

PENENTUAN TALENTA KARYAWAN BERDASARKAN MENGUNAKAN KONSEP DATA MINING

Ni G. A. P. Harry Saptarini

Abstrak :

The human resources competency is one aspect that most affect company performances. Knowing its competence will help the decision maker to place the right man on the right place. However it can be done by analyzing the human talent. One of the conventional method that common to use is C4.5 algorithm. Conventional C4.5 method uses data crisp. In this study, it used a linguistic term as an input data because the talent test expressed using the language (linguistic term) and it is a set of data in fuzzy form. To generate the input data in the form of fuzzy was done by fuzzify the data preprocessing and further preprocessing of data results will be used for the construction of decision trees using C4.5 algorithm is then the process is called fuzzy C4.5.

The result of this research is that the number of linguistic terms of attribute effect directly and significantly to system accuracy. The accuracy of Fuzzy C4.5 algorithm with 5 linguistic terms (90.1099%) is less than Fuzzy C4.5 accuracy with 3 linguistic terms (96.7033%). Fuzzy C4.5 accuracy with 3 linguistic terms has the same accuracy with conventional C4.5 (96.7033 %), so we can use as an alternative solution to build classification of employee talent in PNB.

Keyword : Talent, classification, C4.5 Algorithm, Fuzzy.

PENDAHULUAN

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data..

Data mining terutama digunakan untuk men-cari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery in Data bases (KDD).

Proses pencarian pengetahuan ini menggunakan berbagai teknik-teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksikannya. Proses pencarian bersifat iteratif dan interaktif untuk menemukan pola atau model yang sah, baru, bermanfaat, dan dimengerti. Dalam pene-

rapannya data mining memerlukan berbagai perangkat lunak analisis data untuk menemukan pola dan relasi data agar dapat digunakan untuk membuat prediksi dengan akurat.

Penelitian yang dilakukan oleh Jantan dkk. (2010) untuk klasifikasi talenta dengan algoritma C4.5 menggunakan data dalam bentuk himpunan tegas/nyata (*crisp*) yang berupa bilangan numerik/angka sebagai input untuk membangun pohon keputusan, namun pada kenyataannya penilaian talenta terutama di PNB dijustifikasi oleh pakar psikolog dalam bentuk bahasa (*linguistic term*) dan merupakan data dalam bentuk himpunan tidak tegas (*fuzzy*). Pada penelitian yang dilakukan oleh Jantan dkk. (2010) ini, tingkat bakat seseorang dinyatakan dalam bentuk himpunan tegas/nyata (*crisp*) seperti misalkan untuk talenta dengan kriteria *personal1* digunakan range angka yaitu

$personal1 \leq 22.26$ dan $personal1 > 22.26$ untuk mendapatkan kesimpulan, sedangkan dalam kenyataannya nilai talenta dapat berupa himpunan tidak tegas (*fuzzy*) yang dapat digolongkan ke dalam nilai baik, cukup dan kurang, sehingga lebih mudah dibaca dan dipahami oleh orang awam (Novak, 2005). Jika kita definisikan bahwa “kemampuan personal yang dimiliki oleh seseorang adalah baik” yaitu orang yang memiliki kemampuan personal lebih besar dari 22.26, maka orang yang memiliki kemampuan personal 22.25 menurut definisi tersebut termasuk orang yang tidak memiliki kemampuan personal yang baik. Sulit bagi kita untuk menerima bahwa yang memiliki nilai 22.25 itu tidak termasuk orang yang berkemampuan personal baik. Hal ini menunjukkan bahwa memang batas antara kelompok orang yang berkemampuan personal baik dan berkemampuan personal tidak baik tidak dapat ditentukan dengan tegas (Klir dan Yuan, 1995).

