

Penerapan Algoritma C4.5 Pada NUPTK Untuk Menentukan Pola Sertifikasi Guru Dengan Menggunakan Metode Klasifikasi Decision Tree (Studi Kasus : Pada Dinas Pendidikan Pemuda dan Olahraga Kabupaten Rokan Hulu)

Jailani, Sarjon Defit, Gunadi Widi Nurcahyo

Program Studi Pasca Sarjana Teknik Informatika
Universitas Putra Indonesia “YPTK” Padang
Jl. Raya Lubuk Begalung, Padang

Email : jailani.rohul@yahoo.co.id

Abstrak : Nomor Unik Pendidik dan Tenaga Kependidikan (NUPTK) merupakan data penting bagi Departemen Pendidikan Nasional. NUPTK terdiri dari 16 angka yang bersifat tetap karena NUPTK yang dimiliki seorang Pendidik dan Tenaga Kependidikan (PTK) tidak akan berubah meskipun yang bersangkutan telah berpindah tempat mengajar atau terjadi perubahan data riwayatnya. Pada program sertifikasi, PTK diwajibkan mengumpulkan data diri berupa Nomor Identitas (ID), kepemilikan NUPTK, tingkat pendidikan, usia, masa kerja, golongan guru dan jenis sertifikasi. Dari data-data ini, akan ditentukan seseorang lulus atau tidak dalam sertifikasi guru. Dalam menentukan kelulusan, dibuatlah suatu syarat-syarat tertentu yang menjadi acuan penilaian. Tetapi diperlukan suatu syarat yang menjadi prioritas utama penetapan kelulusan sertifikasi guru. Algoritma C4.5 adalah merupakan salah satu algoritma dalam teknik klasifikasi data yang menghasilkan satu pohon keputusan. Dari pohon keputusan yang dihasilkan maka akan terdapat beberapa rule atau pengetahuan untuk suatu kasus. Teknik klasifikasi yang akan digunakan adalah Decision tree, yaitu algoritma C4.5. Dengan implementasi Algoritma C4.5 dapat membantu pihak pengambil keputusan/kebijakan dalam menentukan calon sertifikasi guru layak atau dipertimbangkan untuk mengikuti Pendidikan Profesi Guru Dalam Jabatan (PPGJ).

Kata kunci: NUPTK, Data mining, Algoritma C4.5, Klasifikasi, Decision tree

Abstract : Unique number Teachers and Education Personnel (NUPTK) is an important data for the Ministry of Education. NUPTK consists of 16 numbers that are fixed because NUPTK held a Teachers and Education Personnel (PTK) will not change even though the party concerned has migrated to teach or change the data transmission. In the certification program, PTK are required to collect personal data in the form of Identity Number (ID), NUPTK ownership, education level, age, tenure, class teachers and type of certification. From these data, we determined a person pass or not the teacher certification. In determining graduation, they invented a certain conditions is the reference assessment. But it required a requirement that a top priority-setting graduation teacher certification. C4.5 algorithm is one of the algorithms in data classification technique that produces a decision tree. Of the decision tree is generated there will be some rule or knowledge to a case. Classification techniques that will be used is Decision tree, the algorithm C4.5. With the implementation of the C4.5 algorithm can help the decision maker / policy in determining the worth or teacher certification candidates considered for the Professional Teacher Education In Position (PPGJ)

Keywords: NUPTK, Data mining, C4.5 algorithm, Classification, Decision tree

PENDAHULUAN

Nomor Unik Pendidik dan Tenaga Kependidikan (NUPTK) merupakan data penting bagi Departemen Pendidikan Nasional. NUPTK terdiri dari 16 angka yang bersifat tetap karena NUPTK yang dimiliki seorang Pendidik dan Tenaga Kependidikan (PTK) tidak akan berubah meskipun yang bersangkutan telah berpindah tempat mengajar atau terjadi perubahan data riwayatnya. Kebijakan yang diberikan oleh pemerintah yang berhubungan dengan PTK, salah satunya adalah

sertifikasi guru dan penentuan calon peserta sertifikasi guru bersumber pada informasi yang diperoleh dari data NUPTK.

TINJAUAN PUSTAKA

Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang bermanfaat yang terdiri dari serangkaian tahap transformasi meliputi data preprocessing dan postprocessing.

Dimana data preprocessing merupakan langkah awal untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis selanjutnya. Data preprocessing kemungkinan akan membutuhkan waktu yang sangat lama, hal ini dikarenakan data yang mentah kemungkinan disimpan dengan format dan database yang berbeda. Sedangkan Postprocessing meliputi semua operasi yang harus dilakukan agar hasil *Data Mining* dapat diakses dan lebih mudah diinterpretasikan oleh para pihak analisis (Sandro, 2013).

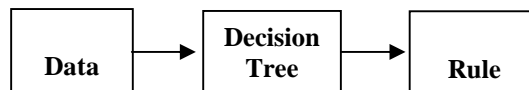
Data Mining

Data Mining adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *Data Mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Liliana, 2013).

Menurut (Sutrisno1 dan Afriyudi, 2013) *Data Mining* adalah proses yang menggunakan teknik statistik, perhitungan, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar.

