

## UJI AKURASI KLASIFIKASI DAN VALIDASI DATA PADA PENGGUNAAN METODE MEMBERSHIP FUNCTION DAN ALGORITMA C4.5 DALAM PENILAIAN PENERIMA BEASISWA

**Ade Surya Budiman**  
Program Studi Teknik Komputer  
AMIK BSI Jakarta  
Email: ade.aum@bsi.ac.id

**Xanty Adhi Parandani**  
Program Studi Manajemen Informatika  
AMIK BSI Bogor  
Email: xanty.xip@bsi.ac.id

### ABSTRAK

Data yang dikumpulkan dalam proses seleksi penerima beasiswa memiliki variabel berbentuk nominal dan numerik serta terdiri atas banyak *item*, sehingga diperlukan metode yang tepat untuk melakukan klasifikasi data secara akurat dan memastikan validitas data yang dipergunakan dalam penilaian penerima beasiswa. Dalam penelitian ini dilakukan pengujian terhadap data dengan dua metode, yaitu (i) Integrasi *Membership Function* dengan Algoritma C4.5, dan (ii) Penerapan Algoritma C4.5 secara langsung pada data. Pengujian dilakukan dengan jenis *10-fold cross validation* dan pengujian kebenaran *instances (correctly/incorrectly)*. Dari hasil perhitungan dan pengujian diperoleh nilai *Mean Square Error (MAE)* dan *Root Mean Square Error (RMSE)* yang lebih rendah pada metode Integrasi *Membership Function* dengan Algoritma C4.5, yaitu masing-masing 0,132 (MAE) dan 0,2714 (RMSE). Untuk pengujian *Classified Instances*, diperoleh persentase kebenaran (*correctly*) yang lebih tinggi pada metode Integrasi *Membership Function* dengan Algoritma C4.5, yaitu sebesar 92,8571%.

**Kata kunci:** klasifikasi data, validitas data, *membership function*, algoritma C4.5, seleksi beasiswa.

### ABSTRACT

*Data collected for grantee selection process has nominal variables and numerical variables which consists of many items. Required the appropriate method to classify the data properly and ensure the validity of the data used in the assessment of the grantee. In this study, conducted tests on the data by two methods, namely (i) Integration of Membership Function and Algorithm C4.5, and (ii) Implementation of C4.5 algorithms directly on the data. Testing is done with a kind of 10-fold cross validation and testing truth instances (correctly/incorrectly). From the calculation and testing of obtained Mean Square Error (MAE) and Root Mean Square Error (RMSE) were lower in the method Integration of Membership Function and Algorithm C4.5, respectively 0.132 (MAE) and 0.2714 (RMSE). For testing Classified Instances, obtained by percentage of truth (Correctly) higher on methods Integration Membership Function Algorithm C4.5, in the amount of 92.8571%.*

**Keywords:** data classification, data validity, *membership function*, C4.5 algorithm, scholarship selection.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Program Beasiswa telah banyak diluncurkan oleh beragam lembaga, baik pemerintah maupun institusi non pemerintah. Seleksi terhadap calon penerima beasiswa diberikan dengan beragam indikator penilaian, diantaranya nilai akademik. Namun, terdapat juga beasiswa yang menitikberatkan kepada kondisi dan status sosial ekonomi calon penerima beasiswa. Salah satu jenis beasiswa yang menitikberatkan kepada kondisi dan status sosial ekonomi siswa, adalah Beasiswa Bantuan Khusus Siswa Miskin / Bantuan khusus Murid Miskin (BKSM/BKMM). BKSM/BKMM banyak diberikan bagi siswa/murid sekolah di daerah-daerah.

SMK PGRI Ploso merupakan sebuah sekolah lanjutan kejuruan yang terletak di Kabupaten Jombang Jawa Timur. Sekolah ini secara rutin menyalurkan beasiswa jenis BKSM bagi siswanya. Mengacu kepada demografis lokasi sekolah yang tergolong wilayah berkembang, dimana banyak siswa yang orangtuanya berpenghasilan menengah ke bawah, maka penilaian kelayakan calon penerima beasiswa harus benar-benar objektif dan dapat dipertanggungjawabkan. Indikator penilaian diperhitungkan dari data siswa dan data mengenai orang tua siswa. Data

dan informasi tersebut ada yang berupa variabel numerik (misal, jumlah tanggungan orang tua), ada pula yang berupa variabel nominal (misal, status kepemilikan tempat tinggal). Mengacu kepada variabel yang berbeda tersebut, maka perlu perhitungan yang tepat sehingga kedua variabel dapat dihitung secara tepat.

Jayus Priyana dan Agus Maman Abadi dalam penelitiannya mengenai peramalan suhu udara [1], diketengahkan proses peramalan suhu udara di Yogyakarta yang didasarkan pada model *fuzzy* dengan metode *table look-up scheme*. Penelitian yang dilakukan menghasilkan prediksi hasil pengolahan data yang lebih baik ketika menggunakan fungsi keanggotaan (*Membership Function*) Gaussian sebagai mesin inferensi minimum yang dikombinasikan dengan model *fuzzy* dengan dua masukan.

Algoritma C4.5 sebagai salah satu algoritma dalam pembentukan pohon keputusan, dipergunakan dalam sebuah penelitian untuk melakukan pengujian akurasi data Diabetes Indian Pima [2]. Pengujian dilakukan menggunakan metode analisa *3-folds Cross Validation*. Dari hasil pengujian diperoleh akurasi data mencapai 78,9%.

## 1.2 Sistem Pengambil Keputusan

Sistem Pendukung Keputusan (*Decision Support System/DSS*) adalah sistem berbasis komputer yang interaktif, yang membantu pengambil keputusan memanfaatkan data dan model untuk menyelesaikan masalah-masalah yang tak terstruktur [3]. Dalam definisi yang lain disebutkan, *Decision Support System* merupakan suatu sistem yang ditujukan untuk mendukung manajemen pengambilan keputusan, dimana sistem ini berbasis model yang terdiri dari prosedur-prosedur dalam pemrosesan data dan pertimbangannya untuk membantu user dalam mengambil keputusan [4].

*Decision Support System* dapat disimpulkan sebagai sebuah sistem yang mengintegrasikan seluruh sumber daya yang dimiliki oleh suatu organisasi untuk memberikan dukungan dalam pengambilan keputusan (*decision support*) bahkan melakukan proses pengambilan keputusan (*decision making*) itu sendiri.

## 1.3 Algoritma C4.5

Algoritma C4.5 ini merupakan perluasan / pengembangan (*extention*) dari algoritma ID3. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah *decision Tree* berdasarkan *training data* yang telah disiapkan [5]. Pada algoritma C4.5 terdapat proses generalisasi dari setiap aturan dengan cara membuang/mengabaikan kondisi awal (*precondition*), sehingga akurasi klasifikasi akan meningkat. Selanjutnya, aturan-aturan tergeneralisasi tersebut diurutkan berdasarkan keakuratannya, untuk digunakan pada uji klasifikasi yang akan datang.

