langkah terakhir adalah membuat tabel conditional probability antara semua kemungkinan nilai dari node yang bersangkutan dengan node parent pada setiap node. Sebagai contoh conditional probability node Perokok tampak pada Tabel 6. Data yang digunakan untuk menghitung conditional probability node tersebut dapat dilihat pada Tabel 3, kolom Perokok dan Gol. Darah. Sebagai contoh

**Tabel 5.** Hasil fuzzy association rule dari studi kasus

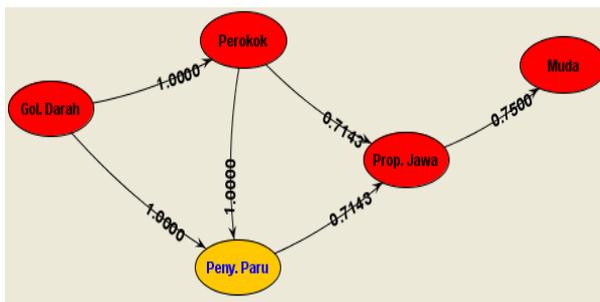
Node 1	Node 2	$P(Node 1   Node 2)$	$P(Node 2   Node 1)$	Arah relasi
Muda	Prop. Jawa	0,750	0,724	Prop. Jawa → Muda
Perokok	Gol. Darah	1,000	0,571	Gol. Darah → Perokok
Perokok	Peny. Paru	1,000	1,000	Perokok → Peny. Paru
Perokok	Prop. Jawa	0,714	0,714	Perokok → Prop. Jawa
Gol. Darah	Peny. Paru	0,571	1,000	Gol. Darah → Peny. Paru
Peny. Paru	Prop. Jawa	0,714	0,714	Peny. Paru → Prop. Jawa



**Gambar 4.** Network untuk data rekam medik

**Tabel 6.** Conditional probability node Perokok

Keterangan	Nilai
$P(\text{Perokok}   \text{Gol. Darah})$ Peluang Pasien adalah Perokok jika diketahui Gol. Darah Pasien (A & B)	$4/4=1$
$P(\sim\text{Perokok}   \text{Gol. Darah})$ Peluang Pasien adalah bukan Perokok jika diketahui Pasien memiliki Gol. Darah (A & B)	$0/4=0$
$P(\text{Perokok}   \sim\text{Gol. Darah})$ Peluang Pasien adalah Perokok jika diketahui Pasien tidak memiliki Gol. Darah (A & B)	$3/6=0,5$
$P(\sim\text{Perokok}   \sim\text{Gol. Darah})$ Peluang Pasien adalah bukan Perokok jika diketahui Pasien tidak memiliki Gol. Darah (A & B)	$3/6=0,5$



**Gambar 5.** Network rekam medik yang dihasilkan oleh aplikasi

Description	Probability
Jika Diketahui Gol. Darah [a, b] maka Probabilitas terjadinya Perokok [ya]	1.0000
Jika Diketahui Gol. Darah [a, b] maka Probabilitas terjadinya Bukan Perokok [ya]	0.0000
Jika Diketahui Bukan Gol. Darah [a, b] maka Probabilitas terjadinya Perokok [ya]	0.5000
Jika Diketahui Bukan Gol. Darah [a, b] maka Probabilitas terjadinya Bukan Perokok [ya]	0.5000

