

Uji Metode Klasifikasi Data Dalam Proses Seleksi Penerima Beasiswa SMK PGRI Ploso

Ade Surya Budiman¹, Xanty Adhi Parandani²

¹ Program Studi Teknik Komputer, AMIK BSI Jakarta
ade.aum@bsi.ac.id

² Program Studi Manajemen Informatika, AMIK BSI Bogor
xanty.xip@bsi.ac.id

Abstrak

Proses seleksi beasiswa melibatkan banyak variabel untuk indikator penilaian kelayakan penerima beasiswa. Terdapat dua jenis variabel penilaian, yaitu variabel Nominal dan variabel Numerik. Kedua variabel ini memiliki kesetaraan perhitungan dan pertimbangan dalam proses seleksi. Terdapat berbagai metode yang bisa dipergunakan untuk melakukan proses klasifikasi dan pengenalan pola data yang dipakai dalam seleksi beasiswa. Didalam penelitian ini dilakukan pengujian terhadap dua metode klasifikasi dan pengenalan pola data, yaitu menggunakan implementasi logika fuzzy dan integrasi logika fuzzy dengan algoritma C4.5. Pengujian dilakukan untuk menguji keakuratan dari metode yang diperhitungkan dengan menggunakan metode pengujian 10-folds cross validation. Dari hasil penelitian diperoleh Root Mean Square Error (RMSE) sebesar 0,216 dan Mean Absolut Error (MAE) sebesar 0,0941 untuk implementasi logika fuzzy. Sedangkan untuk integrasi logika fuzzy dengan algoritma C4.5 diperoleh hasil RMSE sebesar 0,2714 dan MAE sebesar 0,132. Dari penelitian ini disimpulkan bahwa implementasi logika fuzzy memberikan hasil pengujian yang lebih baik.

Kata Kunci : Seleksi Beasiswa, Variabel Nominal, Variabel Numerik, Implementasi Logika Fuzzy, Algoritma C4.5.

Abstract

The scholarship selection process involves many variables for feasibility assessment indicators recipients. There are two types of assessment variables, namely Nominal variables and Numeric variables. Both of these variables have equal calculation and consideration in the selection process. There are various methods that can be used to make the process of classification and pattern recognition data used in the selection of scholarship. In this study tested the two methods of data classification and pattern recognition, which uses fuzzy logic implementation and integration of fuzzy logic with C4.5 algorithm. Tests conducted to test the accuracy of the method were calculated using the method of testing 10-folds cross validation. The results were obtained Root Mean Square Error (RMSE) of 0.216 and Mean Absolute Error (MAE) of 0.0941 to the fuzzy logic implementation. As for the integration of fuzzy logic and C4.5 algorithm's resulting RMSE value 0.2714 and MAE value 0.132. This study suggests that the fuzzy logic implementation provides better test results.

Keywords: Scholarship Selection, Nominal Variables, Numeric Variables, Fuzzy Logic Implementation, C4.5 Algorithm.

1. Pendahuluan

Seleksi penerima beasiswa merupakan salah satu kegiatan pengambilan keputusan yang cukup rumit karena ada beragam variabel yang digunakan sebagai bahan pertimbangan dalam proses seleksi. Variabel yang digunakan juga akan bergantung kepada jenis beasiswa yang diberikan. Beasiswa

Bantuan Khusus Siswa Miskin/Bantuan Khusus Murid Miskin (BKSM/BKMM) lebih melihat kepada aspek kondisi sosial ekonomi calon penerima beasiswa.

SMK PGRI Ploso sebagai salah satu institusi bertugas menyalurkan beasiswa BKSM kepada siswa/i yang berhak menerimanya, memiliki tanggung jawab untuk melakukan seleksi yang

cukup ketat bagi calon penerima beasiswa tersebut.

Pihak penyeleksi dituntut untuk bisa mencari indikator mana yang paling berperan dan berpengaruh bagi pantas atau tidaknya seorang siswa menerima beasiswa BKSM sekaligus memperlihatkan kondisi ekonomi sebenarnya dari seorang siswa, dari beberapa indikator yang terdapat didalam pendataan kondisi sosial ekonomi siswa.

Dalam indikator penilaian tersebut terdapat variabel angka (kuantitatif/numerik) dan variabel linguistik (kualitatif/nominal). Variabel kuantitatif diantaranya adalah nilai akademik siswa dan jumlah tanggungan sekolah orang tua/wali. Variabel kualitatif diantaranya adalah sumber biaya sekolah dan status kepemilikan rumah/tempat tinggal.