Pada algoritma C4.5, model yang dihasilkan oleh proses “belajar” dari *training data* berupa sebuah pohon keputusan. Pohon keputusan ini kemudian dapat dimanfaatkan untuk memprediksi kelas dari kasus yang baru. Untuk mengkonstruksi pohon, algoritma C4.5 membaca seluruh sampel data pelatihan dari *storage* dan memuatnya ke memori. Dalam proses komputasinya, algoritma C4.5 menelaah sampel demi sampel di memori untuk mengkonstruksi pohon (prinsip kerja C4.5 adalah *memory based approach*). Pemuatan sampel ke memori dimaksudkan untuk menghindari pembacaan sampel ke *storage* yang berulang-ulang, yang dapat mengakibatkan tingginya akses ke *input-output* dan memperburuk waktu eksekusi algoritma [6].

## 1.4 Pohon Keputusan (*Decision Tree*)

Pohon keputusan (*decision Tree*) adalah sebuah konstruksi yang berhubungan dengan metode klasifikasi dalam proses *data mining*. Fungsi klasifikasi akan menghasilkan pengetahuan (*knowledge*) yang direpresentasikan oleh pohon keputusan tersebut [7]. Pohon keputusan (*decision Tree*) juga didefinisikan sebagai struktur pohon yang berperan seperti flowchart dimana setiap *node* merupakan sebuah pengujian terhadap sebuah nilai atribut, setiap *branch* merepresentasikan hasil dari pengujian tersebut, dan *tree leaves* merepresentasikan kelas-kelas atau sebaran-sebaran kelas. Pohon keputusan dapat dengan mudah diubah menjadi aturan-aturan klasifikasi [8].

Sebuah pohon keputusan dapat juga didefinisikan sebagai sebuah peta dari proses yang memiliki alasan (sebab akibat). Pohon keputusan mendeskripsikan suatu himpunan data (*data set*) melalui struktur yang mirip pohon [9]. Pohon keputusan sangat baik untuk menyelesaikan permasalahan klasifikasi data. Keuntungan utama dari suatu pohon keputusan adalah kemampuannya untuk melakukan visualisasi data.

Konstruksi pohon keputusan dapat dibentuk dengan mempergunakan perhitungan *Information Gain* (persamaan 1) dan *Entropy* (persamaan 2). Perhitungan *Entropy* (E) menjadi masukan dalam perhitungan *Information Gain* (I) untuk setiap *Fuzzy Sets* (persamaan 3). Persamaan-persamaan tersebut merupakan bagian dari algoritma C4.5 untuk mencapai pembentukan pohon keputusan pada klasifikasi data. Nilai *Information Gain* menentukan posisi dan peran dari setiap atribut pada pohon keputusan. *Information Gain* tertinggi adalah puncak tertinggi dari pohon keputusan atau dapat dikatakan sebagai penentu dari prioritas aturan (*rule*) yang terbentuk untuk sistem pendukung keputusan. *Information Gain* digunakan sebagai penentu batas kepentingan sebuah atribut ataupun sebagai refleksi dari kualitas suatu atribut [10][11].

Berikut adalah langkah-langkah perhitungan *Information Gain* (I) dan *Entropy* (E) yang merupakan representasi algoritma pembentukan pohon keputusan *fuzzy* dengan algoritma ID3/C4.5 [11].

- a. Tentukan komponen himpunan dari data,  
 $U = \{1, \dots, 70\}$   
 $C = \{\text{Nilai, Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan, Jumlah Tanggungan}\}$   
 $D = \{\text{Beasiswa}\}$  ;  $d_1 = \text{Ya} = 50$  ;  $d_2 = \text{Tidak} = 20$  ;  $m=2$   
 $J = \{U, C \cup D\}$
- b. *Information Gain* (I) dari atribut keputusan (D) dirumuskan:  

$$I(S_1, \dots, S_m) = - \sum_{i=1}^m p_i \log_2 p_i \tag{1}$$
- c. *Entropy* (E) dari suatu atribut (c) dirumuskan:  

$$E(C_i) = \sum_{j=i}^v \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \tag{2}$$
- d. *Information Gain* (I) untuk setiap *Fuzzy Sets* (S) dirumuskan:  

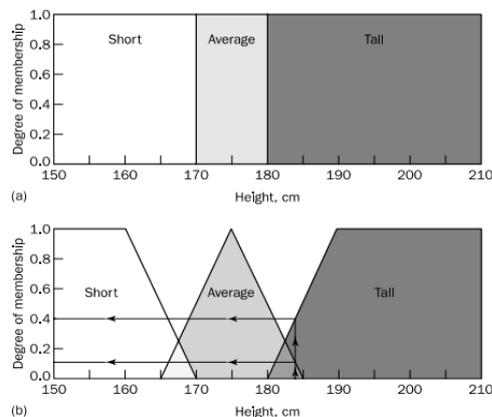
$$I(S_{1j}, \dots, S_{mj}) = \sum_{i=1}^m p_{ij} \log_2 p_{ij} \tag{3}$$

dimana,  
 $U$  = himpunan data sampel,  
 $C$  = himpunan atribut kondisi,  
 $D$  = atribut keputusan,  
 $d$  = kelas pembeda pada atribut keputusan  
 $m$  = jumlah kelas pembeda (*Fuzzy Set*),  
 $J$  = sistem representasi pengetahuan data,  
 $p_i$  = probabilitas sebuah nilai dalam suatu kelas  $i$ ,  
 $S$  = Kelas (*Fuzzy Set*)  $\rightarrow S_j$  = Sub Himpunan (*subset*)  
 $S_{ij}$  = jumlah sampel yang masuk kedalam kelas  $d_i$  dalam *subset*  $j$   
 $c$  = atribut kondisi (*Fuzzy Set*)
- e. Proses perhitungan *Information Gain* dan *Entropy* akan diulangi beberapa tahap hingga mencapai *node* terminasi.

### 1.5 Logika Fuzzy dan Membership Function

Logika *fuzzy* pertama kali dicetuskan oleh Dr. Lotfi Zadeh, Professor pada *Electrical Engineering Department, University of California*, pada tahun 1965, melalui sebuah paper berjudul "*Fuzzy Sets*" atau "Himpunan *Fuzzy*". *Fuzzy Logic System* dapat dikatakan sebagai *blackbox* untuk menghubungkan data yang tidak tepat/tidak presisi (*imprecise data*) dan pernyataan-pernyataan yang bernilai samar (*vague statements*) – misalnya: rendah, sedang, dan tinggi - untuk menghasilkan keputusan-keputusan (*decisions*) yang bisa diterima oleh pengguna/user dalam dunia nyata (*real world*) [12]. Fungsi Keanggotaan (*Membership Function*) adalah suatu kurva yang menunjukkan pemetaan titik-titik input data kedalam nilai keanggotaannya (derajat keanggotaan) yang memiliki interval antara 0 dan 1 [13].

Derajat keanggotaan (*degree of membership*) itu sendiri adalah nilai sebuah elemen dalam himpunannya, dimana nilai tersebut pada himpunan *fuzzy* terletak antara 0 sampai dengan 1, sedangkan pada himpunan *crisp* hanya bernilai 1 atau 0. Prinsip logika *fuzzy* yang mampu menjembatani perbedaan nilai pada garis batas derajat keanggotaan, membuat sebuah elemen dari himpunan *fuzzy* bisa memiliki wilayah keanggotaan yang berbeda. Hal ini dideskripsikan didalam Gambar 1.



