

# ANALISIS PERBANDINGAN TINGKAT AKURASI LOGIKA FUZZY DENGAN NATURAL PARTITION UNTUK PREPROCESSING DATA PERBANKAN

Sumarni Adi<sup>1)</sup>

<sup>1)</sup> Teknik Informatika STMIK AMIKOM Yogyakarta  
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281  
Email : sumarni.a@amikom.ac.id<sup>1)</sup>

## Abstrak

*Preprocessing* adalah kegiatan pembersihan data agar data dapat diproses untuk analisis selanjutnya. Tujuan dari *preprocessing* data adalah untuk mengubah data input mentah menjadi format yang sesuai untuk analisis selanjutnya. Langkah-langkah yang dilakukan antara lain dengan memperbaiki data yang 'kotor', memilih fitur-fitur dari data yang relevan dengan proses pengolahan selanjutnya. Karena banyak cara dalam proses pengumpulan data dan penyimpanan data, maka proses *preprocessing* data mungkin akan memakan waktu yang lama dalam keseluruhan proses penemuan pengetahuan, karena hasil *preprocessing* data dapat meningkatkan kualitas data serta meningkatkan akurasi dan efisiensi proses mining.

*Natural partitioning* dan logika fuzzy merupakan metode yang sering digunakan untuk *preprocessing* data pada tahapan *Discretization and concept hierarchy generation*. Dimana nilai data diubah kedalam bentuk range atau menjadi level yang lebih tinggi. Teknik *discretization* dapat digunakan untuk mengurangi jumlah nilai yang diberikan atribut kontinyu, dengan membagi range pada atribut menjadi suatu interval tertentu. Label interval dapat digunakan untuk merubah nilai data actual.

Hasil akurasi yang didapatkan dengan menggunakan 463 data perbankan dengan 3 variabel yaitu variabel gaji, jaminan dan angsuran adalah hasil pengujian kelas prediksi menggunakan logika fuzzy dan *natural partition* adalah sama yaitu sebesar 312 kelas aman, 150 kelas tidak aman dan 1 data unknow. Sedangkan pengujian akurasi model probabilitas menggunakan logika fuzzy dan *natural partition*, hasil akurasinya juga sama yaitu 0,68 untuk probabilitas aman dan 0,32 untuk probabilitas tidak aman. Sementara pengujian akurasi menggunakan *cross validation* untuk logika fuzzy adalah 66,98 % dan *natural partition* 66,95 %, sedangkan menggunakan *bootstrapping validation* untuk logika fuzzy 64,97 % dan *natural partition* 64,79 %.

**Kata kunci:** *Preprocessing*, Logika fuzzy, *Natural Partition*, Data Perbankan.

## 1. Pendahuluan

Dalam beberapa tahun ini para pengguna *Database Management System* (DBMS) mulai merasakan adanya kebutuhan untuk melakukan penambangan terkait dengan pengetahuan yang ada pada data tersebut. Data mining sendiri merupakan proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru dan bermanfaat dan akhirnya dapat ditemukan suatu corak atau pola dalam data tersebut [1].

Fayyed, dkk (1996) mendefinisikan data mining adalah bagian integral dari penemuan pengetahuan dalam database atau dikenal dengan *Knowledge Discovery in Databases* (KDD) yang merupakan proses keseluruhan mengubah data mentah menjadi pola-pola data yang menarik yang merupakan informasi yang dibutuhkan oleh pengguna sebagai pengetahuan. Berikut adalah langkah-langkah yang terjadi dalam proses KDD :

1. *Selection*  
Dalam tahap ini dilakukan pemahaman terhadap permasalahan yang akan dicari solusinya melalui penemuan pengetahuan, pengetahuan sebelumnya yang relevan, penentuan tujuan dan pemilihan data yang akan dianalisis.
2. *Preprocessing*  
Dalam tahap ini dilakukan proses *preprocessing* data termasuk kegiatan pembersihan data.
3. *Transformation*  
Dalam tahap ini dilakukan pengurangan dan / atau perubahan tipe data menjadi standar, sehingga data siap untuk di presentasikan ke teknik-teknik *data mining*.
4. *Data mining*  
Dalam tahap ini dilakukan pemilihan tugas data mining yang sesuai dengan tujuan dan algoritma data mining untuk pencarian pola.
5. *Interpretation*  
Dalam tahap ini dilakukan visualisasi dan interpretasi terhadap pola yang ditemukan untuk dijadikan pengetahuan.