Pertanyaan yang kemudian akan muncul adalah, bagaimana cara menggunakan variabel yang berbeda jenis tersebut dengan tetap memikirkan proporsi yang seimbang untuk pengambilan keputusan pemberian beasiswa?

Ali Adeli dan Mehdi Neshat (Adeli & Neshat, 2010), melakukan penelitian berjudul A Fuzzy Expert System for Heart Disease Diagnosis. Dalam penelitian tersebut digunakan sistem pakar berbasis aturan logika fuzzy dalam mengklasifikasi dan mengeluarkan kesimpulan pendukung keputusan untuk mendiagnosa penyakit tertentu tanpa bantuan dokter atau pemeriksaan laboratorium.

Tien-Chin Wang dan Hsien-da Lee (Wang & Lee, 2006) dalam penelitian berjudul "Constructing Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy, mengembangkan pohon keputusan (decision tree) berbasis Himpunan Fuzzy (Fuzzy Sets) yang diintegrasikan dengan algoritma ID3/C4.5 dalam kasus penerimaan siswa.

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 (Subakti, 2002).

Decision Support System juga dapat didefinisikan sebagai kumpulan dari aplikasi perangkat lunak yang terintegrasi beserta dengan perangkat keras, yang

kemudian membentuk tulang punggung dalam proses pembuatan keputusan suatu organisasi (www.microstrategy.com).

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 (Hermawan, Gunawan, & Mahono, 2009).

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.

Algoritma C4.5

Algoritma C4.5 merupakan pengembangan (*extention*) dari algoritma ID3. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah *decision tree* berdasarkan training data yang telah disiapkan (Sunjana, 2010).

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 data pelatihan 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 (Moertini, 2007).

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 (Susanto & Suryadi, 2010)

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 (Han & Kamber, 2006).

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 (Negnevitsky, 2005).

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* dan *entropy*. Berikut adalah langkah-langkah perhitungan *entropy* dan *gain* yang merupakan representasi algoritma pembentukan pohon keputusan *fuzzy* dengan algoritma ID3/C4.5 (telah diolah kembali dari (Wang & Lee, 2006)): Tentukan komponen himpunan dari data,

$$\begin{aligned} U &= \{1, \dots, 70\} \\ C &= \{\text{Nilai, Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan, Jumlah Tanggungan}\} \\ D &= \{\text{Beasiswa}\}; d_1 = \text{Ya} = 50; d_2 = \text{Tidak} = 20; m=2 \\ J &= \{U, C \cup D\} \end{aligned}$$

Information Gain (I) dari atribut keputusan (D) dirumuskan:

$$I(S_1, \dots, S_m) = - \sum_{i=1}^m p_i \log_2 p_i \quad (1)$$

Entropy (E) dari atribut (c) dirumuskan:

$$E(C_i) = \sum_{j=i}^v \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \quad (2)$$

Information Gain (I) untuk setiap *Fuzzy Sets* (S) dirumuskan:

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

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*)
- S_j = Sub Himpunan (subset)
- S_{ij} = Jumlah Sampel Yang Masuk Kedalam Kelas Di Dalam Subset j
- c = Atribut Kondisi (*Fuzzy Set*)

Proses perhitungan *gain* dan *entropy* akan diulangi beberapa tahap hingga mencapai node terminasi.

Logika Fuzzy dan Membership Function

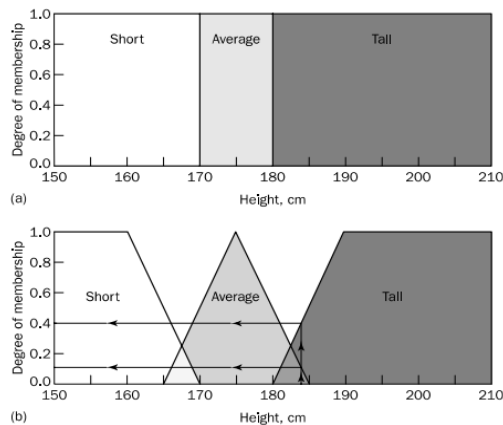
Logika *fuzzy* pertama kali dicetuskan oleh Dr. Lotfi Zadeh, Professor pada *Electrical Engineering Department, University of California*, pada tahun 1965, melalui sebuah paper berjudul "*Fuzzy Sets*" atau "*Himpunan Fuzzy*".