Gambar 1. Representasi Fungsi Keanggotaan: (a) Crisp; (b) Fuzzy [9]

## 2. METODOLOGI PENELITIAN

### 2.1 Metode Pengumpulan Data

Dalam penelitian ini, peneliti mengambil data dan pembahasan melalui metode:

- a. Metode Wawancara  
Untuk mendapatkan pemahaman mengenai aturan yang perlu disusun dalam menentukan calon penerima beasiswa, tim penulis melakukan wawancara dengan Kepala Sekolah dan Wakil Kepala Sekolah bidang kesiswaan serta Ketua Panitia Seleksi Calon Penerima Beasiswa di SMK PGRI Ploso, Jombang, Jawa Timur.
- b. Metode Studi Kepustakaan  
Untuk melakukan studi perbandingan metode dan mencari solusi dari permasalahan penelitian, tim peneliti melakukan studi kepustakaan yang bersumber dari internet, buku teks dan artikel-artikel terkait dari jurnal ilmiah yang valid.

### 2.2 Teknik Analisa Data

Untuk memperoleh hasil yang dilakukan, peneliti melakukan serangkaian tahapan pengolahan data sebelum diperhitungkan dan dianalisa dengan metode yang ditentukan. Secara garis besar, tahapan analisa adalah sebagai berikut:

- a. Pengumpulan data mentah yang bersumber dari pihak sekolah, masih berupa data tertulis/tercetak pada buku data sekolah.
- b. Penyusunan data agar siap diolah (digitalisasi data) dengan menggunakan Microsoft Excel.
- c. Data Tuning dan Data Cleansing. Bertujuan untuk menghilangkan data yang tidak valid atau kekeliruan saat dimasukkan ke buku data siswa. Hal ini untuk mencegah kesalahan penghitungan data oleh sistem dan algoritma yang digunakan.
- d. Mengolah data sesuai algoritma yang ditentukan. Menguji hasil yang diperoleh untuk mendapatkan perbandingan akurat.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Penelitian

Distribusi data hasil pra proses yang menjadi data utama dalam penelitian ini, secara garis besar diperlihatkan dalam Tabel 1.

**Tabel 1. Hasil pengolahan data siswa**

<i>Variabel</i>	<i>Anggota</i>	
Jumlah Data	70	
Nilai Rata-Rata Siswa	Nilai 75	8
	Nilai 76	22
	Nilai 77	13
	Nilai 78	15
	Nilai 79	10
	Nilai 80	2
	Pinjam	7
Biaya Sekolah	Kontrak	6
	Sendiri	57
	Non Permanen	14
Bangunan Rumah	Semi Permanen	10
	Permanen	46
	Tidak Punya	6
Kendaraan	Sepeda Mini	26
	Sepeda Motor	38
	1 Orang	38
Jumlah Tanggungan Orang Tua	2 Orang	25
	3 Orang	6
	4 Orang	1
Keputusan Beasiswa	Tidak	20
	Ya	50

### 3.2 Pengolahan dan Pengujian Metode

#### 3.2.1 Metode 1 (Integrasi Membership Function dan Algoritma C4.5)

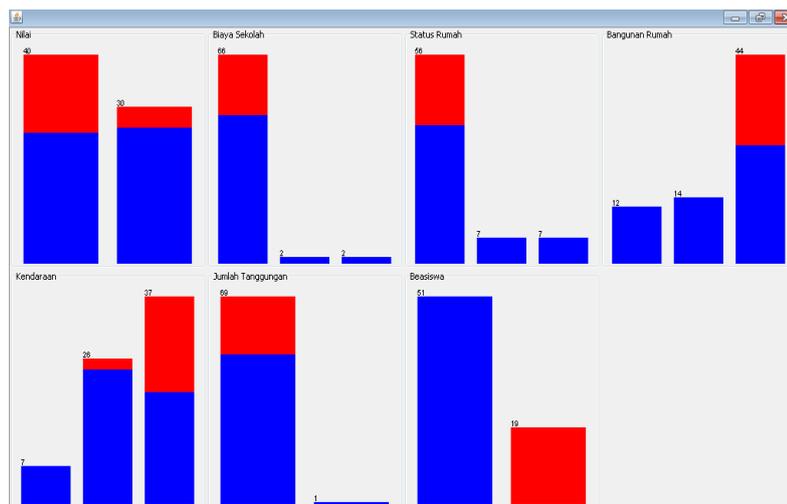
Sebagaimana telah dijelaskan didalam metode penelitian, terdapat 2 jenis variabel data yang digunakan dalam data calon penerima beasiswa, yaitu data kualitatif (nominal) dan data kuantitatif (numerik). Variabel data kuantitatif (numerik) yang berupa atribut nilai siswa dan jumlah tanggungan, diolah dengan memberikan derajat keanggotaan (*Membership Function*) kepada setiap nilai siswa dan jumlah tanggungan. Dengan demikian variabel nilai siswa dan jumlah tanggungan tidak lagi berupa numerik, namun berupa nominal. Pembentukan himpunan *Fuzzy*, semesta pembicaraan dan domain dari setiap atribut diatas diperlihatkan dalam Tabel 2.

**Tabel 2. Pembentukan membership function**

<i>Fungsi</i>	<i>Variabel</i>	<i>Himpunan Fuzzy</i>	<i>Semesta Pembicaraan</i>	<i>Domain</i>
Input	Nilai	Sedang	[65-100]	[65 - 77]
		Tinggi		[75- 100]
	Biaya Sekolah	Ibu	[0-10]	[0-4]
		Wali		[3-7]
		Orang Tua		[6-10]
	Status Rumah	Sendiri	[0-10]	[0-4]
		Kontrak		[3-7]
		Pinjam		[6-10]
	Bangunan Rumah	Permanen	[0-10]	[0-4]
		Semi Permanen		[3-7]
Non Permanen		[6-10]		
Kendaraan	Sepeda Motor	[0-10]	[0-4]	
	Sepeda		[3-7]	
	Tidak Ada		[6-10]	
Jumlah Tanggungan	Sedikit ( $\leq 2$ )	[0-10]	[0-4]	
	Sedang (2 - 4)		[3-7]	
	Banyak ( $\geq 4$ )		[6-10]	
Output	Beasiswa	Ya Tidak	[0 dan 0,5]	[0] [0,5]

Berikut adalah langkah-langkah yang dilakukan untuk melakukan eksplorasi data pada metode ini dari:

- Proses ini membutuhkan proses nominalisasi (perubahan dari variabel numerik/kuantitatif menjadi variabel nominal/kualitatif). Proses tersebut dilakukan melalui penentuan *Membership Function* yang diberikan kepada setiap data pada variabel kuantitatif. Penentuan nilai nominal tersebut mengacu kepada sistem pemberian derajat keanggotaan yang dibuat didalam Tabel 2. Dengan demikian variabel input untuk klasifikasi data seluruhnya akan berupa variabel kualitatif (nominal).
- File* dibuat didalam Microsoft Office Excel 2007 dengan jenis file \*.xlsx. Agar data bisa dibaca dan diproses oleh WEKA, maka dilakukan perubahan dan modifikasi file kedalam bentuk *comma separated values* (pemisah antar nilai menggunakan tanda baca koma) dengan jenis *file* yang digunakan adalah \*.csv.
- Gambar 1 memperlihatkan pembacaan atribut data, yaitu pada WEKA. Pada gambar warna biru menunjukkan kelompok siswa penerima beasiswa dan warna merah sebagai penunjuk kelompok siswa yang bukan penerima beasiswa. Batang yang diperlihatkan didalam grafik, terdiri atas variabel nilai, sumber biaya sekolah, status kepemilikan rumah, bahan bangunan rumah, kepemilikan kendaraan, jumlah tanggungan orang tua, dan keputusan diterima/tidak diterima beasiswa (biru = penerima; merah = bukan penerima)