Tujuan dari proses *preprocessing* data adalah untuk mengubah data input mentah menjadi format yang sesuai untuk analisis selanjutnya. Langkah-langkah yang dilakukan antara lain dengan memperbaiki data yang 'kotor', memilih fitur-fitur dari data yang relevan dengan proses pengolahan selanjutnya. Karena banyak cara

dalam proses pengumpulan data dan penyimpanan data, maka proses preprocessing data mungkin akan memakan waktu yang lama dalam keseluruhan proses penemuan pengetahuan [2].

Oleh karena itu *preprocessing* data sangat penting karena dapat meningkatkan kualitas data serta meningkatkan akurasi dan efisiensi proses *mining* [1].

Menurut Han dan Kamber (2006) Terdapat beberapa teknik pada *data preprocessing* yaitu :

1. *data cleaning*, merupakan tahapan untuk mengisi data yang tidak lengkap, menghaluskan data *random error* atau *outlier*, serta memperbaiki data yang tidak konsisten
2. *data integration dan data transformation*, *Data integration* merupakan tahapan proses *data preprocessing* yang mengkombinasikan data dari berbagai bentuk sumber data menjadi sebuah tempat penyimpanan dalam *data warehousing*. Sumber data tersebut dapat terdiri dari *multiple database*, *data cubes*, atau *flat files*. *Data transformation* merupakan tahapan dimana data ditransformasikan dan dikonsolidasikan menjadi bentuk yang layak untuk *mining*.
3. *data reduction*, Teknik *data reduction* diaplikasikan untuk mengurangi representasi *dataset* sehingga berukuran lebih kecil. Metode yang dapat dilakukan pada teknik *data reduction* adalah :
  - a. *Data cube aggregation*, dimana operasi agregasi diaplikasikan pada data untuk membangun *data cube*.
  - b. *Dimension reduction*, dimana atribut atau dimensi yang tidak relevan dihapus.
  - c. *Data compression*, dimana data diubah atau diestimasi menjadi representasi data yang lebih kecil.
  - d. *Discretization and concept hierarchy generation*, dimana nilai data diubah kedalam bentuk *range* atau menjadi level yang lebih tinggi. Teknik *discretization* dapat digunakan untuk mengurangi jumlah nilai yang diberikan atribut kontinyu, dengan membagi *range* pada atribut menjadi suatu interval tertentu. Label interval dapat digunakan untuk merubah nilai data actual. Konsep hirarki untuk atribut numeric (atribut kontinyu), dapat dibangun berdasarkan analisis distribusi data.

Pada teknik *data reduction* dengan metode *Discretization and concept hierarchy generation* selain teknik *natural partition* dapat juga dilakukan teknik logika fuzzy dalam proses reduksi. Pada himpunan tegas (crisp), nilai keanggotaan suatu item  $x$  dalam suatu himpunan  $A$ , yang sering ditulis dengan  $\mu_A[x]$ , memiliki 2 kemungkinan, yaitu:

1. satu (1), yang berarti bahwa suatu item menjadi anggota dalam suatu himpunan,
2. nol (0), yang berarti bahwa suatu item tidak menjadi anggota dalam suatu himpunan.

Contoh, Jika diketahui:

$S = \{1, 2, 3, 4, 5, 6\}$  adalah semesta pembicaraan.