Penelitian ini menggunakan data awal berupa data dalam bentuk himpunan tegas/nyata (*crisp*) yang berupa bilangan numerik/angka. Untuk menterjemahkan data tersebut ke dalam bentuk *linguistic term* inilah, peneliti menggunakan logika *fuzzy* (proses fuzzifikasi) untuk proses *preprocessing* data yang selanjutnya data yang dihasilkan akan digunakan sebagai input dalam pembangunan pohon keputusan untuk klasifikasi talenta karyawan dengan menggunakan Algoritma C4.5. Dalam penulisan berikutnya, akan digunakan istilah *fuzzy* C4.5 untuk proses pembangunan pohon keputusan menggunakan Algoritma C4.5 dengan data yang telah mengalami proses *preprocessing* menggunakan logika *fuzzy*.

METODE PENELITIAN

Peneliti menguji performansi metode *fuzzy* C4.5 dengan cara mengimplementasikannya pada sistem klasifikasi talenta karyawan Politeknik Negeri Bali. Sistem tersebut memiliki 4 buah komponen utama, yaitu: (i) dataset talenta karyawan (*input*), (ii) proses fuzzifikasi

dan konversi ke *linguistic term* (*preprocessing*), (iii) induksi pohon keputusan dan (iv) pembuatan daftar aturan (*rules*). Sistem yang dibangun peneliti memiliki bagian-bagian yang sama seperti sistem sebelumnya namun perbedaannya terletak pada bagian *preprocessing* dimana peneliti menggunakan metode *fuzzy* C4.5.

1. Dataset yang digunakan

Format dataset berbentuk nilai-nilai numerik dan terdiri dari 9 buah atribut penentu serta 1 buah atribut keputusan. Atribut keputusan diberi nama *Job Function (JF)* dan terdiri dari 4 buah *class*, yaitu: (i) ADMINISTRASI, (ii) HUMAS, (iii) TEKNISI, dan (iv) KEUANGAN, sedangkan identitas dari ke-9 buah atribut penentu yaitu SE, WA, AN, GE, ME, RA, ZR, FA dan WU dengan fungsi keanggotaan seperti pada table 1. Atribut-atribut penentu tersebut dipilih oleh pakar karena dianggap mampu merepresentasikan talenta karyawan secara menyeluruh (*comprehensive*).

2. Tahap *preprocessing*: proses fuzzifikasi dan konversi data ke *linguistic term*

Tahap ini merupakan pembeda antara sistem yang diusulkan peneliti dengan sistem-sistem sebelumnya. Ada 2 proses utama yang terdapat pada tahap *preprocessing* yaitu proses fuzzifikasi untuk merubah data numerik (*crisp*) menjadi *linguistic term* (*fuzzy*). Proses fuzzifikasi bertujuan untuk mengubah nilai numerik menjadi kabur (*fuzzy*) dengan menggunakan bantuan sebuah basis data. Basis data yang digunakan untuk proses fuzzifikasi berisi himpunan fungsi-fungsi keanggotaan dari himpunan-himpunan *fuzzy* yang terkait dengan nilai-nilai linguistik dari seluruh atribut penentu (9 buah atribut) yang terlibat dalam sistem. Oleh karena itu, basis data untuk yang menggunakan 5 *linguistic term* berjumlah 45 buah *record* dan yang menggunakan 3 *linguistic term* berjumlah 27 buah *record*. Konten lengkap dari basis data yang menggunakan 5 *linguistic term* tampak pada Tabel 1 dan yang menggunakan 3

linguistic term tampak pada Tabel 2 dimana untuk fungsi keanggotaan segitiga ditandai oleh adanya tiga parameter {a,b,c} dan untuk fungsi keanggotaan trapezium ditandai dengan empat parameter {a, b, c, d}.