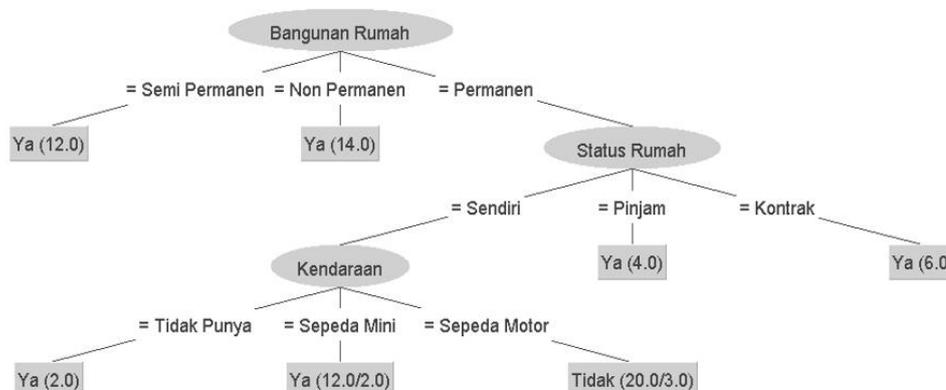


Gambar 1. Tampilan Atribut *Input vs Output* pada Integrasi MF-Algorithm C4.5

- d. Selanjutnya dilakukan proses klasifikasi terhadap data yang dimasukkan dengan menggunakan pengklasifikasi (*classifier*) berupa *trees classifier* berbasis algoritma J.48, yang merupakan implementasi dari algoritma C4.5 pada WEKA.
- e. Proses berikutnya adalah melakukan uji validasi dengan menggunakan *test option* berupa *10-folds cross validation*. Dari hasil pengujian diperoleh hasil sebagai berikut:

$$\begin{aligned} \text{Root Mean Square Error (RMSE)} &= 0.2714 \\ \text{Mean Absolut Error (MAE)} &= 0.132 \end{aligned}$$

- f. Proses terakhir adalah pembentukan pohon keputusan (*decision Tree*) yang dihasilkan (*generated*) oleh WEKA. Pohon keputusan ini, diperlihatkan didalam Gambar 2.



Gambar 2. Pohon Keputusan (*Decision Tree*) pada Integrasi MF – Algoritma C4.5

Dari pohon keputusan tersebut, bisa dilihat bahwa *node* tertinggi adalah bangunan rumah, dengan demikian kondisi bangunan rumah dijadikan sebagai faktor yang paling menentukan (*noctah dasar*) dalam pemberian beasiswa. Selanjutnya terdapat *node* keputusan A yaitu status rumah dan *node* keputusan B adalah kendaraan

### 3.2.2 Metode 2 (Penerapan Algoritma C4.5)

Perbedaan Metode 1 dengan metode 2 terdapat pada perlakuan terhadap variabel kuantitatif (numerik) yaitu data nilai siswa dan jumlah tanggungan, dimana dalam metode ini kedua jenis variabel data yaitu variabel data kualitatif dan variabel data kuantitatif, langsung diolah dengan menggunakan *machine learning* WEKA. Berikut adalah langkah-langkah yang dilakukan untuk melakukan klasifikasi data dalam WEKA dari data mentah hingga diperoleh pohon keputusan (*decision Tree*):

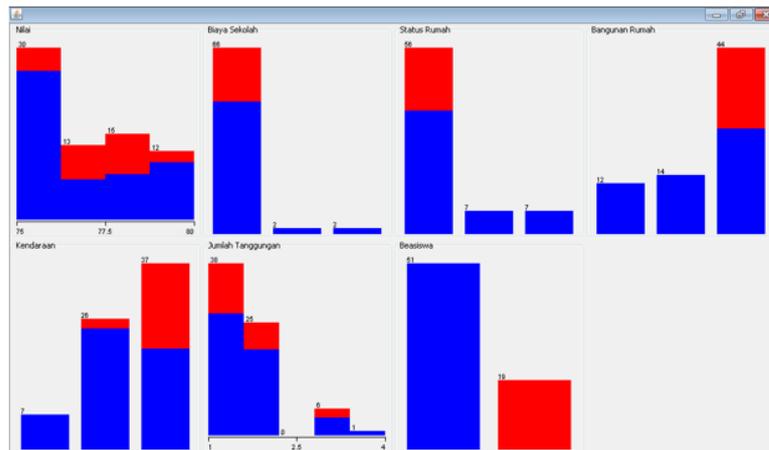
- a. Dilakukan *Preprocessing* terhadap data mentah, untuk memisahkan atribut yang tidak diperlukan dan tidak relevan dengan proses pengolahan data.

- b. File dibuat didalam Microsoft Office Excel 2007 dengan jenis file adalah \*.xlsx. Agar data bisa dibaca oleh WEKA, maka dilakukan perubahan dan modifikasi file kedalam bentuk *comma separated values* (pemisah antar nilai menggunakan tanda baca koma) dengan jenis file yang digunakan adalah \*.csv.
- c. Selanjutnya, WEKA akan membaca jumlah *instances* (data) yang dimasukkan dan atribut yang melekat pada *instances* tersebut.

*Instances: 70*

*Attributes: 8*

Gambar 3 memperlihatkan pengelompokan data masukan dibandingkan dengan data keluaran berdasarkan atribut yang terdapat pada data. Warna biru memperlihatkan kelompok siswa yang mendapatkan beasiswa, sedangkan warna merah memperlihatkan kelompok siswa yang tidak mendapatkan beasiswa



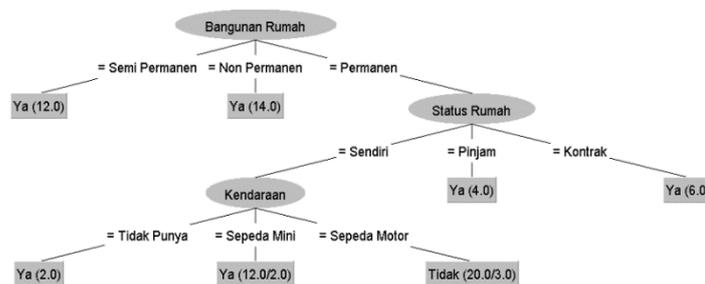
**Gambar 3. Tampilan Setiap Atribut *Input vs Output* Menggunakan WEKA**

- d. Langkah berikutnya adalah melakukan proses klasifikasi data dengan menggunakan classifier jenis Tree berbasis algoritma J48.
- e. Untuk uji validasi (test options) digunakan 10-folds cross validation. Cross validation Folds, yaitu menguji dengan membagi Training Data menjadi beberapa bagian misalkan k bagian (subset).
- f. Dari hasil pengujian diperoleh hasil sebagai berikut:

*Root Mean Square Error (RMSE)* = 0.2914

*Mean Absolut Error (MAE)* = 0.1434

- g. Proses terakhir adalah pembentukan pohon keputusan (*decision Tree*) yang dihasilkan (*generated*) oleh WEKA, seperti yang diperlihatkan dalam gambar 4.