$A = \{1, 2, 3\}$

$B = \{3, 4, 5\}$

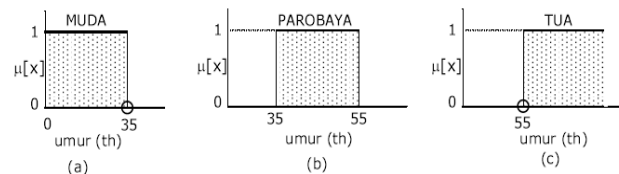
bisa dikatakan bahwa:

1. Nilai keanggotaan 2 pada himpunan  $A$ ,  $\mu_A[2]=1$ , karena  $2 \in A$ .
2. Nilai keanggotaan 3 pada himpunan  $A$ ,  $\mu_A[3]=1$ , karena  $3 \in A$ .
3. Nilai keanggotaan 4 pada himpunan  $A$ ,  $\mu_A[4]=0$ , karena  $4 \notin A$ .
4. Nilai keanggotaan 2 pada himpunan  $B$ ,  $\mu_B[2]=0$ , karena  $2 \notin B$ .
5. Nilai keanggotaan 3 pada himpunan  $B$ ,  $\mu_B[3]=1$ , karena  $3 \in B$ .

Misalkan variabel umur dibagi menjadi 3 kategori, yaitu:

1. MUDA umur  $< 35$  tahun
2. PAROBAYA  $35 \leq$  umur  $\leq 55$  tahun
3. TUA umur  $> 55$  tahun

Nilai keanggotaan secara grafis, himpunan MUDA, PAROBAYA dan TUA ini dapat dilihat pada Gambar 1.

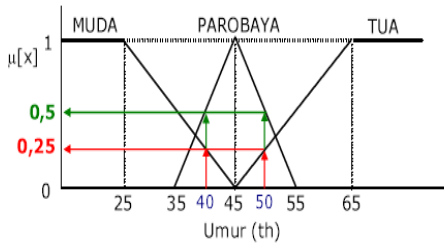


**Gambar 1 Himpunan: MUDA, PAROBAYA, dan TUA.**

Pada Gambar 1, dapat dilihat bahwa:

1. apabila seseorang berusia 34 tahun, maka ia dikatakan MUDA ( $\mu_{MUDA}[34]=1$ );
2. apabila seseorang berusia 35 tahun, maka ia dikatakan TIDAK MUDA ( $\mu_{MUDA}[35]=0$ );
3. apabila seseorang berusia 35 tahun kurang 1 hari, maka ia dikatakan TIDAK MUDA ( $\mu_{MUDA}[35 \text{ th} - 1 \text{ hr}]=0$ );
4. apabila seseorang berusia 35 tahun, maka ia dikatakan PAROBAYA ( $\mu_{PAROBAYA}[35]=1$ );
5. apabila seseorang berusia 34 tahun, maka ia dikatakan TIDAK PAROBAYA ( $\mu_{PAROBAYA}[34] = 0$ );
6. apabila seseorang berusia 35 tahun, maka ia dikatakan PAROBAYA ( $\mu_{PAROBAYA}[35]=1$ );
7. apabila seseorang berusia 35 tahun kurang 1 hari, maka ia dikatakan TIDAK PAROBAYA ( $\mu_{PAROBAYA}[35 \text{ th} - 1 \text{ hr}] = 0$ );

Seseorang dapat masuk dalam 2 himpunan yang berbeda, MUDA dan PAROBAYA, PAROBAYA dan TUA, dan sebagainya. Seberapa besar eksistensinya dalam himpunan tersebut dapat dilihat pada nilai keanggotaannya. Gambar 2 menunjukkan himpunan fuzzy untuk variabel umur.



**Gambar 2 Himpunan fuzzy untuk variabel Umur.**

Pada Gambar 2, dapat dilihat bahwa:

1. Seseorang yang berumur 40 tahun, termasuk dalam himpunan MUDA dengan  $\mu_{MUDA} [40] = 0,25$ ; namun dia juga termasuk dalam himpunan PAROBAYA dengan  $\mu_{PAROBAYA} [40] = 0,5$ .
2. Seseorang yang berumur 50 tahun, termasuk dalam himpunan MUDA dengan  $\mu_{TUA} [50] = 0,25$ ; namun dia juga termasuk dalam himpunan PAROBAYA dengan  $\mu_{PAROBAYA} [50] = 0,5$ .