Pohon Keputusan (Decision tree)

Menurut (Yusuf , 2012) Pohon keputusan adalah pohon yang ada dalam analisis pemecahan masalah, pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah [6]. Pohon Keputusan dapat juga dikatakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*).



Gambar 1. Konsep Decision Tree

Decision Tree sesuai digunakan untuk kasus- kasus dimana outputnya bernilai diskrit. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Algoritma C4.5

Algoritma C4.5 merupakan bagian dari kelompok algoritma decision trees dan merupakan katerogi 10 algoritma yang paling populer . *Algoritma C4.5* di perkenalkan oleh J. Ross Quinlan diakhir tahun 1970 hingga awal tahun 1980-an. J. Ross Quinlan seorang peneliti dibidang mesin pembelajaran yang merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser), algoritma tersebut digunakan untuk membentuk pohon keputusan.

Untuk menghitung entropi menggunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \tag{1}$$

Di mana :

S : himpunan kasus

A : atribut

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

Kemudian hitung nilai gain yang menggunakan rumus:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot Entropy(S_i) \quad (2)$$

Di mana :

S : himpunan kasus

A : fitur

n : jumlah partisi atribut A

$[S_i]$: proporsi S_i terhadap S

$[S]$: jumlah kasus dalam S

METODE PENELITIAN

Menganalisa teknik pengolahan data menggunakan algoritma C4.5, Data yang diperoleh dari tempat penelitian selanjutnya dilakukan analisa dan pengolahan menggunakan algoritma C4.5, Merancang *Data Mining* menggunakan algoritma C4.5, Analisa data dengan penerapan algoritma C4.5 selanjutnya dirancang dalam bentuk pohon keputusan (*decision tree*) sehingga menghasilkan suatu klasifikasi untuk menentukan pola sertifikasi guru. Melakukan Pengujian Pada tahap ini, penulis mengimplementasikan model dan hasil perancangan system menggunakan software *RapidMiner*. Sistem diuji dengan prosedur-prosedur untuk melakukan eksplorasi dan permodelan dari data-data yang ada sehingga mendapatkan suatu hubungan tersembunyi dari data tersebut. Mekanisme yang digunakan saat pengujian sistem adalah: Mula-mula data manual "Data Calon Sertifikasi Guru" yang sudah ada sebelumnya, di transformasikan ke dalam *Microsoft Excel*. Sebagai tahap awal di dalam proses pengujian yaitu mengimpor Data Calon Sertifikasi Guru yang akan dijadikan sebagai data tabel keputusan di dalam *tools Rapidminer Versi 5.3* tersebut. Kemudian tahap selanjutnya yaitu, melakukan proses *Import data excel* dan *Decision tree* dari data tabel keputusan tersebut. Dari proses *Import data excel* dan *Decision tree* tersebut akan terdapat pola atau kriteria tertentu berdasarkan karakteristik setiap Peserta Sertifikasi. Hasil dari Proses *Import data excel* dan *Decision tree* tersebut akan diolah kembali untuk menghasilkan suatu *Decision Tree*. Dari *Decision Tree* tersebut akan terlihat hasilnya. Hasilnya tersebut merupakan pengetahuan baru (*knowledge*) dan juga merupakan sebagai acuan untuk menentukan calon atau pola sertifikasi guru yang disebut dengan Pendidikan dan Pelatihan Profesi Guru dalam Jabatan (PPGJ).

HASIL DAN PEMBAHASAN

Analisa Data

Data Mining yang diterapkan dalam penelitian ini adalah berbentuk klasifikasi dengan menggunakan algoritma C4.5 dalam melakukan klasifikasi terhadap data sertifikasi guru. Pengklasifikasian yang dilakukan dengan menggunakan 4 atribut sebagai kriteria dalam menentukan status kelayakan terhadap guru untuk mendapatkan sertifikasi. Atribut yang digunakan

dalam penelitian ini adalah pendidikan. Atribut kedua adalah masa kerja. Atribut ketiga adalah golongan. Dan Atribut keempat adalah Usia. Keempat atribut tersebut dilihat dari data yang telah ada sebelumnya.