**Gambar 4. Pohon Keputusan (*Decision Tree*) Pada Algoritma C4.5**

Dari pohon keputusan yang dihasilkan, *node* tertinggi adalah bangunan rumah, sehingga kondisi bangunan rumah dijadikan sebagai faktor yang paling menentukan (noktah dasar) dalam pemberian beasiswa. Selanjutnya terdapat *node* keputusan A yaitu status rumah dan *node* keputusan B adalah kendaraan.

### 3.3 Evaluasi dan Analisa

#### 3.3.1 Hasil Uji Cross Validation

Dalam Tabel 3 diperlihatkan perbandingan hasil uji validasi data yang telah dirincikan dalam subbab sebelumnya.

**Tabel 3. Perbandingan hasil uji cross validation**

<i>Jenis Error</i>	<i>Metode 1</i>	<i>Metode 2</i>
<i>Mean Absolut Error</i>	0.132	0.1434
<i>Root Mean Square Error</i>	0.2714	0.2914

Hasil yang diperoleh diatas, diperoleh dengan menggunakan 2 metode yang berbeda, yaitu (1) *Membership Function* dengan algoritma C4.5 (diimplementasi oleh algoritma J.48 pada WEKA), dan (2) Penerapan algoritma C4.5 (pada WEKA).

Pada Metode 1, diperoleh nilai *error* yang cukup rendah. Dalam metode ini data diperlakukan secara berbeda dengan data dalam metode pertama. Variabel data kualitatif (nominal) langsung diarahkan kedalam WEKA untuk diolah, sedangkan data kuantitatif (numerik) diproses terlebih dahulu oleh sistem *Fuzzy*. Variabel yang bersifat numerik, dimasukkan kedalam sistem *Fuzzy* untuk memperoleh fungsi keanggotaan (*Membership Function*) terlebih dahulu sebelum diolah oleh WEKA. Hal ini membuat variabel data numerik akan memiliki kekuatan nilai dan cakupan keanggotaan yang lebih fokus daripada variabel nominal. Terbukti dari hasil yang diperoleh dimana *error* yang diperoleh masih lebih baik daripada *error* pada metode 2.

Metode 2 memusatkan pengolahan data dari data olahan awal hingga terbentuk keluaran didalam WEKA. Seluruh data, baik variabel nominal maupun numerik langsung diproses oleh WEKA tanpa diberikan pembedaan kekuatan keanggotaan sebuah data dalam atribut yang memilikinya. Dengan demikian kemungkinan data tidak terklasifikasi dengan baik, menjadi lebih besar.

#### 3.3.2 Perhitungan Entropy serta Pembentukan Pohon Keputusan

Proses perhitungan *Gain* dan *Entropy* akan diulangi beberapa tahap hingga *node* terminasi secara keseluruhan.

##### a. Perhitungan Tahap Pertama

Tabel 4 memperlihatkan hubungan atribut dengan keputusan dari data awal.

Diketahui: jumlah data = 70 data

**Tabel 4. Hubungan atribut dan keputusan pertama**

<i>Variabel Atribut</i>	<i>Himpunan Fuzzy</i>	<i>Total Data</i>	<i>Keputusan Beasiswa</i>	
			<i>Ya</i>	<i>Tidak</i>
Nilai	Sedang	30	26	4
	Tinggi	40	25	15
Biaya Sekolah	Ibu	2	2	0
	Wali	2	2	0
	Orang Tua	66	47	19
Status Rumah	Pinjam	7	7	0
	Kontrak	7	7	0
	Sendiri	56	37	19
Bangunan Rumah	Non Permanen	14	14	0
	Semi Permanen	12	12	0
	Permanen	44	25	19
Kendaraan	Sepeda	26	24	2
	SepedaMotor	37	20	17
	Tidak Ada	7	7	0
JumlahTanggungan	Sedikit ( $\leq 2$ )	69	50	19
	Sedang (2 - 4)	1	1	0
	Banyak ( $\geq 4$ )	0	0	0
<b><i>Jumlah Kelas Pembeda (<math>d_1</math> dan <math>d_2</math>)</i></b>			<b><i>51</i></b>	<b><i>19</i></b>

Dengan mempergunakan persamaan (1), (2) dan (3), diperoleh perhitungan sebagai berikut:

(i) Hitung *Information Gain* dari atribut beasiswa (*decision*):

$$I(S1,S2) = - 51/70 \log_2 51/70 - 19/70 \log_2 19/70 = 0.863$$

(ii) Hitung *Information Gain* (I) dan *Entropy* (E) dari setiap atribut.

*Atribut Nilai*

*Fuzzy Set* = Sedang

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 26/30 \log_2 26/30 - 4/30 \log_2 4/30 = 0.567$$

*Fuzzy Set* = Tinggi

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12,S22) = - 25/40 \log_2 25/40 - 15/40 \log_2 15/40 = 0.954$$

$$E(\text{Nilai}) = 30/70 * I(S11,S21) + 40/70 * I(S12,S22) = 0.788$$

$$\text{Gain}(\text{Nilai}) = I(S1,S2) - E(\text{Nilai}) = 0.055$$

*Atribut Biaya Sekolah*

*Fuzzy Set* = Ibu

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 2/2 \log_2 2/2 - 0/2 \log_2 0/2 = 0$$

*Fuzzy Set* = Wali

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12,S22) = - 2/2 \log_2 2/2 - 0/2 \log_2 0/2 = 0$$

*Fuzzy Set* = Orang Tua

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S13,S23) = - 47/66 \log_2 47/66 - 19/66 \log_2 19/66 = 0.867$$

$$E(\text{Biaya}) = 2/70 * I(S11,S21) + 2/70 * I(S12,S22) + 66/70 * I(S13,S23) = 0.817$$

$$\text{Gain}(\text{Biaya}) = I(S1,S2) - E(\text{Biaya}) = 0.027$$

*Atribut Status Rumah*

*Fuzzy Set* = Pinjam

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 7/7 \log_2 7/7 - 0/7 \log_2 0/7 = 0$$

*Fuzzy Set* = Kontrak

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12,S22) = - 7/7 \log_2 7/7 - 0/7 \log_2 0/7 = 0$$

*Fuzzy Set* = Sendiri

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S13,S23) = - 37/56 \log_2 37/56 - 19/56 \log_2 19/56 = 0.924$$

$$E(\text{Rumah}) = 7/70 * I(S11,S21) + 7/70 * I(S12,S22) + 56/70 * I(S13,S23) = 0.739$$

$$\text{Gain}(\text{Rumah}) = I(S1,S2) - E(\text{Rumah}) = 0.104$$

*Atribut Bangunan Rumah*

*Fuzzy Set* = Non Permanen

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 14/14 \log_2 14/14 - 0/14 \log_2 0/14 = 0$$