Kalau pada himpunan crisp, nilai keanggotaan hanya ada 2 kemungkinan, yaitu 0 atau 1, pada himpunan fuzzy nilai keanggotaan terletak pada rentang 0 sampai 1. Apabila  $x$  memiliki nilai keanggotaan fuzzy  $\mu_A[x]=0$  berarti  $x$  tidak menjadi anggota himpunan A, demikian pula apabila  $x$  memiliki nilai keanggotaan fuzzy  $\mu_A[x] = 1$  berarti  $x$  menjadi anggota penuh pada himpunan A.

Himpunan fuzzy memiliki 2 atribut, yaitu:

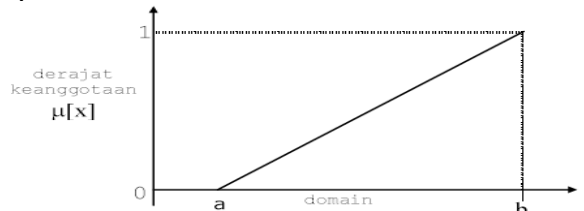
1. Linguistik, yaitu penamaan suatu grup yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami, seperti: MUDA, PAROBAYA, TUA.
2. Numeris, yaitu suatu nilai (angka) yang menunjukkan ukuran dari suatu variabel seperti: 40, 25, 50, dsb.

Ada beberapa hal yang perlu diketahui dalam memahami sistem fuzzy, yaitu:

1. Variabel fuzzy, merupakan variabel yang hendak dibahas dalam suatu sistem fuzzy. Contoh: umur, temperatur, permintaan, dan sebagainya.
2. Himpunan fuzzy, merupakan suatu grup yang mewakili suatu kondisi atau keadaan tertentu dalam suatu variabel fuzzy.
3. Semesta Pembicaraan, adalah keseluruhan nilai yang diperbolehkan untuk dioperasikan dalam suatu variabel fuzzy. Semesta pembicaraan merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai semesta pembicaraan dapat berupa bilangan positif maupun negatif. Adakalanya nilai semesta pembicaraan ini tidak dibatasi batas atasnya.
4. Domain, Domain himpunan fuzzy adalah keseluruhan nilai yang diijinkan dalam semesta pembicaraan dan boleh dioperasikan dalam suatu himpunan fuzzy. Seperti halnya semesta pembicaraan, domain merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai domain dapat berupa bilangan positif maupun negatif.

5. Fungsi Keanggotaan (membership function) adalah suatu kurva yang menunjukkan pemetaan titik-titik input data ke dalam nilai keanggotaannya (sering juga disebut dengan derajat keanggotaan) yang memiliki interval antara 0 sampai 1. Salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi. Fungsi yang biasa digunakan adalah Representasi Linear. Pada representasi linear, pemetaan input ke derajat keanggotaannya digambarkan sebagai suatu garis lurus. Bentuk ini paling sederhana dan menjadi pilihan yang baik untuk mendekati suatu konsep yang kurang jelas.

Ada 2 keadaan himpunan fuzzy yang linear. Pertama, kenaikan himpunan dimulai pada nilai domain yang memiliki derajat keanggotaan nol [0] bergerak ke kanan menuju ke nilai domain yang memiliki derajat keanggotaan lebih tinggi, yang ditunjukkan pada Gambar 3.



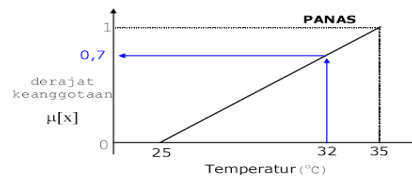
**Gambar 3 Representasi Linear Naik**

Persamaan Keanggotaan Gambar 3, ditunjukkan pada persamaan 1

$$\mu[x] = \begin{cases} 0; & x \leq a \\ (x - a) / (b - a); & a \leq x \leq b \\ 1; & x \geq b \end{cases} \dots\dots\dots (1)$$

Contoh Fungsi keanggotaan untuk himpunan PANAS pada variabel temperature ruangan seperti terlihat pada Gambar 4.

$$\mu_{PANAS}[32] = (32-25)/(35-25) = 7/10 = 0,7$$



**Gambar 4 Himpunan fuzzy: PANAS.**

Kedua, merupakan kebalikan yang pertama. Garis lurus dimulai dari nilai domain dengan derajat keanggotaan tertinggi pada sisi kiri, kemudian bergerak menurun ke nilai domain yang memiliki derajat keanggotaan lebih rendah, seperti yang ditunjukkan pada Gambar 5.