Tabel 1. Basis Data untuk Proses Fuzzifikasi (5 linguistic term)

Atribut	Simbol	Linguistik	Parameter
<i>Practical thinking</i>	SE	Kurang	0;0;89;99
	SE	Rendah	0;0;100.5;113.5
	SE	Cukup	100;114;128
	SE	Tinggi	114;128;180;180
	SE	Excellent	128;142;180;180
<i>Verbal integrative</i>	WA	Kurang	0;0;76.5;89.5
	WA	Rendah	0;0;89.5;102.5
	WA	Cukup	90;102;114
	WA	Tinggi	101.5;114.5;135;135
<i>Flexibility thinking</i>	AN	Kurang	0;0;63.5;92.5
	AN	Rendah	0;0;92.5;121.5
	AN	Cukup	92.5;121.5;150.5
	AN	Tinggi	121.5;150.5;204;204
	AN	Excellent	150.5;179.5;204;204
<i>Verbal abstrack</i>	GE	Kurang	0;0;84;96
	GE	Rendah	0;0;96;108
	GE	Cukup	96.5;107.5;118.5
	GE	Tinggi	107;119;143;143
	GE	Excellent	119;131;143;143
<i>Working memory</i>	ME	Kurang	0;0;72;90
	ME	Rendah	0;0;90;108
	ME	Cukup	90;108;126
	ME	Tinggi	108;121.5;132;132
	ME	Excellent	121.5;130.5;132;132
<i>Numerikal reasoning</i>	RA	Kurang	0;0;73.5;90.5
	RA	Rendah	0;0;90.5;107.5
	RA	Cukup	90;108;126
	RA	Tinggi	107.5;126.5;160;160
	RA	Excellent	126.5;145.5;160;160
<i>Numerikal theoretic</i>	ZR	Kurang	0;0;74.5;91.5
	ZR	Rendah	0;0;91.5;108.5
	ZR	Cukup	91;109;127
	ZR	Tinggi	109;127;200;200
	ZR	Excellent	127;145;200;200

<i>Constructive thinking</i>	FA	Kurang	0;0;68;84
	FA	Rendah	0;0;84;100
	FA	Cukup	83.5;100.5;117.5
	FA	Tinggi	101;117;149;149
	FA	Excellent	117;133;149;149
<i>Technical abstract</i>	WU	Kurang	0;0;91;101
	WU	Rendah	0;0;101;111
	WU	Cukup	101;111;121
	WU	Tinggi	111;121;141;141
	WU	Excellent	121;131;141;141

Tabel 2. Basis Data untuk Proses Fuzzifikasi (3 linguistic term)

Atribut	Simbol	Linguistik	Parameter
<i>Practical thinking</i>	SE	Rendah	0;0;100.5;113.5
	SE	Cukup	100;114;128
	SE	Tinggi	114;128;180;180
<i>Verbal integrative</i>	WA	Rendah	0;0;89.5;102.5
	WA	Cukup	90;102;114
	WA	Tinggi	101.5;114.5;135;135
<i>Flexibility thinking</i>	AN	Rendah	0;0;92.5;121.5
	AN	Cukup	92.5;121.5;150.5
	AN	Tinggi	121.5;150.5;204;204
<i>Verbal abstrack</i>	GE	Rendah	0;0;96;108
	GE	Cukup	96.5;107.5;118.5
	GE	Tinggi	107;119;143;143
<i>Working memory</i>	ME	Rendah	0;0;90;108
	ME	Cukup	90;108;126
	ME	Tinggi	112.5;121.5;132;132
<i>Numerical reasoning</i>	RA	Rendah	0;0;90.5;107.5
	RA	Cukup	90;108;126
	RA	Tinggi	107.5;126.5;160;160
<i>Numerical theoretic</i>	ZR	Rendah	0;0;91.5;108.5
	ZR	Cukup	91;109;127
	ZR	Tinggi	109;127;200;200
<i>Constructive thinking</i>	FA	Rendah	0;0;84;100
	FA	Cukup	83.5;100.5;117.5
	FA	Tinggi	101;117;149;149
<i>Technical abstract</i>	WU	Rendah	0;0;101;111
	WU	Cukup	101;111;121
	WU	Tinggi	111;121;141;141