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*) (Sivanandam, Sumathi, & Deepa, 2007).

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 (Kusumadewi & Purnomo,

2010). 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. Fungsi Keanggotaan: (a) *Crisp*; (b) *Fuzzy*

Sumber: (Negnevitsky, 2005)

Rule Base System

Rule based dibagi menjadi dua bagian yaitu pengetahuan faktual yang relevan dengan masalah yang dihadapi, direalisasikan sebagai data atau database atau basis dan keterampilan yang dibutuhkan dengan menggunakan pengetahuan yang tersedia untuk memecahkan masalah, dan membimbing perolehan fakta-fakta baru dan keterampilan (*skill*) baru, yang diwujudkan sebagai aturan (Siler & Buckley, 2005).

Rule based system mengharuskan *expert's knowledge* dan pola berpikir ditentukan secara eksplisit. Dalam hal ini (biasanya), dua orang (atau kelompok) mengembangkan sistem bersama. *Domain expert*, yaitu orang yang tahu bagaimana memecahkan masalah tapi tidak memahami pemrograman komputer, dan *knowledge engineer*, yaitu orang yang mengetahui dengan baik teknologi komputer dan ahli sistem, tetapi hanya memiliki sedikit atau tidak mengetahui masalah yang sedang dihadapi. Memproleh pengetahuan dan peraturan

dengan tepat berdasarkan cara diatas disebut fase akuisisi pengetahuan.

Fuzzy rule based system memproses *capability sequential* dan *parallel rule* dengan menambahkan fitur tambahan untuk bahasa *conventional expert system*, yaitu penanganan terhadap ketidakpastian, ambiguitas dan kontradiksi, dan memodifikasi kata-kata seperti "sangat" dan "agak".

Jadi, *fuzzy system* meningkatkan kemampuan dalam meniru pola berpikir yang tidak kaku seperti dokter dan ahli biologi serta pola yang relatif kaku seperti ilmuwan komputer (Siler & Buckley, 2005).

Fuzzy Expert System

Fuzzy expert system merupakan penggabungan sistem pakar dan sistem *fuzzy*. Penerapan sistem *Fuzzy* dalam sistem pakar bertujuan untuk merepresentasikan pengetahuan pakar pada lingkungan yang tidak pasti, tidak lengkap, dan sangat kompleks ((Kandel, 2001); (Marimin, 2005)).

Fuzzy expert system mengembangkan sistem pakar yang menggunakan logika *fuzzy* secara keseluruhan ((Negnevitsky, 2005); (Siler & Buckley, 2005)), yang meliputi gugus *fuzzy*, aturan *fuzzy if-then*, serta proses inferensi. Gugus *fuzzy* merupakan perangkat yang tepat untuk mengekspresikan hal yang ambigu, yang diperlukan oleh komputer untuk mengerti bahasa manusia yang tidak dapat diselesaikan dengan logika biasa.

2. Metode Penelitian

Kerangka Penelitian

Secara garis besar, kerangka penelitian dirincikan sebagai berikut:

- Data yang terkumpul dari objek penelitian dilakukan pra proses terlebih dahulu, dimana data mentah akan diolah dan diperbaiki untuk menghilangkan variabel yang tidak relevan dan tidak diperlukan dalam penelitian. Proses ini juga melibatkan pakar atau guru yang berhak menangani masalah pendataan dan penentuan siswa calon penerima beasiswa.
- Data hasil pra proses selanjutnya dipilah berdasarkan jenis variabel yaitu variabel kualitatif (nominal/linguistik) atau variabel kuantitatif (numerik).
- Perlakuan data selanjutnya dipisah kedalam 2 metode berbeda, yaitu