Tabel 1. Daftar Calon Peserta Sertifikasi Guru

No	Nama	NUPTK	Tempat Tugas	Pe nd.	Usia	MK	Gol.
1	ABDUL MANAP	6758737641 200002	SD N 022 TAMBUSAI	SM A	55_0 8	32_ 03	III/C
2	BONTOR MANIK	2559738640 200022	SD 014 PAGARAN TAPAH	D2	54_1 0	22_ 04	IV/A
3	NIZAR.SS	3737738640 200072	SMP N 4 RAMBAH HILIR	S1	54_0 8	06_ 09	III/D
4	ASRUN SAN	6937738640 200112	SMA S TIGA PUTRA PERSADA TAMBUSAI	S1	54_0 6	27_ 06	
5	SYAHARUDD IN	2741739641 200042	SD N 022 RAMBAH SAMO	D2	53_0 8	32_ 03	IV/A
6	ROSMERI SARAGI	9936739641 300062	SMP ADVENT UJUNGBATU	S1	53_0 6	06_ 05	
7	SARIATI.Y	9442740643 300022	SD N 024 RAMBAH SAMO	S1	52_1 1	13_ 04	III/A
8	MUSLIM. N	3559740642 200042	SD N 010 RAMBAH HILIR	S1	52_1 0	31_ 03	IV/A
9	ERMA	9554740644 300003	SD N 001 ROKAN IV KOTO	S1	52_0 0	30_ 03	IV/A
10	ROBERT SITUMORAN G	2555740646 200003	SD S YAYASAN TIGA PUTRA PERSADA PT. HUTAHAEAN	SM A	52_0 0	13_ 04	
11	ERLINA KASMA	1433741643 300252	SD N 003 RAMBAH	SM A	52_0 0	29_ 00	IV/A
12	M. SALEH	6637741644 200042	SMA N 1 TAMBUSAI	S1	51_0 9	08_ 01	NON PNS
13	SRI APRILLA	9745741644 300022	SMK NEGERI 1 TANDUN	S1	51_0 8	08_ 06	
14	HAMZAH	6847741643 200082	SD N 012 KABUN	S1	51_0 7	30_ 04	IV/A
15	NURIDA SARAGI, S.Pd	6962741643 300052	SMP N 2 UJUNGBATU	S1	51_0 6	08_ 03	NON PNS
16	AFRIZAL	1041741646 200013	SMP N 2 KEPENUHAN	D3	51_0 3	19_ 01	IV/A
17	DAHHERMAN	1554741642 200033	SD N 003 ROKAN IV KOTO	S1	51_0 0	13_ 04	II/B
18	NURLELA SARI	3433741642 300322	SD N 020 RAMBAH SAMO	S1	51_0 0	28_ 07	IV/A
19	MUZAKIR	8433742644 200272	SD N 010 ROKAN IV KOTO	S1	51_0 0	26_ 09	IV/A
20	SYAFRIZAL	6539742644 200072	SD N 014 TAMBUSAI UTARA	S1	50_1 0	26_ 09	IV/A

Proses Pengklasifikasian *Data Mining* Menggunakan C4.5

Setelah data calon sertifikasi guru dikumpulkan, tahap selanjutnya adalah dilakukan pengolahan data terhadap data yang telah terkumpul tersebut. Proses pengolahan *Data Mining* yang dilakukan dalam penelitian ini terdiri dari beberapa tahapan diantaranya adalah *data selection*, *data cleaning*, *data transformation*, dan pengolahan *Data Mining* dengan menggunakan algoritma C4.5.

Data Cleaning

Data-data yang tidak lengkap (*incomplete*) dan tidak konsisten (*inconsisten*) serta data-data yang salah dihilangkan terlebih dahulu agar data menjadi *valid*, *complete*, dan konsisten, karena data harus bersih dari kesalahan, tidak lengkap dan tidak konsisten. Karena data di atas sudah konsisten, maka penulis tidak melakukan proses *data cleaning*.

Data Selection

Data yang terkumpul pada tabel 2 terdapat beberapa atribut yang tidak terpakai dalam proses pengolahan data. Atribut yang tidak memiliki kontribusi terhadap tujuan penelitian, tidak digunakan dalam penelitian ini. Tabel 2 merupakan hasil dari proses *data selection* terhadap data anggota sebagaimana terlihat pada tabel 2.

Tabel 2 Data Anggota Setelah Dilakukan *Data Selection*

No	Nama	Pend.	Usia	MK	Gol.
1	ABDUL MANAP	SMA	55_08	32_03	III/C
2	BONTOR MANIK	D2	54_10	22_04	IV/A
3	NIZAR.SS	S1	54_08	06_09	III/D
4	ASRUN SAN	S1	54_06	27_06	
5	SYAHARUDDIN	D2	53_08	32_03	IV/A
6	ROSMERI SARAGI	S1	53_06	06_05	
7	SARIATI.Y	S1	52_11	13_04	III/A
8	MUSLIM. N	S1	52_10	31_03	IV/A
9	ERMA	S1	52_00	30_03	IV/A
10	ROBERT SITUMORANG	SMA	52_00	13_04	
11	ERLINA KASMA	SMA	52_00	29_00	IV/A
12	M. SALEH	S1	51_09	08_01	NON PNS
13	SRI APRILLA	S1	51_08	08_06	
14	HAMZAH	S1	51_07	30_04	IV/A
15	NURIDA SARAGI, S.Pd	S1	51_06	08_03	NON PNS
16	AFRIZAL	D3	51_03	19_01	IV/A
17	DAHERMAN	S1	51_00	13_04	II/B

18	NURLELA SARI	S1	51_00	28_07	IV/A
19	MUZAKIR	S1	51_00	26_09	IV/A
20	SYAFRIZAL	S1	50_10	26_09	IV/A

Data Transformation

Langkah selanjutnya adalah melakukan proses transformasi dengan cara mengubah pengetahuan dari data *real* yang terdapat pada tabel 2. Transformasi data dilakukan pada atribut pendidikan, masa kerja, golongan dan usia dengan mengubah data yang berbentuk angka ke dalam data bentuk huruf berdasarkan kriteria-kriteria yang telah ditentukan. Kriteria yang telah ditetapkan pada ketentuan *calon yang mendapatkan sertifikasi guru* adalah sebagai berikut :

1. Tingkat Pendidikan

Mengelompokan pendidikan berdasarkan data pendidikan seperti terlihat pada tabel 3 berikut :

Tabel 3 Klasifikasi Tingkat Pendidikan

Tingkat Pendidikan	Kriteria
SMA	Rendah
D2-D3	Menengah
S1-S2	Tinggi

2. Masa Kerja

Mengelompokan Lamanya Masa Kerja berdasarkan pada data Masa kerja Calon sertifikasi guru seperti terlihat pada tabel 4.