Tabel 3. Contoh Dataset Talenta Karyawan

ID	Daftar Atribut					Atribut Tujuan
	SE	WA	AN	GE	ME	JOB
1	115	61	155	125	84	Humas
2	138	131	200	120	121	Humas
3	147	87	116	90	81	Teknisi
4	138	40	194	142	52	Humas
5	149	117	148	127	117	Humas

Dengan menggunakan basis data seperti pada Tabel 1, maka dataset numeric pada Tabel 3 setelah mengalami proses *preprocessing* dan akan menghasilkan dataset nominal yang mengandung data dalam bentuk *fuzzy* dengan nilai untuk setiap atribut adalah K untuk kurang, R untuk rendah, C untuk cukup, T untuk Tinggi dan E untuk excellent seperti pada Tabel 4 untuk 5 *linguistic term* dan Tabel 5 untuk 3 *linguistic term*. Misalkan diberikan data numerik dari atribut “*Practical Thinking*” (SE) sebesar 115 dan fungsi keanggotaan seperti pada Tabel 1. Proses fuzzifikasi dan konversi ke 5 *linguistic term* sebagai berikut:

$$\mu_{\text{kurang}} = \frac{115 - 0}{115 - 0} = 1$$

$$\mu_{\text{rendah}} = \frac{115 - 115}{115 - 115} = 0$$

$$\mu_{\text{cukup}} = \frac{115 - 115}{115 - 115} = 0$$

$$\mu_{\text{tinggi}} = \frac{115 - 115}{115 - 115} = 0$$

$$\mu_{\text{excellent}} = \frac{115 - 115}{115 - 115} = 0$$

$$\mu_{\text{kurang}} \cup \mu_{\text{rendah}} \cup \mu_{\text{cukup}} \cup \mu_{\text{tinggi}} \cup \mu_{\text{excellent}} = \max[0, 0, 0, 0, 0] = 0$$

Dengan demikian, nilai 115 akan dikonversi ke variabel linguistik **C (cukup)**.

Tabel 4. Dataset nominal (5 linguistic term)

ID	Daftar Atribut					Atribut Tujuan
	SE	WA	AN	GE	ME	JOB
1	Cukup (0.928)	Kurang (1)	Tinggi (1)	Tinggi (1)	Rendah (1)	Humas (0.928)
2	Tinggi (1)	Excellent (1)	Excellent (1)	Tinggi (1)	Tinggi (0.962)	Humas (0.962)
3	Excellent (1)	Rendah (1)	Cukup (0.810)	Rendah (1)	Rendah (1)	Teknisi (0.81)
4	Tinggi (1)	Kurang (1)	Excellent (1)	Excellent (1)	Kurang (1)	Humas (1)
5	Excellent (1)	Tinggi (1)	Tinggi (0.913)	Tinggi (1)	Tinggi (0.666)	Humas (0.666)

Tabel 5. Contoh Dataset nominal (3 linguistic term)

ID	SE	WA	AN	GE	ME	JOB
1	Cukup (0.928)	Rendah (1)	Tinggi (1)	Tinggi (1)	Rendah (1)	Humas (0.928)
2	Tinggi (1)	Tinggi (1)	Tinggi (1)	Tinggi (1)	Tinggi (0.944)	Humas (0.944)
3	Tinggi (1)	Rendah (1)	Cukup (0.810)	Rendah (1)	Rendah (1)	Teknisi (0.81)
4	Tinggi (1)	Rendah (1)	Tinggi (1)	Tinggi (1)	Rendah (1)	Humas (1)
5	Tinggi (1)	Tinggi (1)	Tinggi (0.913)	Tinggi (1)	Cukup (0.5)	Humas (0.5)

3. Tahap pembentukan pohon keputusan dengan menggunakan algoritma C4.5

Peneliti menggunakan teknik rekursif dalam mengklasifikasi talenta karyawan sehingga penjelasan mengenai pembuatan pohon keputusan berbasis algoritma C4.5, dimana prinsip kerja dari algoritma C4.5 adalah sebagai berikut:

1. Dataset yang berformat nominal, atribut keputusan dan atribut-atribut penentu berfungsi sebagai data masukan. Struktur data untuk menyimpan atribut keputusan, atribut-atribut penentu dan dataset adalah *ActiveX Data Object (ADO) recordset*. *ADO recordset* yang berisi dataset dinamakan *SAMPLE*.
2. Buat *Root Node* berdasarkan pada *SAMPLE* tersebut.
3. Periksa, apakah isi dari *SAMPLE* terdiri dari *Class* yang sama? Jika ya, maka buat *node* dengan *Class* tersebut sebagai *node*-nya kemudian *return* dan *exit*. Jika tidak, maka lanjut ke langkah 4.
4. Periksa, apakah atribut penentu dalam *ADO recordset* kosong? Jika ya, maka buat *node* dengan nilai terbanyak dari atribut keputusan sebagai *node*-nya kemudian *return* dan *exit*. Jika tidak, maka lanjut ke langkah 5.
5. Cari atribut dengan *gain ratio* terbesar

Misal *S* berisi *s* data samples. Anggap atribut untuk class memiliki *m* nilai yang berbeda, C_i (untuk $i=1, \dots, I$). Anggap s_i menjadi jumlah samples *S* pada class C_i , maka menurut Han dan Kamber (2006) besar informasi-nya dapat dihitung dengan:

$$Info(s_1, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

dimana:

$p_i = \frac{s_i}{s}$ adalah probabilitas dari sample yang mempunyai class C_i .

Misalkan atribut A mempunyai v nilai yang berbeda $\{a_1, a_2, \dots, a_v\}$. Atribut A dapat digunakan untuk mempartisi S menjadi v subset $\{S_1, S_2, \dots, S_v\}$, dimana S_j berisi sample pada S yang memiliki nilai a_j dari A. Jika A terpilih menjadi data test atribut (yaitu best atribut untuk plitting), maka subset-subset akan berhubungan dengan pertumbuhan node-node cabang yang berisi S. Anggap s_{ij} sebagai jumlah samples class C_i pada subset S_j . Menurut Han dan Kamber (2006), entropy atau nilai information dari subset A adalah:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} \times I(s_1, \dots, s_m) \quad (2)$$

$\frac{s_{1j} + \dots + s_{mj}}{s}$ adalah bobot dari subset j dan jumlah samples pada subset (yang mempunyai nilai a_j dari A) dibagi dengan jumlah total samples pada S. Menurut Han dan Kamber (2006), nilai information gain atribut A pada subset S adalah

$$Gain(A) = Info(s_1, \dots, s_m) - E(A) \quad (3)$$

Untuk menghitung gain digunakan persamaan seperti tertera dalam persamaan(3.14) (Han dan Kamber, 2006) adalah

$$GainRatio(A) = \frac{Gain(A)}{SplitInformation(A)} \quad (4)$$

$$SplitInformation(A) = - \sum_{j=1}^v p_j \times \log_2(p_j) \quad (5)$$

dimana :

$p_j = \frac{s_j}{|s|}$ adalah probabilitas sample S_j yang diketahui nilainya.

6. Buat sebuah *group* yang merupakan *subsets* dari SAMPLE untuk setiap nilai (partisi) dari atribut terpilih.
7. Untuk setiap *subset* yang tidak kosong, jadikan *subset* tersebut sebagai SAMPLE kemudian kurangkan daftar atribut dengan atribut yang terpilih.
8. Buat *node* berdasarkan pada SAMPLE tersebut kemudian lompat ke langkah 3.

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini diperoleh bahwa pengujian Classifier dilakukan dengan tujuan sebagai berikut :

1. Memeriksa perbandingan tingkat akurasi yang diperoleh antara pengklasifikasian talenta pegawai menggunakan algoritma *fuzzy* C4.5 dengan algoritma C4.5 biasa (konvensional). Kedua sistem tersebut akan diuji dengan menggunakan dataset referensi yang sama.
2. Memeriksa pengaruh jumlah himpunan *fuzzy* terhadap akurasi *classifier*

Dataset yang digunakan terdiri dari 182 records, 9 buah atribut penentu yang berformat *numeric* dan sebuah atribut keputusan yang berformat bersifat *categorical*.