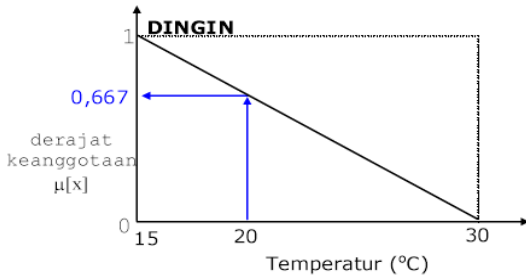
**Gambar 5 Representasi Linear Turun.**

Persamaan Keanggotaan Gambar 5, ditunjukkan pada persamaan 2.

$$\mu[x] = \begin{cases} (b-x)/(b-a); & a \leq x \leq b \\ 0; & x \geq b \end{cases} \dots\dots\dots(2)$$

Contoh Fungsi keanggotaan untuk himpunan DINGIN pada variabel temperatur ruangan seperti terlihat pada Gambar 6.

$$\mu_{DINGIN}[20] = (30-20)/(30-15) = 10/15 = 0,667$$



Gambar 6 Himpunan fuzzy: DINGIN.

Sementara untuk *natural partition* dapat dihitung menggunakan persamaan [3] :

$$\text{Min } \tau(a) = \min_{x \in T} a(x) \dots\dots\dots(3)$$

$$\text{Max } \tau(a) = \max_{x \in T} a(x) \dots\dots\dots(4)$$

jika akan membagi sebanyak k bagian maka :

$$\Delta_T^k(a) = \frac{\text{max}_T(a) - \text{min}_T(a)}{k} \dots\dots\dots(5)$$

**2. Pembahasan**

Secara umum *preprocessing* ini melewati beberapa tahap antara lain :

1. Data *transformation* dengan metode *normalization* yaitu mengubah data ke range spesifik. Metode ini dilakukan pada variabel gaji, variabel nilai Jaminan, dan variabel angsuran. Contoh penerapan *normalization* pada ketiga variabel tersebut adalah memberikan range tinggi, sedang atau rendah pada nilai variabel-variabel tersebut.
2. Data *reduction* dengan metode *Discretization and concept hierarchy generation*, yaitu mengurangi jumlah nilai yang diberikan atribut kontinyu dengan membagi *range* pada atribut menjadi suatu interval tertentu. Atribut yang diterapkan teknik ini adalah variabel gaji, variabel nilai jaminan, dan variabel angsuran. Contoh penerapan *Discretization and concept hierarchy generation* yaitu ketiga variabel tersebut diberikan nilai batas atas dan batas bawah untuk menentukan range tinggi, sedang dan rendah.

3. Data *reduction* dengan metode *compression*, yaitu menghilangkan beberapa variabel data dengan cara mengubah data dengan melakukan fungsi agregasi agar menjadi representasi data yang lebih kecil. Contohnya adalah variabel angsuran didapatkan dengan cara menjumlahkan variabel jumlah pinjam dengan variabel margin pembiayaan dibagi dengan tempo pinjam. Sehingga setelah dilakukan metode ini, variabel yang tersisa adalah variabel angsuran yang akan digunakan pada proses selanjutnya yaitu klasifikasi.

4. Data *Reduction* dengan Metode *Dimension reduction*, dimana atribut atau dimensi yang tidak relevan dihapus. Contoh penerapan metode *Dimension reduction* adalah variabel tujuan pinjam dihapus karena semua nilai tujuan pinjam sesuai syariah, sehingga tidak digunakan dalam proses klasifikasi karena hanya akan menambah nilai komputasi namun tidak menambah nilai akurasi sistem yang dibangun.