*Fuzzy Set* = Semi Permanen

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S12,S22) = - 12/12 \log_2 12/12 - 0/12 \log_2 0/12 = 0$$

*Fuzzy Set* = Permanen

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S13,S23) = - 25/44 \log_2 25/44 - 19/44 \log_2 19/44 = 0.783$$

$$E(\text{Bangunan}) = 14/70 * I(S11,S21) + 12/70 * I(S12,S22) + 44/70 * I(S13,S23) = 0.492$$

$$\text{Gain}(\text{Bangunan}) = I(S1,S2) - E(\text{Bangunan}) = 0.351$$

*Atribut Kendaraan*

*Fuzzy Set* = Tidak Ada

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 7/7 \log_2 7/7 - 0/7 \log_2 0/7 = 0$$

*Fuzzy Set* = Sepeda

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S12,S22) = - 24/26 \log_2 24/26 - 2/26 \log_2 2/26 = 0.391$$

*Fuzzy Set* = Sepeda Motor

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S13,S23) = - 20/37 \log_2 20/37 - 17/37 \log_2 17/37 = 0.995$$

$$E(\text{Kendaraan}) = 7/70 * I(S11,S21) + 26/70 * I(S12,S22) + 37/70 * I(S13,S23) = 0.671$$

$$\text{Gain}(\text{Kendaraan}) = I(S1,S2) - E(\text{Kendaraan}) = 0.172$$

*Atribut Jumlah Tanggungan*

*Fuzzy Set* = Sedang

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11, S21) = - 1/1 \log_2 1/1 - 0/1 \log_2 0/1 = 0$$

*Fuzzy Set* = Sedikit

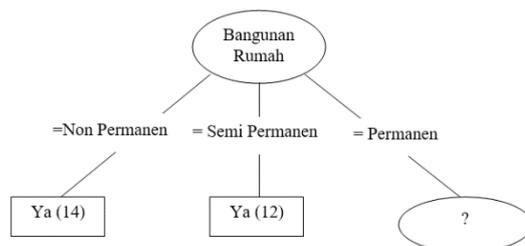
S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12, S22) = - 50/69 \log_2 50/69 - 19/69 \log_2 19/69 = 0.936$$

$$E(\text{Tanggungan}) = 1/70 * I(S11, S21) + 69/70 * I(S12, S22) = 0.923$$

$$\text{Gain}(\text{Tanggungan}) = I(S1, S2) - E(\text{Tanggungan}) = -0.079$$

Dari hasil perhitungan diatas, bangunan rumah digunakan sebagai *node* awal karena memiliki *Entropy* terkecil atau *Gain* yang paling besar, dengan *leafnode* adalah non permanen, semi permanen dan permanen. Gambar 5 memperlihatkan pembentukan tersebut.



**Gambar 5. Pembentukan Node Awal**

b. Perhitungan Tahap Kedua

*Node* berikutnya dihitung dengan menggunakan rumus yang sama dengan rumus yang digunakan untuk menghitung *node* pertama. Tabel 5 memperlihatkan hubungan atribut dengan keputusan dari data sisa proses perhitungan pertama, yaitu data awal dikurangi dengan data dari variabel atribut bangunan rumah. Diketahui: jumlah data = 44 data

**Tabel 5. Hubungan atribut dan keputusan kedua**

Variabel Atribut	Himpunan Fuzzy	Total Data	Keputusan Beasiswa	
			Ya	Tidak
Nilai	Sedang	15	11	4
	Tinggi	29	14	15
Biaya Sekolah	Ibu	1	1	0
	Wali	2	2	0
	Orang Tua	41	22	19
Status Rumah	Pinjam	3	3	0
	Kontrak	6	6	0
	Sendiri	35	16	19
Kendaraan	Sepeda	15	13	2
	SepedaMotor	26	9	17
	Tidak Ada	3	3	0
JumlahTanggungan	Sedikit ( $\leq 2$ )	44	25	19
	Sedang (2 - 4)	0	0	0
	Banyak ( $\geq 4$ )	0	0	0
<b>Jumlah Kelas Pembeda (<math>d_1</math> dan <math>d_2</math>)</b>			<b>25</b>	<b>19</b>

Dengan mempergunakan persamaan (1), (2) dan (3), diperoleh perhitungan sebagai berikut:

(i) Hitung *Information Gain* dari atribut beasiswa (decision):

$$I(S1, S2) = - 25/44 \log_2 25/44 - 19/44 \log_2 19/44 = 0.987$$

(ii) Hitung *Information Gain* (I) dan *Entropy* (E) dari setiap atribut.

*Atribut Nilai*

*Fuzzy Set* = Sedang

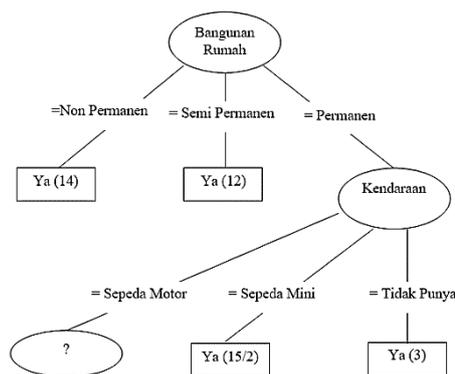
S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11, S21) = - 11/15 \log_2 11/15 - 4/15 \log_2 5/(15) = 0.837$$