Pada pengujian tahap I, fungsi keanggotaan digunakan sebagai referensi untuk mengubah format dataset dari bentuk *numeric* menjadi bentuk nominal (*ling-uistic term*). Bentuk fisik dari fungsi keanggotaan adalah sebuah file Microsoft Excel yang bernama setting.xls yang terdiri dari 45 records (5 buah *linguistic term* x 9 atribut). Salah satu atribut yang digunakan adalah atribut *practical thinking* (SE). Atribut SE dibagi ke dalam 5 buah *linguistic term*, yaitu "Tidak Mampu" (berbentuk trapesium: 0;0;89;99), "Rendah" (berbentuk triangular: 87.5;100.5;113.5), "Cukup" (berbentuk triangular: 100;114;128), "Tinggi" (berbentuk triangular: 114;128;142) dan "Excellent" (berbentuk trapesium: 128;142;180;180).

Proses pengujian sebagai berikut:

1. Sistem klasifikasi talenta pegawai akan diberikan 2 buah data masukan yaitu:
 - (i) dataset talenta pegawai yang berformat *numeric* dan terdiri dari 182 record dan
 - (ii) dataset fungsi keanggotaan (berisi 5 buah *linguistic term*) yang terdiri dari 45 record.
2. Sistem (*classifier*) akan membaca kedua dataset tersebut kemudian melakukan proses transformasi data (fuzzifikasi). Luaran system adalah sebuah file yang bernama nominal.xls yang berisi 182 record data talenta pegawai yang telah berubah format datanya dari *numeric* menjadi berformat nominal.

3. Dataset nominal (hasil transformasi) dibagi menjadi *training set* dan *testing set*. Pembagian data ini menggunakan metode *10-folds cross validation*. Data akan dibagi menjadi 10 *folds* yang berbeda dengan jumlah yang sama besar. Setiap kali sebuah *fold* digunakan sebagai *testing set* maka 9 buah *fold* yang lain akan berfungsi sebagai *training set*.

4. Hitung akurasi *classifier*.

5. Ulangi langkah 1 dengan menggunakan dataset fungsi keanggotaan yang lain (berisi 3 buah *linguistic term*).

Jalannya proses pengujian untuk *classifier* berbasis algoritma C4.5 konvensional berbeda dengan *classifier* berbasis *fuzzy* C4.5 karena tidak menggunakan dataset fungsi keanggotaan sehingga dataset masukan hanya satu saja yaitu dataset talenta pegawai. Selain itu, pada *classifier* berbasis algoritma C4.5 konvensional tidak terdapat proses transformasi data (fuzzifikasi).

Hasil pengujian sistem (*classifier*) berbasis algoritma *fuzzy* C4.5 dengan menggunakan dua dataset:

- (i) Dataset talenta karyawan yang berformat numerik dan
- (ii) Data masukan berbentuk fungsi keanggotaan. Dataset fungsi keanggotaan yang digunakan adalah fungsi keanggotaan dengan 5 buah *linguistic term* dan 3 buah *linguistic term*.

Sedangkan pada *classifier* berbasis algoritma C4.5 konvensional digunakan satu dataset masukan saja yaitu dataset talenta karyawan.

Tujuan penggunaan himpunan *fuzzy* (*linguistic term*) yang berbeda-beda pada *classifier* berbasis algoritma *fuzzy* C4.5 adalah ingin melihat berapa banyak jumlah himpunan *fuzzy* (*linguistic term*) yang ideal agar akurasi *classifier* tinggi. Tabel 6 menunjukkan table perbandingan akurasi *classifier* terhadap jumlah *linguistic term* yang berbeda-beda.