- penerapan logika *fuzzy* dan integrasi *fuzzy* dengan algoritma klasifikasi C4.5 pada WEKA.
- d. Variabel kuantitatif (numerik) pada Metode 1 dan Metode 2 akan melalui proses pembentukan fungsi keanggotaan (membership function) terlebih dahulu.
 - e. Variabel kualitatif (nominal/linguistik) pada Metode 1 dan 2 juga akan melalui proses pembentukan fungsi keanggotaan (membership function) terlebih dahulu.
 - f. Hasil membership function untuk Metode 1, kemudian diolah menggunakan *Fuzzy Inference System* (FIS) Model Sugeno didalam *Fuzzy Toolbox* MATLAB.
 - g. Hasil membership function untuk Metode 2, kemudian langsung diolah menggunakan algoritma klasifikasi C4.5 pada *machine learning system* WEKA.
 - h. Data pada Metode 1 selanjutnya diproses kembali dengan menggunakan *Adaptive-Neuro* FIS (ANFIS). Hal ini bertujuan untuk memperbaiki keluaran FIS ketika dilakukan pengujian validasi silang (*cross validation*), dimana untuk melakukan uji validasi silang, data yang dimasukkan secara manual pada FIS, harus melalui penyesuaian/pengaturan (*adjustment/tuning*) secara otomatis oleh ANFIS.
 - i. Kedua Metode kemudian diberlakukan uji validasi silang dengan iterasi sebanyak 10 kali (*10-folds cross validation*).
 - j. Dari hasil pengujian, dilakukan evaluasi.
 - k. Untuk metode 1, sebelum dimasukkan kedalam ANFIS, dilakukan evaluasi FIS *Rule Base*, untuk menguji keakuratan aturan yang dimasukkan.
 - l. Untuk Metode 2, dilakukan visualisasi pohon keputusan (*decision tree*) yang bisa dihasilkan (*generated*) secara otomatis oleh WEKA. Hal ini akan mempermudah pembacaan hasil klasifikasi data dan ekstraksi aturan dari data penelitian.

Populasi dan Sampel Penelitian

Populasi siswa SMK PGRI Ploso untuk setiap tahun ajaran mencapai 1200 siswa, yang dibagi ke dalam kelas pagi dan kelas siang. Perolehan jatah beasiswa

pertahun berada pada kisaran 200-300 siswa.

Jumlah sampel data adalah sebanyak 70 data siswa yang diambil dari 2 kelas berbeda. Penggunaan 70 data siswa ini bisa mewakili keseluruhan siswa calon penerima beasiswa (siswa yang diamati untuk pemberian beasiswa). Sesuai pendapat pakar/guru penyeleksi, 70 data siswa telah dapat mewakili calon penerima beasiswa secara keseluruhan, karena indikator penilaian yang sama untuk setiap proses seleksi.

Metode Pengumpulan Data

Data mentah diperoleh dari Daftar Nilai Siswa, dan Biodata Siswa Tahun Diklat 2010/2011, yaitu rekam kondisi pribadi siswa yang dilakukan oleh setiap wali kelas dalam satu tahun ajaran, serta data penerima beasiswa pada tahun ajaran/diklat 2010/2011.

3. Pembahasan Pra Proses Data

Sebelum memulai proses eksplorasi data, dilakukan proses pendahuluan (*preprocessing*) terhadap data, yaitu berupa pengabaian atribut pada data mentah yang dianggap tidak relevan dengan hasil pengujian.

Dari keseluruhan variabel yang terdapat dalam data yang diperoleh dari pihak sekolah, dipilih variabel yang paling mendekati kriteria/indikator syarat penerima beasiswa BKSM seperti yang dijelaskan dalam latar belakang penelitian.

Pemilihan variabel tersebut didasari oleh variabel standar pendataan siswa yang berkaitan dengan kondisi sosial ekonomi yang digunakan di objek penelitian (SMK PGRI Ploso). Variabel standar pendataan siswa tersebut tidak semuanya bisa digunakan sebagai variabel pembeda tingkatan sosial ekonomi antar siswa, sehingga dipilih variabel tertentu yang dianggap memiliki peran signifikan untuk menunjukkan kondisi ekonomi seorang siswa.

Atribut pada data asli yang akan digunakan dalam penelitian, terdiri atas 10 variabel, yaitu, (i) Input, terdiri atas: No., NIPD, Nama, Nilai, Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan, Jumlah Tanggungan dan (ii) Output, Beasiswa.

Selanjutnya data yang akan digunakan dalam ketiga metode diatas akan menggunakan data olahan yang

hanya akan terdiri atas Nama, Nilai, Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan, Jumlah Tanggungan dan Beasiswa.

Secara umum, variabel data dalam penelitian ini dibedakan atas 2 jenis variabel, yaitu (i) variabel kualitatif (nominal) yang terdiri atas Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan dan Beasiswa, serta (ii) variabel kuantitatif (numerik) yaitu Nilai dan Jumlah Tanggungan.