Tabel 4 Klasifikasi Lamanya Masa Kerja

Lama Masa Kerja (thn)	Kriteria
< 8	Rendah
8-20	Sedang
> 20	Tinggi

3. Golongan

Mengelompokkan Jenis Golongan yang dilihat dari golongan data calon sertifikasi guru yang dapat dilihat pada tabel 5.

Tabel 5 Kriteria Jenis Golongan Jabatan

Jenis Golongan	Kriteria
Gol I A-D	Rendah
Gol II A- D	Sedang
Gol III & IV A-D	Tinggi

4. Usia

Mengelompokkan Kriteria usia yang dilihat dari data usia calon sertifikasi guru yang dapat dilihat pada tabel 6.

Tabel 6 Kriteria Jenis Usia

Jenis Golongan	Kriteria
< 35	Muda
35-54	Parobaya

>54	Tua
-----	-----

Data anggota pada tabel 4.2 ditransformasi dan ditentukan tingkat potensi calon yang akan mendapatkan sertifikasi guru. Hasil dari proses *data transformation* dapat dilihat pada tabel 7.

7 Tabel Hasil Transformation

Nama	Pend	MK	Gol	Usia	Prediksi
Abdul Manap	rendah	tinggi	tinggi	tua	layak
Bontor Manik	menengah	tinggi	tinggi	parobaya	layak
Nizar.Ss	tinggi	rendah	tinggi	parobaya	dipertimbangkan
Asrun San	tinggi	tinggi	tinggi	parobaya	layak
Syharuddin	menengah	tinggi	tinggi	parobaya	layak
Rosmeri Saragi	tinggi	rendah	sedang	parobaya	dipertimbangkan
Sariati.Y	tinggi	tinggi	tinggi	parobaya	layak
Muslim. N	tinggi	tinggi	tinggi	parobaya	layak
Erma	tinggi	tinggi	tinggi	parobaya	layak
Robert Situmorang	rendah	sedang	sedang	parobaya	layak
Erlina Kasma	rendah	tinggi	tinggi	parobaya	layak
M. Saleh	tinggi	sedang	tinggi	parobaya	layak
Sri Aprilla	tinggi	sedang	tinggi	parobaya	layak
Hamzah	tinggi	tinggi	tinggi	parobaya	layak
Nurida Saragi, S.Pd	tinggi	tinggi	tinggi	parobaya	layak
Afrizal	menengah	sedang	tinggi	parobaya	layak
Daherman	tinggi	sedang	sedang	parobaya	dipertimbangkan
Nurlela Sari	tinggi	tinggi	tinggi	parobaya	layak
Muzakir	tinggi	tinggi	tinggi	parobaya	layak
Syafrizal	tinggi	tinggi	tinggi	parobaya	layak
Edison Panggabean	tinggi	sedang	tinggi	parobaya	layak
Sondang Mariana	tinggi	sedang	tinggi	parobaya	layak

3.1. Data Mining dengan Menggunakan Algoritma C.45

Setelah dilakukan proses *data transformation* tahapan selanjutnya adalah melakukan proses pengolahan *Data Mining* dengan menggunakan algoritma C4.5. Pengolahan *Data Mining* dengan menggunakan algoritma C4.5 dilakukan dengan perhitungan *Entropy* dan *Gain* untuk menentukan akar (*root*) dari pohon keputusan dalam menentukan klasifikasi anggota berdasarkan tingkat potensi anggota dalam memberikan kontribusi pada perusahaan.

Langkah-langkah pengolahan *Data Mining* dengan menggunakan algoritma C4.5 adalah sebagai berikut:

- a. Pilih atribut sebagai akar.

Data dikelompokkan terlebih dahulu dengan menghitung jumlah kasus, jumlah kasus untuk anggota yang berpotensi tinggi, sedang dan rendah. Selanjutnya dihitung nilai *Entropy* dan *Gain* dari semua atribut. Untuk menentukan akar ditentukan dari nilai *Gain* tertinggi. Hasil pengelompokkan data anggota dapat dilihat pada tabel 8 berikut :

Tabel 8 Pengelompokkan Data Anggota

No de			Jumlah Kasus	Potensi Anggota	
				Layak	Dipertimbangkan
1	Total		150	47	103
	Pendidikan				
		Rendah	3	3	0
		Menengah	12	5	7
		Tinggi	135	39	96
	Masa Kerja				
		Rendah	35	0	35
		Sedang	92	24	68
		Tinggi	23	23	0
	Golongan				
		Rendah	16	0	16
		Sedang	85	4	81
		Tinggi	49	43	6
	Usia				
		Muda	1	0	1
		Parobaya	145	43	102
		Tua	4	4	0

- b. Buat cabang untuk tiap-tiap nilai
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

3.6 Perhitungan Node 1

Setelah data kita kelompokkan seperti pada tabel 8, langkah selanjutnya adalah mencari nilai *Entropy* total dan juga nilai semua atribut yang digunakan. Kemudian dihitung *Gain* dari masing-masing atribut yang digunakan.