Tabel 6. Hasil Pengujian

Metode yang Digunakan	10 Dataset	25 Dataset	182 Dataset
	Akurasi (%)	Akurasi (%)	Akurasi (%)
<i>Fuzzy</i> C4.5 dengan 5 buah <i>linguistic term</i>	90	60	90.1099
<i>Fuzzy</i> C4.5 dengan 3 buah <i>linguistic term</i>	90	68	96.7033
C4.5 konvensional dengan <i>binary split</i>	80	64	96.7033

Hasil penelitian menemukan bahwa untuk kasus klasifikasi talenta karyawan PNB dengan menggunakan *fuzzy* C4.5 bahwa akurasi algoritma *fuzzy* C4.5 dengan 3 buah *linguistic term* memiliki tingkat akurasi tertinggi dibandingkan dengan metode yang lainnya dan dapat dijadikan sebagai rekomendasi untuk pembangunan system klasifikasi talenta karyawan di PNB.

Proses pencarian dan penentuan keputusan (*class*) untuk dataset baru yang belum pernah digunakan sebelumnya berdasarkan pada *rule-rule* yang dihasilkan oleh model *classifier* menggunakan *fuzzy* C4.5 dengan 3 *linguistic term* dimana untuk data

SE=131;WA=92;AN=170;GE=132;ME=122;RA=135;ZR=65;FA=113;WU=81 diperoleh hasil klasifikasi adalah SE = Tinggi (1); WA = Rendah (0.807); AN = Tinggi (1); GE = Tinggi (1); ME = Tinggi (1); RA = Tinggi (1); ZR = Rendah (1); FA = Tinggi (0.75); WU= Rendah (1); → HUMAS (0.75)

KESIMPULAN

Penelitian ini menghasilkan kesimpulan bahwa tahap *preprocessing* dengan proses fuzzifikasi terdiri dari dua proses utama yaitu:

1. Proses untuk mendapatkan derajat keanggotaan dan
2. Proses konversi ke *linguistic term*.
3. Proses fuzzifikasi bertujuan untuk mengubah nilai *crisp(numeric)* menjadi *fuzzy* (*linguistic term*) dengan menggunakan bantuan sebuah basis data. Basis data tersebut berisi himpunan fungsi-fungsi keanggotaan dari himpunan-himpunan *fuzzy*

yang terkait dengan nilai-nilai linguistik dari seluruh atribut penentu yang terlibat dalam sistem. Penggunaan logika *fuzzy* dalam tahap *pre-processing* dan algoritma C4.5 dalam membentuk pohon keputusan (*fuzzy* C4.5) dapat digunakan sebagai solusi alternatif untuk mengolah dataset masukan yang berformat numeric, dimana jumlah *linguistic term* dari suatu atribut berpengaruh secara langsung dan signifikan terhadap akurasi sistem.

DAFTAR PUSTAKA:

- Abidin, A.Z.Z., 2011, Implementasi Algoritma C 4.5 Untuk Menentukan Tingkat Bahaya Tsunami, *Seminar Nasional Informatika*, Yogyakarta.
- Capelli, P., 2009, *Talent on Demand : Metode Baru Mendapatkan SDM Bertalenta Tepat Jumlah, Tepat Kualifikasi, dan Tepat Waktu*, (diterjemahkan oleh : Wendra Triana Rafelina), PPM, Jakarta.
- Cintra, M.E., Monard, M.C., 2010, *Evaluation of The Pruning Impact on Fuzzy C4.5*, *Congresso Brasileiro de Sistemas Fuzzy, Brasil*.
- Cunningham, I., 2007, *Talent Management: Making it Real, Development and Learning in Organizations*, 2, 21, 4-5.
- Firqiani, H.N., Kustiyo, A., dan Giri, E. P., 2011, Seleksi Fitur Menggunakan Fast Correlation Based Filter Pada Algoritma Voting Feature Intervals 5, *Jurnal Ilmiah Ilmu Komputer*, IPB, Bogor.
- Han, J., dan Kamber, M., 2006, *Data Mining Concept and Technique*, Second Edition, Morgan Kaufman Publisher, San Francisco.
- Jantan, H., Razak, H.A., dan Zulaiha, A.O., 2009, Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application, *Proceedings Of World Academy Of Science, Engineering And Technology*, 38, 803-811.
- Jantan, H., Razak, H.A., dan Zulaiha, A.O., 2009, Potential Data Mining Classification Techniques for Academic Talent Forecasting, *International Conference on Intelligent Systems Design and Applications*, 9, 1173-1178.
- Jantan, H., Razak, H.A., dan Zulaiha, A.O., 2010, Human Talent Prediction in SDMM using C4.5 Classification Algorithm, *International Journal on Computer Science and Engineering*, 2, 8, 2526-2534.
- Jing, H., 2009, Application of Fuzzy Data Mining Algorithm in Performance Evaluation of Human Resource, *International Forum on Computer Science-Technology and Applications*, 343-346.
- Klir, G.J. dan Yuan, B., 1995, *Fuzzy Sets and Fuzzy Logic : Theory and Applications*, Prentice Hall, USA.
- Kusrini, Hartati, S., Wardoyo, R., dan Harjoko, A., 2009, Perbandingan Metode Nearest Neighbor dan Algoritma C4.5 Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa di STMIK AMIKOM Yogyakarta, *DASI*, 10, 1, 114-132.
- Kusrini, 2010, Pemodelan Analisis Diagnosis Banding Pada Sistem Pendukung Keputusan Kelompok Untuk Asesmen Geriatri, *Disertasi*, S3 Ilmu Komputer FMIPA UGM, Yogyakarta.
- Larose, D.T., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Willey & Sons. Inc., Canada.
- Maharani, W., 2009, Klasifikasi Data Menggunakan JST Backpropagation Momentum dengan Adaptive Learning Rate, *Seminar Nasional Informatika*, Yogyakarta.

- Manopo, C., 2011, *Competency Based Talent and Performance Management System*, Salemba Empat, Jakarta.
- Morgan, H. dan Jardin, D. (2010) HR + OD = Integrated Talent Management, *OD Practitioner*, 42, 4, 23-30.
- Negnevitsky, M., 2005, *Artificial Intelligence*, Second Edition, Addison-Wesley, Inggris.
- Novak, V., Perfilieva, I., Dvorak, A., Chen, G., Wei, Q., dan Yan, P., 2005, *Discovering Linguistic Associations from Numerical Data*, Elsevier Science, Cina.
- Pella, D.A. dan Inayati, A., 2011, *Talent Management Mengembangkan SDM Untuk Mencapai Pertumbuhan dan Kinerja Prima*, PT Gramedia Pustaka Utama, Jakarta.
- Rachmadhan, A.E., 2011, Pengaruh Talent Management dan Performance Appraisal Terhadap Succession Planning Pada Karyawan Direktorat HCGA & Unit HRC di Kantor Pusat PT Telekomunikasi Indonesia, Tbk, *Tesis*, Magister Manajemen, Universitas Padjadjaran, Bandung.
- Romansyah, F., Sitanggang, S., Nurdiati, S., 2009, Fuzzy Decision Tree dengan Algoritma ID3 pada Data Diabetes, *Internetworking Indonesia Jurnal*, 1, 2,45-52.
- Sujana, 2010, Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5, *Seminar Nasional Aplikasi Teknologi Informasi*, Yogyakarta.
- Wang, L.-X., 1997, *A Course in Fuzzy Systems and Control*, Prentice Hall, Hong Kong.
- Wang, T.C., dan Lee, H.D., 2006, Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy, *WSEAS Transactions on Information Science and Applications*, 3, 8, 1547–1552.
- Williams, G.J., Simmof, S.J., 1998, *Data Mining Theory, Methodology, Techniques and Applications*, Spinger, New York.
- Witten, I.H., Frank, E., 2005, *Data Mining Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufman Publisher, San Francisco.