*Fuzzy Set* = Tinggi

$S_{12} = \text{Beasiswa Ya}, S_{22} = \text{Beasiswa Tidak}$   
 $I(S_{12}, S_{22}) = -14/29 \log_2 14/29 - 15/29 \log_2 15/29 = 0.999$   
 $E(\text{Nilai}) = 15/44 * I(S_{11}, S_{21}) + 29/44 * I(S_{12}, S_{22}) = 0.944$   
 $\text{Gain}(\text{Nilai}) = I(S_1, S_2) - E(\text{Nilai}) = 0.043$   
*Atribut Biaya Sekolah*  
*Fuzzy Set = Ibu*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{11}, S_{21}) = -1/1 \log_2 1/1 - 0/1 \log_2 0/1 = 0$   
*Fuzzy Set = Wali*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{12}, S_{22}) = -2/2 \log_2 2/2 - 0/2 \log_2 0/2 = 0$   
*Fuzzy Set = Orang Tua*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{13}, S_{23}) = -22/41 \log_2 22/41 - 19/41 \log_2 19/41 = 0.996$   
 $E(\text{Biaya}) = 1/44 * I(S_{11}, S_{21}) + 2/44 * I(S_{12}, S_{22}) + 41/44 * I(S_{13}, S_{23}) = 0.928$   
 $\text{Gain}(\text{Biaya}) = I(S_1, S_2) - E(\text{Biaya}) = 0.058$   
*Atribut Status Rumah*  
*Fuzzy Set = Pinjam*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{11}, S_{21}) = -4/4 \log_2 4/4 - 0/4 \log_2 0/4 = 0$   
*Fuzzy Set = Kontrak*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{12}, S_{22}) = -6/6 \log_2 6/6 - 0/6 \log_2 0/6 = 0$   
*Fuzzy Set = Sendiri*  
 $S_{13} = \text{Beasiswa Ya}, S_{23} = \text{Beasiswa Tidak}$   
 $I(S_{13}, S_{23}) = -16/34 \log_2 16/34 - 19/34 \log_2 19/34 = 0.970$   
 $E(\text{Rumah}) = 4/44 * I(S_{11}, S_{21}) + 6/44 * I(S_{12}, S_{22}) + 34/44 * I(S_{13}, S_{23}) = 0.765$   
 $\text{Gain}(\text{Rumah}) = I(S_1, S_2) - E(\text{Rumah}) = 0.222$   
*Atribut Kendaraan*  
*Fuzzy Set = Tidak Ada*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{11}, S_{21}) = -3/3 \log_2 3/3 - 0/3 \log_2 0/3 = 0$   
*Fuzzy Set = Sepeda*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{12}, S_{22}) = -13/15 \log_2 13/15 - 2/15 \log_2 2/15 = 0.567$   
*Fuzzy Set = Sepeda Motor*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{13}, S_{23}) = -9/26 \log_2 9/26 - 17/26 \log_2 17/26 = 0.931$   
 $E(\text{Kendaraan}) = 3/44 * I(S_{11}, S_{21}) + 15/44 * I(S_{12}, S_{22}) + 26/44 * I(S_{13}, S_{23}) = 0.743$   
 $\text{Gain}(\text{Kendaraan}) = I(S_1, S_2) - E(\text{Kendaraan}) = 0.244$   
*Atribut Jumlah Tanggungan*  
*Fuzzy Set = Sedikit*  
 $S_{11} = \text{Beasiswa Ya}, S_{21} = \text{Beasiswa Tidak}$   
 $I(S_{12}, S_{22}) = -25/44 \log_2 25/44 - 19/44 \log_2 19/44 = 0.987$   
 $E(\text{Tanggungan}) = 44/44 * I(S_{21}, S_{22}) = 0.987$   
 $\text{Gain}(\text{Tanggungan}) = I(S_1, S_2) - E(\text{Tanggungan}) = 0$

Dari hasil perhitungan diatas, kendaraan digunakan sebagai *node* kedua karena memiliki *Entropy* terkecil atau *Gain* yang paling besar, dengan *leaf node* adalah tidak punya, sepeda mini dan sepeda motor. Dimana pada *leaf* sepeda mini (15/2) terdapat 15 data 2 diantaranya merupakan data bukan penerima beasiswa. Gambar 6 memperlihatkan pembentukan tersebut.



**Gambar 6. Pembentukan Node Kedua**

c. Perhitungan Tahap Ketiga

Node berikutnya dihitung dengan menggunakan rumus yang sama dengan rumus yang digunakan untuk menghitung node pertama dan kedua. Tabel 6 memperlihatkan hubungan atribut dengan keputusan dari data sisa proses perhitungan kedua, yaitu data awal dikurangi dengan data pada variabel atribut bangunan rumah dan kendaraan. Diketahui: jumlah data = 26 data.

**Tabel 6. Hubungan atribut dan keputusan ketiga**

Variabel Atribut	Himpunan Fuzzy	Total Data	Keputusan Beasiswa	
			Ya	Tidak
Nilai	Sedang	9	4	5
	Tinggi	17	5	12
Biaya Sekolah	Ibu	0	0	0
	Wali	1	1	0
	Orang Tua	25	8	17
Status Rumah	Pinjam	1	0	1
	Kontrak	5	5	0
	Sendiri	20	4	16
<b>Jumlah Kelas Pembeda (<math>d_1</math> dan <math>d_2</math>)</b>			<b>9</b>	<b>17</b>

Dengan mempergunakan persamaan (1), (2) dan (3), diperoleh perhitungan sebagai berikut:

(i) Hitung *Information Gain* dari atribut beasiswa (decision):

$$I(S1,S2) = - 9/26 \log_2 9/26 - 17/26 \log_2 17/26 = 0.931$$

(ii) Hitung *Information Gain* (I) dan *Entropy* (E) dari setiap atribut;

*Atribut Nilai*

*Fuzzy Set* = Sedang

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 5/9 \log_2 5/9 - 4/9 \log_2 4/9 = 0.991$$

*Fuzzy Set* = Tinggi

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12,S22) = - 4/17 \log_2 4/17 - 13/17 \log_2 13/17 = 0.787$$

$$E(\text{Nilai}) = 9/26 * I(S11,S21) + 17/26 * I(S21,S22) = 0.858$$

$$\text{Gain}(\text{Nilai}) = I(S1,S2) - E(\text{Nilai}) = 0.073$$

*Atribut Biaya Sekolah*

*Fuzzy Set* = Wali

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 2/2 \log_2 2/2 - 0/2 \log_2 0/2 = 0$$

*Fuzzy Set* = Orang Tua

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S12,S22) = - 7/24 \log_2 7/24 - 17/24 \log_2 17/24 = 0.871$$

$$E(\text{Biaya}) = 2/26 * I(S11,S21) + 24/26 * I(S21,S22) = 0.804$$

$$\text{Gain}(\text{Biaya}) = I(S1,S2) - E(\text{Biaya}) = 0.127$$

*Atribut Status Rumah*

*Fuzzy Set* = Pinjam

S11 = Beasiswa Ya, S21 = Beasiswa Tidak

$$I(S11,S21) = - 1/1 \log_2 1/1 - 0/1 \log_2 0/1 = 0$$

*Fuzzy Set* = Kontrak

S12 = Beasiswa Ya, S22 = Beasiswa Tidak

$$I(S_{12}, S_{22}) = -5/5 \log_2 5/5 - 0/5 \log_2 0/5 = 0$$

Fuzzy Set = Sendiri

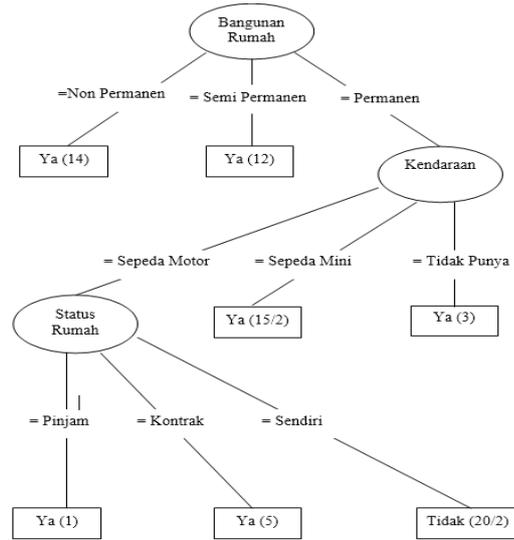
$$S_{13} = \text{Beasiswa Ya}, S_{23} = \text{Beasiswa Tidak}$$

$$I(S_{13}, S_{23}) = -2/20 \log_2 2/20 - 18/20 \log_2 18/20 = 0.469$$

$$E(\text{Rumah}) = 1/26 * I(S_{11}, S_{22}) + 5/26 * I(S_{12}, S_{22}) + 20/26 * I(S_{13}, S_{23}) = 0.361$$

$$\text{Gain}(\text{Rumah}) = I(S_1, S_2) - E(\text{Rumah}) = 0.57$$

Dari hasil perhitungan diatas, status rumah digunakan sebagai *node* ketiga karena memiliki *Entropy* terkecil atau *Gain* yang paling besar, dengan *leaf node* pinjam, kontrak dan sendiri. Perhitungan pembentukan *node* berakhir pada perhitungan pembentukan *node* ketiga karena sisa data sudah tidak bisa dihitung untuk pembentukan *node* keempat. Dimana pada *leaf* sendiri (20/2) terdapat 20 data 2 diantaranya merupakan data penerima beasiswa. Sehingga pohon keputusan (*decision Tree*) terbentuk dengan jumlah 3 *node* dan 7 *leaf*. Gambar 7 memperlihatkan pembentukan tersebut.