Misalkan, 6 data perbankan akan dilakukan *preprocessing*, yang hasilnya akan disimpan di database dengan nama “data bersih” yang ditunjukkan pada Tabel 1.

Tabel 1 DataHasil *Preprocessing*

Id	Pendidikan	Pekerjaan	Gaji	Angsuran	Nilai Jaminan	Kelas
1050070629	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
7044291538	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050060436	SMA	PNS Umum	Sedang	Sedang	Kecil	Tidak Aman
1050074990	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050049194	Sarjana	PNS Umum	Kecil	Kecil	Kecil	Aman
7043990416	Sarjana	PNS Umum	Sedang	Kecil	Kecil	Aman

**1. Pengujian kelas Prediksi dengan *Natural Partition***

Pada pengujian ini dilakukan perhitungan jumlah kelas prediksi berdasarkan pembagian level berdasarkan distribusi nilai pada variabel gaji, jaminan dan angsuran menjadi 3 level. Jumlah sampel data yang digunakan adalah 463 data debitur. Tabel 2 tertuang hasil pengujian kelas prediksi dengan *natural partition* berdasarkan besar kecilnya nilai variabel gaji, jaminan dan angsuran yang ditetapkan berdasarkan levelnya.

Tabel 2 Pengujian Kelas Prediksi Teknik Standar Deviasi

No	Level Gaji			Level Ansuran			Level Jaminan			Jumlah Sampel		
	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Aman	Tidak Aman	Unknown
1	1.000.000-	2.000.000-	3.000.000-	500.000-	1.500.000-	2.500.000-	0-	25.000.000-	50.000.000-	223	135	105
	2.000.000	3.000.000	4.000.000	1.500.000	2.500.000	3.500.000	25.000.000	50.000.000	75.000.000			
2	1.000.000-	3.000.000-	5.000.000-	500.000-	2.500.000-	4.500.000-	0-	50.000.000-	100.000.000-	302	150	11
	3.000.000	5.000.000	7.000.000	2.500.000	4.500.000	6.500.000	50.000.000	100.000.000	150.000.000			
3	1.000.000-	4.000.000-	7.000.000-	500.000-	3.500.000-	6.500.000-	0-	75.000.000-	150.000.000-	312	150	1
	4.000.000	7.000.000	10.000.000	3.500.000	6.500.000	9.500.000	75.000.000	150.000.000	225.000.000			
4	1.000.000-	5.000.000-	9.000.000-	500.000-	4.500.000-	8.500.000-	0-	100.000.000-	200.000.000-	312	150	1
	5.000.000	9.000.000	13.000.000	4.500.000	8.500.000	12.500.000	100.000.000	200.000.000	300.000.000			

### 2. Pengujian kelas Prediksi dengan Logika Fuzzy

Pada pengujian ini dilakukan perhitungan jumlah kelas prediksi berdasarkan logika fuzzy, pada variabel gaji, variabel angsuran dan variabel jaminan. Ketiga variabel ini dibagi ke dalam tiga himpunan yaitu kecil, sedang dan besar. Derajat keanggotaan yaitu bernilai 0, 0-1 dan 1. Setelah itu menentukan batasan dari masing-masing himpunan. Setelah proses pengaturan nilai derajat keanggotaan dan batasan dari setiap himpunan pada masing-masing variabel selesai. Hasil dari pengujian kelas prediksi dengan teknik logika fuzzy ditunjukkan pada Tabel 3.