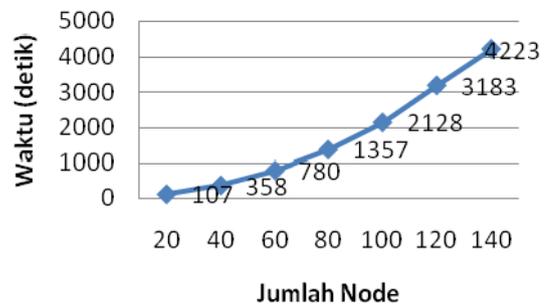
  

Description	Probability
$P(\text{Gol. Darah}   \text{Perokok})$	0.5714

**Gambar 6.** Conditional probability node Perokok dihasilkan oleh aplikasi

**Tabel 7.** Kecepatan proses pembuatan network

Jumlah record	Jumlah node	Waktu
12000	2 (Diag1, Kompl1)	3 detik
	3 (Diag1, Kompl1, Umur)	7 detik
	4 (Diag1, Kompl1, Umur, Gender)	12 detik
	5 (Diag1, Kompl1, Umur, Gender, Kebangsaan)	23 detik
	2 (Diag1, Kompl1)	4 detik
24000	3 (Diag1, Kompl1, Umur)	11 detik
	4 (Diag1, Kompl1, Umur, Gender)	19 detik
	5 (Diag1, Kompl1, Umur, Gender, Kebangsaan)	28 detik
	2 (Diag1, Kompl1)	6 detik
	3 (Diag1, Kompl1, Umur)	15 detik
36000	4 (Diag1, Kompl1, Umur, Gender)	25 detik
	5 (Diag1, Kompl1, Umur, Gender, Kebangsaan)	40 detik
	2 (Diag1, Kompl1)	7 detik
	3 (Diag1, Kompl1, Umur)	22 detik
	4 (Diag1, Kompl1, Umur, Gender)	28 detik
48000	5 (Diag1, Kompl1, Umur, Gender, Kebangsaan)	54 detik
	2 (Diag1, Kompl1)	10 detik
	3 (Diag1, Kompl1, Umur)	27 detik
	4 (Diag1, Kompl1, Umur, Gender)	39 detik
	5 (Diag1, Kompl1, Umur, Gender, Kebangsaan)	75 detik



**Gambar 7.** Perbandingan waktu dan jumlah node yang dapat ditampilkan

$P(\text{Perokok} | \text{Gol. Darah})$  dihitung menggunakan persamaan (7). Jumlah pasien perokok bergolongan darah A dan B adalah 4. Pasien bergolongan darah A dan B berjumlah 4. Untuk itu  $P(\text{Perokok} | \text{Gol. Darah}) = 4/4=1$ . Hal ini dapat diartikan menurut data simulasi bahwa pasien yang bergolongan darah A atau B pasti adalah perokok. Dengan cara yang sama dilakukan perhitungan conditional probability untuk semua kemungkinan nilai dapat dilihat pada Tabel 6.

**Pengujian Aplikasi**

Pengujian aplikasi dilakukan dengan menggunakan komputer dengan processor AMD Turion X2, memory 2 GB, dan hard disk 80 GB. Sebagai contoh sederhana untuk menguji jalannya algoritma, Data

pada Tabel 2 dimasukkan ke dalam Oracle Database. BBN yang dihasilkan aplikasi pada Gambar 5 sama dengan BBN yang dihasilkan dari perhitungan manual. Begitu juga dengan *conditional probability node* Perokok hasil perhitungan manual pada Tabel 6 sama dengan *conditional probability* yang dihasilkan oleh aplikasi pada Gambar 6. Selain itu, juga dilakukan pengujian atas kecepatan proses pembuatan *network* dengan jumlah *record* dan *node* yang bervariasi. Hasil pengujian dapat dilihat pada Tabel 7. Pengujian dilakukan dengan menggunakan data-data yang berasal dari data rekam medik RSU Dr. Sutomo. Tampak kecepatan proses pembuatan *network* pada Tabel 7 dipengaruhi oleh banyaknya *record* dan *node* yang dianalisa. Selain itu juga dilakukan pengujian terhadap seberapa banyak *node* yang dapat ditampilkan oleh aplikasi. Jumlah *record* yang digunakan dalam pengujian ini sebanyak 74000 *records*. Hasil pengujian jumlah *node* yang mampu digambarkan tampak pada Gambar 7. Berdasarkan pengujian tersebut tampak aplikasi mampu menampilkan *network* dengan variasi jumlah *node* yang beragam. Semakin banyak *node* yang dianalisa, maka semakin banyak waktu dan memory yang digunakan aplikasi untuk menampilkan *network*.