Metode yang Digunakan dan Dievaluasi

Terdapat 3 (tiga) metode yang akan digunakan dan dievaluasi dalam penelitian ini, yaitu:

1. Metode Penerapan Logika Fuzzy.

Dalam metode ini, data diproses dengan menerapkan logika *fuzzy* dengan menggunakan perangkat lunak MATLAB *Fuzzy Toolbox*. Hasil yang diperoleh merupakan intisari berupa informasi yang diperoleh dari ekstraksi data menggunakan aturan-aturan (*rules*) yang dibuat oleh pakar maupun pengumpulan *experience* data (data valid yang diperoleh).

2. Metode Integrasi Membership Function dengan Algoritma C4.5.

Data yang memiliki nilai kuantitatif (numerik) diberikan derajat dan fungsi keanggotaan, sehingga kesemuanya data akan berupa data kualitatif (nominal). Setelah keseluruhan berupa data nominal, maka data diolah dengan algoritma C4.5 menggunakan perangkat lunak WEKA.

Metode Evaluasi

Proses evaluasi yang digunakan dalam penelitian ini untuk menguji keakuratan proses klasifikasi data oleh ketiga metode yang digunakan, adalah metode evaluasi dan pengujian *10-Folds Cross Validation*. Metode pengujian *10-Folds Cross Validation* atau *k-Folds Cross Validation* adalah metode pengujian yang diterapkan kepada *training data* untuk menguji tingkat validasi (termasuk besaran *error*) dari data yang diuji.

Metode ini dapat dikatakan sebagai suatu metode pengujian sejauh mana data yang diperoleh *valid* dan benar untuk menunjukkan hubungan antara masukan, proses dan keluaran pada sistem yang berlaku saat ini.

Dengan metode ini, data dibagi menjadi beberapa bagian. Pembagian ini dinamakan sebagai *subset* (misalnya k-bagian). Dimana k merupakan nilai dari

fold. Sehingga jika k bernilai 10 maka akan dinamakan *10-folds*. Jika dilakukan *10-folds*, berarti diberlakukan 10 kali pengujian. Dimana *training data* sebanyak k-1 dan *testing data* adalah 1 bagian, selanjutnya setiap 1 bagian *training data* akan diuji oleh *testing data*. Nilai *error* akan diperoleh dari rata-rata setiap pengujian.

Pada WEKA, nilai *default* dari pengujian *cross validation* adalah 10. Sehingga *test options* pada WEKA memiliki nilai awal *10-folds cross validation*.

Pada *Fuzzy Toolbox* MATLAB pengujian *cross validation*, istilah *fold* sinonim dengan istilah *epochs*. Namun untuk pengujian ini diperlukan penerapan model *Fuzzy Sugeno* (Model TSK).

Evaluasi Hasil Pengujian yang dilakukan akan dilihat dari 3 indikator, yaitu: (i) Nilai kesalahan data yang dicerminkan oleh *Root Mean Square Error* (RMSE) setelah dilakukan iterasi terhadap data, (ii) Nilai kesalahan yang dicerminkan oleh *Mean Absolut Error* (MAE), (iii) Evaluasi kebenaran data berdasarkan aturan-aturan (*rules*) dengan menggunakan GUI *rule viewer* MATLAB.

Hasil Penelitian

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

Tabel 1. Distribusi Data Penelitian (Hasil Pra Proses)

Variabel	Anggota	
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
Biaya Sekolah	Pinjam	7
	Kontrak	6
	Sendiri	57
Bangunan Rumah	Non Permanen	14
	Semi Permanen	10
	Permanen	46
Kendaraan	Tidak Punya	6

	Sepeda Mini	26
	Sepeda Motor	38

		Banyak (≥ 4)		[6-10]
Output	Beasiswa	Ya	[0 dan 0,5]	[0]
		Tidak		[0,5]

Jumlah Tanggungan Orang Tua	1 Orang	38
	2 Orang	25
	3 Orang	6
	4 Orang	1
Keputusan Beasiswa	Tidak	20
	Ya	50

Pengolahan dan Pengujian Metode

1. Metode 1 (Penerapan Logika Fuzzy)

Data hasil pengolahan selanjutnya diproses untuk memiliki fungsi keanggotaan (membership function). Atribut data hasil pengolahan yang diberikan derajat keanggotaan adalah Nilai, Biaya Sekolah, Status Rumah, Bangunan Rumah, Kendaraan, Jumlah Tanggungan dan Beasiswa.