Tabel 3 Pengujian Kelas Prediksi Teknik Logika Fuzzy

	Gaji									Jumlah Sampel			
	Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow	
	0	0.1	1	0	0.1	1	0	0.1	1				
>= 7.000.000	4.000.000	0 <= gaji	Gaji <= 10.000.000	7.000.000 <=	Gaji <= 10.000.000	Gaji <= 10.000.000	Gaji >= 13.000.000	0	11.000.000	0	312	150	1
<= 7.000.000	<= 7.000.000	4.000.000	atau gaji <= 13.000.000	<= 10.000.000	0	0	0	0	0	0			
0	0	0	>= 13.000.000	0	0	0	0	0	0	0			
	Jaminan									Jumlah Sampel			
	Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow	
	0	0.1	1	0	0.1	1	0	0.1	1				
>= 100.000.000	90.000.000	0 >= jaminan	Jaminan >= 150.000,00	100.000,00 >=	Jaminan <= 150.000,00	Jaminan <= 150.000,00	Jaminan >= 160,000,00	0	0	0	312	150	1
0	<= 100.000.000	90.000.000	0	Jaminan <= 160.000,00	0	0	0	0	0	0			
	Ansuran									Jumlah Sampel			
	Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow	
	0	0.1	1	0	0.1	1	0	0.1	1				
>= 3.500.000	2.000,00	0 <= Ansuran	Ansuran >= 3.500,000	2.000,00 <=	Ansuran <= 3.500.000	Ansuran <= 3.500.000	Ansuran >= 8.500.000	0	0	0	312	150	1
0	<= 3.500,00	2.000,00	Ansuran >= 4.000,000	<= 4.000,000	<= 3.500.000	<= 3.500.000	<= 5.000.000	0	0	0			

### 3. Pengujian Akurasi Model probabilitas kelas teknik Natural Partition

Pada pengujian ini dilakukan pengujian akurasi model probabilitas kelas terhadap data bersih yang terbentuk dengan berbagai range dengan tingkatan nilai tertentu. Seperti pada preprocessing, range nilai level dibagi dengan interval tertentu sampai mendapat nilai terbaik. Tabel 4 merupakan rincian hasil pengujian model probabilitas kelas dengan teknik *natural partition* dengan berbagai range nilai level.

Tabel 4 Pengujian Akurasi Model Probabilitas Kelas teknik *natural partition*

No	Level Gaji			Level Ansuran			Level Jaminan			Probabilitas Kelas	
	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Aman	Tidak Aman
1	1.000.000-	2.000.000-	3.000.000-	500.000-	1.500.000-	2.500.000-	0-	25.000.000-	50.000.000-	0,62	0,38
	2.000.000	3.000.000	4.000.000	1.500.000	2.500.000	3.500.000	25.000.000	50.000.000	75.000.000		
2	1.000.000-	3.000.000-	5.000.000-	500.000-	2.500.000-	4.500.000-	0-	50.000.000-	100.000.000-	0,67	0,33
	3.000.000	5.000.000	7.000.000	2.500.000	4.500.000	6.500.000	50.000.000	100.000.000	150.000.000		
3	1.000.000-	4.000.000-	7.000.000-	500.000-	3.500.000-	6.500.000-	0-	75.000.000-	150.000.000-	0,68	0,32
	4.000.000	7.000.000	10.000.000	3.500.000	6.500.000	9.500.000	75.000.000	150.000.000	225.000.000		
4	1.000.000-	5.000.000-	9.000.000-	500.000-	4.500.000-	8.500.000-	0-	100.000.000-	200.000.000-	0,68	0,32
	5.000.000	9.000.000	13.000.000	4.500.000	8.500.000	12.500.000	100.000.000	200.000.000	300.000.000		

### 4. Pengujian Akurasi Model probabilitas kelas teknik Logika Fuzzy

Pada pengujian ini dilakukan pengujian akurasi model probabilitas kelas terhadap data bersih yang terbentuk dengan teknik logika fuzzy. Pada variabel gaji, variabel angsuran dan variabel jaminan dibagi ke dalam tiga himpunan yaitu kecil, sedang dan besar. Setelah itu mengatur derajat keanggotaan yaitu bernilai 0, 0-1 dan 1. Terakhir menentukan batasan dari masing-masing himpunan. Tabel 5 merupakan rincian hasil pengujian model probabilitas kelas dengan teknik logika Fuzzy.