Berikut perhitungan nilai *Entropy* dan *Gain* pada *node* 1:

- a. Nilai *Entropy* Total

Nilai *Entropy* total merupakan nilai *Entropy* yang mewakili dari seluruh jumlah total variabel atribut yang ada. Total kasus pada data sampel dalam penelitian ini adalah berjumlah 150 buah data. Menghitung nilai *Entropy* total dilakukan berdasarkan nilai total potensi anggota layak (47), dipertimbangkan (103), sedangkan jumlah kasus keseluruhan adalah sebanyak 150. Perhitungan nilai *Entropy* total adalah sebagai berikut :

$$\begin{aligned}
 & \text{Entropy (total)} \\
 & = \left(-\frac{47}{150} * \log_2 \left(\frac{47}{150} \right) \right) + \left(-\frac{103}{150} * \log_2 \left(\frac{103}{150} \right) \right) \\
 & = 0.896986
 \end{aligned}$$

- b. Nilai *Entropy* Pendidikan

Kriteria Pendidikan terdiri dari 3 nilai yaitu rendah, sedang dan, tinggi. Perhitungan nilai *Entropy* pada atribut pendidikan adalah sebagai berikut :

Entropy (rendah)

$$= \left(-\frac{3}{3} * \log_2\left(\frac{3}{3}\right)\right) + \left(-\frac{0}{3} * \log_2\left(\frac{0}{3}\right)\right)$$

$$= 0$$

Entropy (Menengah)

$$= \left(-\frac{5}{12} * \log_2\left(\frac{5}{12}\right)\right) + \left(-\frac{7}{12} * \log_2\left(\frac{7}{12}\right)\right)$$

$$= 0.979868757$$

Entropy (Tinggi)

$$= \left(-\frac{39}{135} * \log_2\left(\frac{39}{135}\right)\right) + \left(-\frac{96}{135} * \log_2\left(\frac{96}{135}\right)\right)$$

$$= 0,967281622$$

c. Nilai *Entropy* Masa Kerja

Atribut yang terdapat pada Masa Kerja yakni rendah (35), sedang (92) dan , tinggi (23). Berikut adalah perhitungan nilai *Entropy* Masa Kerja :

Entropy (rendah)

$$= \left(-\frac{0}{35} * \log_2\left(\frac{0}{35}\right)\right) + \left(-\frac{35}{35} * \log_2\left(\frac{35}{35}\right)\right) = 0$$

Entropy (sedang)

$$= \left(-\frac{24}{92} * \log_2\left(\frac{24}{92}\right)\right) + \left(-\frac{68}{92} * \log_2\left(\frac{68}{92}\right)\right)$$

$$= 0.82806$$

Entropy (tinggi)

$$= \left(-\frac{23}{23} * \log_2\left(\frac{23}{23}\right)\right) + \left(-\frac{0}{23} * \log_2\left(\frac{0}{23}\right)\right) = 0$$

d. Nilai *Entropy* Golongan

Nilai *Entropy* Golongan terdiri atas 3 nilai yaitu Rendah (16), sedang (85), dan tinggi (49). Perhitungan nilai *Entropy* Golongan adalah sebagai berikut :

Entropy (rendah)

$$= \left(-\frac{0}{16} * \log_2\left(\frac{0}{16}\right)\right) + \left(-\frac{16}{16} * \log_2\left(\frac{16}{16}\right)\right) = 0$$

Entropy (sedang)

$$= \left(-\frac{4}{85} * \log_2\left(\frac{4}{85}\right)\right) + \left(-\frac{81}{85} * \log_2\left(\frac{81}{85}\right)\right)$$

$$= 0.27377$$

Entropy (tinggi)

$$= \left(-\frac{43}{49} * \log_2\left(\frac{43}{49}\right)\right) + \left(-\frac{6}{49} * \log_2\left(\frac{6}{49}\right)\right)$$

$$= 0.53636$$

e. Nilai *Entropy* Usia

Nilai *Entropy* Usia terdiri atas 3 nilai yaitu Muda (1), Parobaya (145), dan Tua (4). Perhitungan nilai *Entropy* Golongan adalah sebagai berikut :