### Simpulan

Naskah ini memperkenalkan pengembangan MI dengan menggunakan *fuzzy* label untuk memperoleh *fuzzy association rules* dalam upaya mengembangkan BBN. Hasil pengembangan metode BBN diimplementasikan pada program aplikasi. Aplikasi tersebut dapat menghasilkan diagram BBN dan tabel kemungkinan bersyarat yang berkontribusi untuk menganalisa catatan rekam medik. MI bukanlah satu-satunya metode yang dapat dipakai untuk membangun BBN. Untuk itu perlu adanya suatu penelitian lanjutan untuk membangun BBN dengan berbagai metode selain penggunaan MI.

### Ucapan Terima Kasih

Penelitian ini didanai oleh Hibah Kompetensi DP2M, Direktorat Jendral Pendidikan Tinggi, Departemen Pendidikan Nasional, tahun anggaran 2009, kontrak nomor 251/SP2H/PP/DP2M/V/2009. Ucapan terima kasih kami sampaikan juga kepada Rekam Medik dan Intalasi Teknologi Informasi RSU Dr. Seotomo atas kerja sama dalam penelitian dengan Jurusan Teknik Informatika Universitas Kristen Petra.

### Daftar Pustaka

1. Cheng, J., Bell, D., and Liu, W. Learning Bayesian Networks from Data: An Efficient Approach Based on Information Theory. *Proceedings of the 6th Conference on Information and Knowledge Management*, ACM Press, 1997.
2. Chiu, C. Y., Lo, C. C., and Hsu, Y. X., Integrating Bayesian Theory and Fuzzy Logics with Case-Based Reasoning for Car-Diagnosing Problems.” *Proceedings of the 4th Fuzzy System and Knowledge Discovery*, IEEE Press, 2007, pp. 344-348.
3. Codd, E. F., A Relational Model of Data for Large Shared Data Bank. *Journal of the ACM Communication*, 13(6), 1970, pp. 377-387.
4. Han, J., Kamber, M., and Pei, J., *Data mining: Concepts and techniques*, 2<sup>nd</sup> ed., Morgan Kaufmann, San Fransisco, 2005.
5. Intan, R., and Mukaidono, M., Fuzzy Conditional Probability Relations and Its Applications in Fuzzy Information System. *Knowledge and Information systems, an International Journal*, Springer-Verlag, 6(3), 2004, pp. 345-365.
6. Klir, J., and Yuan, B., *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, New Delhi, 1995.
7. Kristanto, D., *Perancangan dan Pembuatan Aplikasi untuk Menganalisa Data pada Relational Database dengan Metode Bayesian Belief Network*, Tugas Akhir, Jurusan Teknik Informatika, Universitas Kristen Petra, Surabaya, 2009.
8. Kullback, S., and Leibler, R. A., On Information and Sufficiency, *The Annals of Mathematical Statistics*, 22(1), 1951, pp. 79-86.
9. Lin, C. Y., Yin, J. X., Ma, L. H., and Chen, J. Y., An Intelligent Model Based on Fuzzy Bayesian Networks to Predict Astrocytoma Malignant Degree. *Proceedings of the 2nd Cybernetics and Intelligent System*, IEEE Press, 2006, pp. 1-5.
10. Lin, C. Y., Yin, J. X., Ma, L. H., and Chen, J. Y., Fuzzy Bayesian Network-Based Inference in Predicting Astrocytoma Malignant Degree. *Proceedings of the 6th World Congress on Intelligent Control and Automation*, IEEE Press, 2006, pp. 10251-10255.
11. Neapolitan, R. E., *Learning Bayesian Network*, Prentice Hall, New Jersey, 2003.
12. Rényi, A., On Measures of Information and Entropy. *Proceedings of the 4th Berkeley symposium on Mathematics, Statistics and Probability*, 1961, pp. 547-561.