Tabel 5 Pengujian Akurasi Model Probabilitas Kelas teknik Logika Fuzzy

Gaji									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0..1	1	0	0..1	1	0	0..1	1		
>= 7.000.000	4.000.000 <= gaji	0 <= gaji	Gaji <= 10.000.000	7.000.000 <= gaji	10.000.000 <= gaji	Gaji <= 10.000.000	10.000.000 <= gaji	Gaji >= 13.000.000	0,68	0,32
0	<= 7.000.000	4.000.000	atau gaji >= 13.000.000	0	13.000.000 <= 10.000.000	0	11.000.000	00		

Jaminan									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0..1	1	0	0..1	1	0	0..1	1		
>= 100.000.000	90.000.000 >= jaminan	0 >= jaminan	Jaminan >= 150.000,00	100.000,00 >= Jaminan	150.000,00 >= Jaminan	Jaminan <= 150.000,000	150.000,000 >= Jaminan	Jaminan >= 160.000,000	0,68	0,32
000	<= 100.000.000	90.000.000	0	Jaminan <= 160.000,00	<= 150.000,000	00	Jaminan <= 160.000,000	160.000,000		

Anesuran									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0..1	1	0	0..1	1	0	0..1	1		
Anesuran >= 3.500.000	2.000,00 <= Anesuran	0 <= Anesuran	Anesuran >= 3.500,000	2.000,000 <= Anesuran	3.500,000 <= Anesuran	Anesuran <= 3.500.000	3.500.000 <= anesuran	anesuran >= 8.500.000	0,68	0,32
0	<= 2.000,000	2.000,000	Anesuran >= 4.000,000	>= 4.000,000	>= 3.500,000	0	<= 5.000.000	0		

yang dapat diolah, semakin besar range maka semakin besar data yang dapat diolah dan sebaliknya. Hal ini berlaku baik menggunakan logika fuzzy maupun *natural partition*

- Selain preprocessing data, ada faktor lain sebagai penentu meningkatnya nilai akurasi, misalnya pemilihan algoritma yang pas dengan jenis data yang akan diolah

**Saran**

penelitian ini masih memiliki keterbatasan yang dapat dijadikan acuan untuk pengembangan dimasa yang akan datang, sehingga dapat disarankan beberapa hal sebagai berikut :

- Perlu dilakukan analisis hubungan antar variabel, apakah sebagai salah satu penentu meningkatnya nilai akurasi atau tidak.
- Perlu dilakukan perbandingan dengan teknik lain pada proses *preprocessing* variabel continue
- Perlu dilakuakn uji akurasi dengan teknik lainnya.

**Daftar Pustaka**

[1] Han, J., & Kamber, M., *Data Mining Concept and Technique*, San Fransisco: Morgan Kaufman Publisher, 2006.  
 [2] Tan, P. N., Stenbach, M., & Kumar, V., *Introduction to Data Mining*, Boston: Pearson Education, 2006.  
 [3] Cichosz, P, *Data Mining Algorithms : Explained Using R*, United Kingdom : John Wiley& Sons, Ltd, 2015

**5. Pengujian akurasi menggunakan Cross Validation dan Bootstrapping Validation**

Teknik	Akurasi	
	Cross Validation	Bootstrapping Validation
Logika Fuzzy	66.98 %	66.97 %
Natural Partition	66.95 %	66.79 %

**3. Kesimpulan**

Dari penelitian yang telah dilakukan terdapat beberapa kesimpulan yaitu :

- Dari hasil pengujian dengan menggunakan teknik *natural partition* dan logika fuzzy pada proses preprocessing, dapat disimpulkan bahwa hasil data yang dapat diolah adalah sama. Namun uji akurasi menggunakan Cross validation dan Bootstrapping Validation hasilnya berbeda yaitu lebih tinggi teknik logika fuzzy
- Dari hasil pengujian *training* yang dilakukan, dapat disimpulkan bahwa semakin banyak data yang dapat diolah maka semakin besar nilai probabilitas kelas.
- Pembagian range nilai sangat berpengaruh terhadap data

**Biodata Penulis**

**Sumarni Adi**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2009. Memperoleh gelar *Master Of Computer Science(M.Cs)* Program Pasca Sarjana Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2013. Saat ini menjadi Dosen di jurusan S1 Teknik Informatika STMIK AMIKOM Yogyakarta. Penelitian-penelitian dan publikasinya banyak terdapat di bidang data mining, information retrieval dan teori komputasi.