Gambar 7. Pembentukan *Node* Ketiga

### 3.3.3 Hasil Uji *Correctly/incorrectly Classified Instances*

Untuk tingkat kesalahan validasi data pada kedua metode, pengujian dilakukan secara otomatis oleh WEKA. Hasil uji *Correctly/incorrectly Classified Instances* ini sangat bergantung kepada kepada konsistensi hubungan sebab akibat antara data masukan dan data keluaran. Sehingga, semakin banyak data yang tidak konsisten, maka semakin rendah data yang terklasifikasi dengan benar. Nilai *correctly Classified Instances* dan *incorrectly Classified Instances* tersebut diperoleh dari hasil pembacaan confusion matrix dalam WEKA sebagai berikut:

- Metode 1
  - Pembacaan Benar = 65
  - Pembacaan Salah = 5
  - Bernilai Ya = 51
  - Terbaca Ya = 48
  - Terbaca Tidak = 3
  - Bernilai Tidak = 19
  - Terbaca Ya = 2
  - Terbaca Tidak = 17
- Metode 2
  - Pembacaan Benar = 63
  - Pembacaan Salah = 7
  - Bernilai Ya = 51
  - Terbaca Ya = 48
  - Terbaca Tidak = 3
  - Bernilai Tidak = 19
  - Terbaca Ya = 4
  - Terbaca Tidak = 15

Dengan menggunakan data yang sama, dari hasil pengujian terlihat metode 1 memiliki tingkat kebenaran klasifikasi lebih tinggi daripada metode 2. Hal ini dikarenakan metode 1 menggunakan data yang telah dibentuk

dalam *Membership Function*, yang menyebabkan cakupan derajat suatu variabel lebih terfokus dan bernilai pasti, sedangkan pada metode 2, variabel data memiliki cakupan yang tidak terfokus dan nilainya tidak bisa dipastikan. Hasil perbandingan dari kedua metode diatas diperlihatkan dalam Tabel 7.

**Tabel 7. Perbandingan uji *correctly/incorrectly classified instances***

<i>Jenis Instances</i>	<i>Metode 1</i>	<i>Metode 2</i>
<i>Correctly</i>	92.8571%	90%
<i>Incorrectly</i>	7.1429%	10%

#### 4. KESIMPULAN

Proses seleksi penerima beasiswa dipengaruhi oleh beberapa variabel, yang secara garis besar terbagi atas 2 jenis, yaitu variabel kualitatif dan variabel kuantitatif. Proses komputasi terhadap kedua jenis variabel tersebut dapat dilakukan secara bersamaan dengan baik dan proporsional. Dari hasil uji validasi silang (cross validation) diperoleh nilai *error* terendah adalah pada metode Integrasi *Membership Function* dengan Algoritma C4.5, yaitu sebesar 0.2714 untuk *Root Mean Square Error* (RMSE) dan 0.132 untuk *Mean Square Error* (MAE),

Sementara dari hasil Perbandingan Uji *Correctly/incorrectly Classified Instances*, diperoleh persentase 92.8571% untuk *instances* menggunakan metode Integrasi *Membership Function* dengan Algoritma C4.5, dimana persentase tersebut lebih baik dari metode kedua. Metode integrasi *Membership Function* dan Metode Penerapan Algoritma C4.5, kurang sesuai untuk kasus dimana variabel yang diperhitungkan melibatkan variabel kualitatif dan kuantitatif seperti kasus seleksi penerimaan beasiswa yang memiliki variabel kualitatif dan kuantitatif.

#### UCAPAN TERIMA KASIH

Tim penulis mengucapkan Terima Kasih yang sebesar-besarnya kepada Dr. Said Mirza Pahlevi, M.Eng., yang telah memberikan sangat banyak masukan dan kritik terkait teknik pengujian, metodologi penelitian dan teori klasifikasi dan pengujian data, yang menjadi dasar dan pondasi bagi penelitian ini. Terima kasih juga disampaikan kepada Kepala Sekolah dan Wakil Kepala Sekolah serta seluruh pihak terkait di SMK PGRI Ploso yang telah bersedia memberikan data dan informasi yang dibutuhkan bagi penelitian ini.

#### DAFTAR PUSTAKA

- [1] Priyana J. Abadi AM. Peramalan Suhu Udara di Yogyakarta dengan Menggunakan Model *Fuzzy*. Yogyakarta: Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, Universitas Negeri Yogyakarta; 2011.
- [2] Saikhu, A., Lianto, J., & Hanik, U. *Fuzzy Decision Tree Dengan Algoritma C4.5 Pada Data Diabetes Indian Pima*. Bali: Konferensi Nasional Sistem dan Informatika; 2011.
- [3] Subakti I. *Buku Panduan IF1524: Sistem Pendukung Keputusan*. Surabaya: FTI - Institut Teknologi Sepuluh Nopember; 2002.
- [4] Hermawan AT, Gunawan. Mahono YC. *Decision Support System Tool untuk Penyelesaian Permasalahan Linear Berbasis Simplex dan Revised Simplex*. Yogyakarta: Seminar Nasional Aplikasi Teknologi Informasi (SNATI); 2009.
- [5] Sunjana. *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree*. Yogyakarta: SNATI 2010; 2010.
- [6] Moertini VS. *Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi (Studi Kasus: Pra-Pengolahan Dan Klasifikasi Citra Batik)*. Bandung: (Disertasi) Program Studi Teknik Informatika –ITB; 2007.
- [7] Susanto S. Suryadi D. *Pengantar Data Mining; Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi Offset; 2010.
- [8] Han J, Kamber M. *Data Mining: Concepts and Techniques*, 2nd Edition. San Francisco: Morgan Kaufmann Publisher; 2006.
- [9] Negnevitsky M. *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd Edition. Harlow: Pearson Education Limited; 2005.
- [10] Wang TC. Lee HD. *Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy*. Taiwan: I-Shou University & Fortune Institute of Technology; 2006.
- [11] Maulana, M. R., & Al Karomi, M. A. *Information Gain Untuk Mengetahui Pengaruh Atribut Terhadap Klasifikasi Persetujuan Kredit*. Pekalongan: Jurnal Litbang Kota Pekalongan Vol. 9; 2015.
- [12] Sivanandam SN. Sumathi S. Deepa SN. *Introduction to Fuzzy Logic using MATLAB*. Berlin: Springer; 2007.
- [13] Kusumadewi S. Purnomo H. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*, Edisi 2. Yogyakarta: Graha Ilmu; 2010.