Entropy (Muda)

$$= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right)$$

$$= 0$$

Entropy (Parobaya)

$$= \left(-\frac{43}{145} * \log_2\left(\frac{43}{145}\right)\right) + \left(-\frac{102}{145} * \log_2\left(\frac{102}{145}\right)\right)$$

$$= 0.87703$$

Entropy (tua)

$$= \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right)$$

$$= 0$$

f. Nilai *Gain*

Nilai *Gain* diambil dari hasil *Entropy* total dikurang jumlah hasil kali dari $\frac{S_i}{S}$ dikali *Entropy* dari masing-masing kasus. Pada kasus ini ada tiga atribut yang akan kita cari nilai *Gain*-nya yaitu, *Gain* Pendidikan, *Gain* Masa Kerja, *Gain* Golongan dan *Gain* Usia. Perhitungan masing-masing nilai *Gain* dari 4 atribut tersebut adalah sebagai berikut:

Gain (Total,Pendidikan)

$$= 0.896986 - \left(\left(\frac{3}{150} \times 0\right) + \left(\frac{12}{150} \times 0.979308757\right) + \left(\frac{135}{150} \times 0.967231022\right)\right)$$

$$= 0.05196$$

Gain (Total,Masa Kerja)

$$= 0.896986 - \left(\left(\frac{35}{65} \times 0\right) + \left(\frac{92}{65} \times 0.82806\right) + \left(\frac{23}{65} \times 0\right)\right)$$

$$= 0.389109$$

Gain (Total,Golongan)

$$= 0.896986 - \left(\left(\frac{16}{150} \times 0\right) + \left(\frac{85}{150} \times 0.27377\right) + \left(\frac{49}{150} \times 0.53636\right)\right)$$

$$= 0.566639$$

Gain (Total,Usia)

$$= 0.896986 - \left(\left(\frac{1}{150} \times 0\right) + \left(\frac{145}{150} \times 0.87703\right) + \left(\frac{4}{150} \times 0\right)\right)$$

$$= 0.04919033$$

Hasil perhitungan nilai *Entropy* dan *Gain node* 1 dapat dilihat pada tabel 9.

Tabel 9 Tabel Hasil Perhitungan *Node* 1

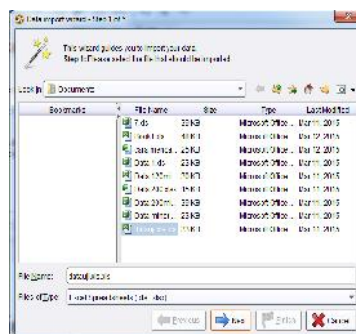
Node		Jlh Kasus	Potensi Anggota		Entropy	Gain
			L	D		
1	Total	150	47	103	0.896986	
	Pendidikan					0.05196
	Rendah	3	3	0	0	
	Menengah	12	5	7	0.979868757	
	Tinggi	135	39	96	0.967281622	
	Masa Kerja					0.389109
	Rendah	35	0	35	0	
	Sedang	92	24	68	0.82806	
	Tinggi	23	23	0	0	
	Golongan					0.566639
	Rendah	16	0	16	0	
	Sedang	85	4	81	0.27377	
	Tinggi	49	43	6	0.53636	
	Usia					0.04919033
	Muda	1	0	1	0	
	Parobaya	145	43	102	0.87703	
	Tua	4	4	0	0	

Dari hasil perhitungan nilai *Entropy* dan *Gain* seperti terlihat pada tabel 4.9 dapat diketahui bahwa atribut yang memiliki *Gain* tertinggi adalah atribut Golongan yaitu sebesar 0.566639. Dengan demikian, atribut Golongan adalah atribut yang menjadi *node* akar. Ada 3 nilai yang terdapat pada atribut Golongan yaitu rendah, sedang, dan tinggi. Dari ketiga atribut, nilai atribut Rendah adalah 1, yaitu potensi anggota adalah dipertimbangkan, sehingga tidak perlu perhitungan lanjut, tetapi untuk nilai atribut sedang dan tinggi masih perlu dilakukan perhitungan lagi.

Pengujian Dan Implementasi

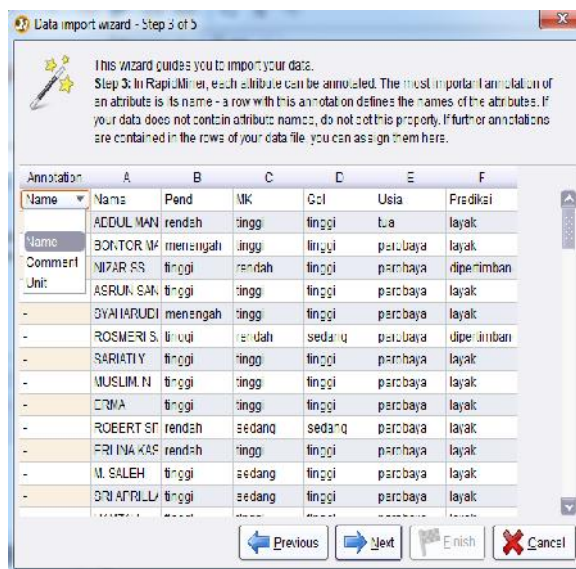
Pada tahap pengujian dilakukan pengolahan *Data Mining* menggunakan aplikasi RapidMiner berdasarkan data anggota sebanyak 150 buah data. Berikut terlihat dataset yang terdiri 150 record data dan berisi variabel-variabel atribut kondisi dan atribut keputusan untuk menganalisa Data calon Peserta Sertifikasi Guru disimpan dalam aplikasi *microsoft excel* dengan nama file datauji.xls