Pembentukan himpunan Fuzzy, semesta pembicaraan dan domain dari setiap atribut diatas diperlihatkan dalam Tabel 2.

Tabel 2. Pembentukan Fungsi Keanggotaan

Fungsi	Variabel	Himpunan Fuzzy	Semesta Pembicaraan	Domain	
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 (≤ 2)	[0-10]		[0-4]
		Sedang (2 - 4)			[3-7]

Setelah setiap atribut ditentukan semesta pembicaraannya, maka data dipindahkan kedalam *Fuzzy Logic Toolbox* MATLAB untuk memperoleh bentuk derajat keanggotaan sesuai dengan yang diinginkan.

Untuk menguji validasi data yang dimiliki dengan aturan-aturan yang telah dibuat, maka dilakukan uji validasi silang. Uji validasi silang ini akan membagi data hasil pengolahan menjadi beberapa *subset*. Uji validasi yang dilakukan dalam penelitian ini diatur pada posisi iterasi (pendekatan) sebanyak 10 kali (*10 folds*). Karena jumlah data total berjumlah 70 data, maka setiap *subset* akan dibagi menjadi 7 data untuk dilakukan pengujian validasi silang sebanyak 10 kali. Nilai Kesalahan dari uji validasi data ini adalah:

$$\begin{aligned} \text{Root Mean Square Error (RMSE)} &= 0.2616 \\ \text{Mean Absolut Error (MAE)} &= 0.0941 \end{aligned}$$

Nilai RMSE diperoleh dari *error* hasil pengujian data *testing* (*Checking Data* pada MATLAB) jika dibandingkan dengan keluaran FIS (*FIS Output*), sedangkan nilai MAE diperoleh dari *error* hasil pengujian data *training* jika dibandingkan dengan keluaran FIS (*FIS Output*).

2. Metode 2 (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). Dalam Metode kedua, variabel data kuantitatif (numerik) – berupa atribut nilai siswa dan jumlah tanggungan - diolah terlebih dahulu, yaitu dengan memberikan derajat keanggotaan kepada setiap nilai siswa dan jumlah tanggungan. Dengan demikian variabel nilai siswa dan jumlah tanggungan tidak lagi berupa numerik, namun berupa nominal.

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 dengan ekstensi 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 ekstensi file yang digunakan adalah *.csv.

Proses berikutnya adalah melakukan uji validasi dengan menggunakan *test option* berupa *10-folds cross validation*. Dari hasil pengujian diperoleh hasil sebagai berikut:
Root Mean Square Error (RMSE) = 0.2714
Mean Absolut Error (MAE) = 0.132
 Proses terakhir adalah pembentukan pohon keputusan (*decision tree*) yang dihasilkan (*generated*) secara otomatis oleh WEKA.

Evaluasi dan Analisa

1. Hasil Uji Cross Validation

Dalam Tabel 3 diperlihatkan perbandingan hasil uji validasi data yang telah dijabarkan didalam subbab sebelumnya.

Tabel 3.Perbandingan Hasil Uji Validasi Silang

Indikator	Metode-1 FUZZY	Metode-2 MF-WEKA
<i>Mean Absolut Error</i>	0.0941	0.132
<i>Root Mean Square Error</i>	0.2616	0.2714

Hasil yang diperoleh diatas, diperoleh dengan menggunakan 2 metode yang berbeda, yaitu (1) *Fuzzy logic* (MATLAB *Fuzzy Toolbox*), dan (2) *Membership function* dengan algoritma C4.5 (diimplementasi oleh algoritma J.48 pada WEKA).

2. Hasil Uji Rule Viewer

Dengan menggunakan fasilitas *rule viewer* yang merepresentasikan sistem pengujian data terhadap aturan yang telah dimasukkan pada MATLAB, dimana pengujian ini dikhususkan untuk melihat

validasi data jika menggunakan metode 1 (penerapan logika *fuzzy*), diperoleh hasil berupa 4 kesalahan data dari 70 data keseluruhan. Dari hasil tersebut dapat dipersentasekan kesalahan sebesar,

$$4/70 \times 100 \% = 5,71 \%$$

Kesalahan ini diakibatkan oleh masih terdapat inkonsistensi dalam penentuan keputusan pemberian beasiswa sebelumnya, sehingga berimbas kepada terdapatnya kesalahan pada data pada saat diuji dengan *rule viewer*.

4. Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa proses seleksi penerima beasiswa dipengaruhi oleh beberapa variabel, yang secara garis besar terbagi atas 2 jenis, yaitu variabel kualitatif dan variabel kuantitatif. Sedangkan 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 penerapan logika *fuzzy*, yaitu sebesar 0,261 untuk *Root Mean Square Error* (RMSE) dan 0,094 untuk *Mean Square Error* (MAE).

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.

Referensi

- Adeli, A., & Neshat, M. (2010). A Fuzzy Expert System for Heart Disease Diagnosis. *Proceedings of The International Multi Conference of Engineers and Computer Scientist (IMECS) 2010 Vol.1*. Hongkong.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, 2nd Edition*. San Francisco: Morgan Kaufmann Publisher.
- Hermawan, A. T., Gunawan, & Mahono, Y. C. (2009). Decision Support System Tool untuk Penyelesaian Permasalahan Linear Berbasis Simplex dan Revised Simplex.

- Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2009*. Yogyakarta.
- Kandel, A. (2001). *Fuzzy Expert System*. Florida: CRC Press.
- Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan, Edisi 2*. Yogyakarta: Graha Ilmu.
- Marimin. (2005). *Teori dan aplikasi Sistem Pakar dalam Teknologi Manajerial*. Bogor: IPB Press.
- Microstrategy. (n.d.). *Decision Support System*. Retrieved April 22, 2012, from www.microstrategy.com/decision-support-system/.
- Moertini, V. S. (2007). *Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi (Studi Kasus: Pra-Pengolahan Dan Klasifikasi Citra Batik)*. Bandung: Program Studi Teknik Informatika – ITB.
- Negnevitsky, M. (2005). *Artificial Intelligence; A Guide to Intelligent Systems, 2nd Edition*. Harlow: Pearson Education Limited.
- Setiawan, B. (2010). *Perancangan Sistem Pendukung Keputusan (SPK) Untuk Menentukan Kelaiklautan Kapal*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Siler, W., & Buckley, J. (2005). *Fuzzy Expert Systems and Fuzzy Reasoning*. New Jersey: John Wiley & Sons.
- Sivanandam, S. N., Sumathi, S., & Deepa, S. N. (2007). *Introduction to Fuzzy Logic using MATLAB*. Berlin: Springer.
- Subakti, I. (2002). *Buku Panduan IF1524: Sistem Pendukung Keputusan*. Surabaya: FTI – Institut Teknologi Sepuluh Nopember.
- Sunjana. (2010). *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree*. SNATI 2010. Yogyakarta.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining; Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi Offset.
- Wang, T.-C., & Lee, H.-D. (2006). *Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy*. Taiwan: I-Shou University & Fortune Institute of Technology.

BIODATA PENULIS

Ade Surya Budiman, S.T., M.Kom.

Tahun 2007 lulus dari Program Strata Satu (S1) Program Studi Teknik Elektro, Fakultas Teknik, Universitas Andalas, Sumatera Barat. Tahun 2010 lulus dari Program Strata Dua (S2) Program Studi Ilmu Komputer, konsentrasi *e-business* pada STMIK Nusa Mandiri Jakarta. Bidang penelitian yang pernah dan sedang dikerjakan adalah manajemen pengetahuan, pengolahan citra digital, perbaikan berkelanjutan (*continuous improvement*) pada industri manufaktur, *data mining* dan *E-Government*. Bekerja sebagai staf *Engineering* di PT. Indonesia Nippon Steel Pipe, Cikampek, Jawa Barat hingga tahun 2009. Dari tahun 2009 hingga saat ini, bekerja sebagai dosen dengan jabatan akademik Asisten Ahli di AMIK BSI Jakarta.

Xanty Adhi Parandani, S.T., M.Kom.

Tahun 2008 lulus dari Program Strata Satu (S1) Program Studi Teknik Elektro, Universitas Gajayana, Malang, Jawa Timur. Melanjutkan pendidikan di jenjang Strata Dua (S2) di STMIK Nusa Mandiri, Jakarta dan lulus pada tahun 2014. Dari tahun 2009 hingga saat ini bekerja sebagai dosen di AMIK BSI Bogor. Mendalami bidang penelitian Data Mining Sistem Pakar dan Implementasi Komputasi Statistik.