1. Pilih Menu *File*, pilih *Import Data* dan pilih *Import Excel Sheet*, sehingga akan muncul tampilan seperti berikut :



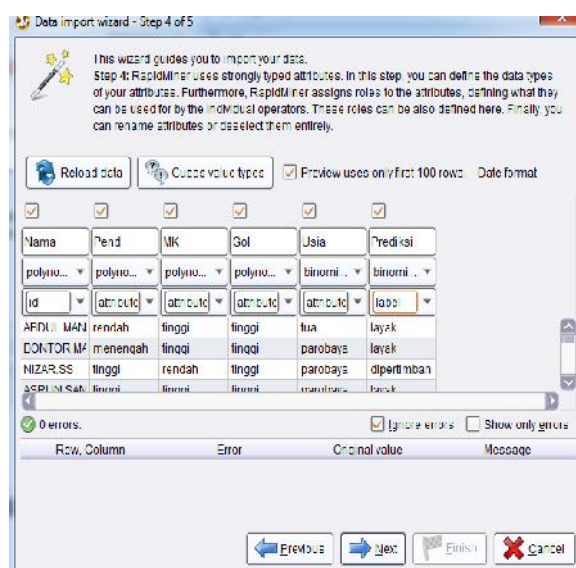
Gambar 3 Tampilan *Import Excel Sheet*

2. Arahkan ke folder tempat file datauji.xls disimpan, kemudian klik pada file datauji.xls kemudian klik tombol *Next* dan klik *Next* sekali lagi.
3. Klik pada baris pertama di kolom Annotation, pilih name dan klik *Next* seperti tampilan berikut :



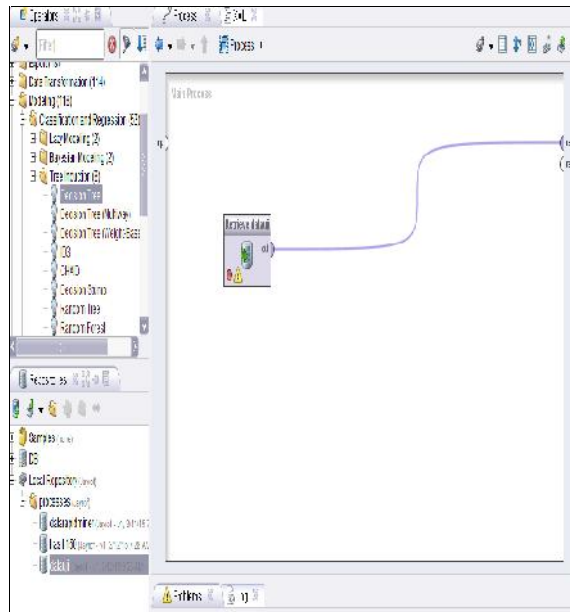
Gambar 4 Pilihan Annotation

4. Selanjutnya ubah properties atribut norek menjadi *id* dan atribut Status menjadi *label*, dan sesuaikan tipe tiap field atribut lalu klik *Next* seperti tampilan berikut :



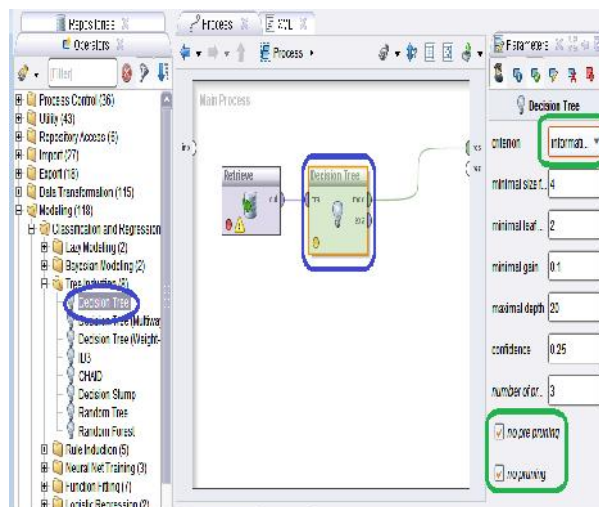
Gambar 5 Pilihan Properties atribut dan field

5. Ketikkan DataUji di *textbox* Name, kemudian klik tombol *Finish* untuk menyimpan *dataset* yang akan kita pakai.
6. Untuk menghasilkan pohon keputusan dengan algoritma C4.5, pilih *dataset* yang kita sudah simpan sebelumnya di *Tab Repositories*, kemudian *drag dataset* datauji ke *Tab Process*, seperti tampilan berikut :



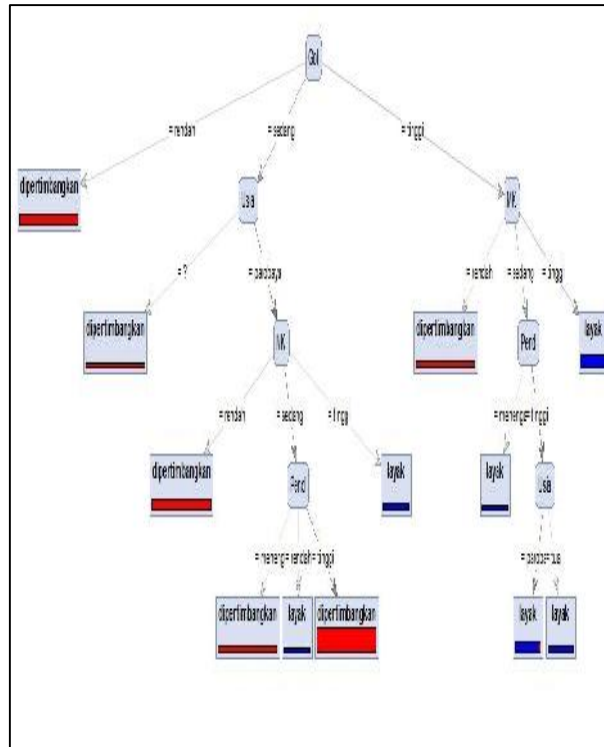
Gambar 6. Tampilan Repositories

- Selanjutnya tambahkan algoritma decision tree dari Tab Operators, pilih Modelling(118), pilih Classification and Regression (53), pilih Tree Induction(8) , kemudian drag Decision Tree ke Tab Process, pada Tab Parameters pilih information_gain pada criterion, dan cek pada pilihan no pre pruning dan no pruning.



Gambar 7. Tampilan Process Decision Tree

- Pilih menu Process dan pilih Run, pilih No dan pilih Yes, secara otomatis Rapidminer akan memproses untuk menghasilkan pohon keputusan seperti gambar dibawah :



Gambar 8. Pohon Keputusan Hasil Process Decision Tree

KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat diambil kesimpulan sebagai berikut : 1.) Metode pohon keputusan menggunakan Algoritma C4.5 bekerja sangat baik dalam membangun pohon keputusan dan proses ekstraksi rule atau knowledge dari data sample calon peserta sertifikasi guru. 2). Penentuan data training sangat menentukan tingkat akurasi tree yang dibuat. 3). Berdasarkan prosentase kebenaran tree sangat dipengaruhi oleh data training yang digunakan untuk membangun model tree tersebut. 4). Melalui pohon keputusan diperoleh 11 aturan-aturan (rule) dalam menentukan pola sertifikasi guru. 5). Atribut yang menjadi akar (root) dalam penelitian ini adalah Gol (Golongan). Hal ini menyatakan bahwa atribut golongan guru dalam menentukan pola merupakan atribut yang paling menentukan dalam penentuan klasifikasi calon sertifikasi guru. Saran yang dapat diberikan untuk pengembangan lebih lanjut adalah sebagai berikut : pertama Pada penelitian ini, penulis hanya mencoba salah satu teknik yang digunakan untuk membangun pohon keputusan dalam menentukan calon sertifikasi guru yaitu teknik klasifikasi decision tree menggunakan Algoritma C4.5. Untuk mendapatkan hasil prediksi yang baik dapat digunakan beberapa atau penggabungan beberapa teknik prediksi seperti Penerapan metode pincer search dan Paradigma apriori digunakan untuk mencari large itemset dalam penetapan association rule kedua Untuk mengimplementasikan rule yang dihasilkan dari teknik klasifikasi decision tree tersebut perlu dilakukan konsultasi dengan tim penentuan calon sertifikasi guru pada Dinas Pendidikan Kabupaten Rokan Hulu, sehingga dapat dipilih rule-rule yang terbaik dalam menentukan calon sertifikasi guru.

DAFTAR PUSTAKA

Angga Ginanjar Mabur dan Riani Lubis (2012). Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit, Edisi 1 Volume. 1 Maret 2012.
 Budanis Dwi Meilani Achmad dan Fauzi Slamet (2012). Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree, Iptek Vol 16 No.1 Mei 2012.

- Jonh Fredrik Ulysses (2013). **Data Mining Clasificasion Untuk Prediksi Lama Masa Studi Mahasiswa Berdasarkan Jalur Penerimaan Dengan Metode Naïve Bayes.**
- Khafis Hastuti (2012). **Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Mahasiswa Non Aktif.** ISBN 979-26-0255-0
- Liliana Swastina (2013). **Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa.** Gema Aktualita, Vol. 2 No. 1, Juni 2013.
- Mohammed M. Abu Tair dan Alaa M. El-Halees (2012). **Mining Educational Data to Improve Students' Performance: A Case Study.** ISSN 2223-4985 Volume 2. No.2, February 2012.
- Nuqson Masykur Huda (2010). **Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa,**
- Rizal Amegia Saputra (2014). **Komprasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (TB),** SNIT 2014.
- Sandro Romario. S (2-013). **Penerapan Data Mining Pada RSUP DR. Moh Hoesin Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin,** TUV NORD 2013.
- Strisno¹, Afriyudi², Widiyanto³ (2013). **Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering : Studi Kasus PT. Indomarco Palembang,** Vol.x No.x, 4 November 2013: 1-11