

**PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI TINGKAT KOMPETENSI KARYAWAN PT
MULTISTRADA ARAH SARANA**

Wahyono, Agung Nugroho

Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Pelita Bangsa

Jl. Inspeksi Kalimalang Tegal Arah Deltamas, Bekasi

Telp. (021) 28518181, Fax (021) 28518180

Email: wahyono4628@gmail.com, agung@pelitabangsa.ac.id

Abstrak

The success of an organization or company is influenced by the competence of human resources in it in carrying out its duties. Assessment of competence in the company is very important because it is necessary to determine the level of competence of employees. So far, the competency level assessment is still done manually by providing certain criteria by the Personnel Department (HRD). The number of criteria and employees employed is quite difficult in the assessment of employee competency. In addition, the quality of employee work is also often changing. This is very influential on the future prospects of the company. The purpose of this study is to analyze and know the pattern of employee competency level. The research method used is Decision Tree Learning (DCT). The C4.5 algorithm is part of DCT. Algorithm C4.5, the data mining technique to create a decision tree that represents rules or rules. Of the 257 overall employee data is divided into two partitions, 80% for training data and 20% data testing. The result of application of algorithm C4.5 got its accuracy 80,39% with calculation of RapidMiner 5.3. This research produces a pattern of competency level that can be used by PT Multistrada Arah Sarana for employee monitoring process to determine whether or not to need training or promotion promotion.

Keywords: Competence, Data mining, Algorithm C4.5, RapidMiner

1. Pendahuluan

Salah satu elemen dalam perusahaan yang sangat penting adalah Sumber Daya Manusia (SDM). Pengelolaan SDM dari suatu perusahaan sangat mempengaruhi banyak aspek penentu keberhasilan kerja dari perusahaan tersebut. Jika SDM dapat diorganisir dengan baik, maka diharapkan perusahaan dapat menjalankan semua proses usahanya dengan baik.

Monitoring dan evaluasi yang tidak tepat sasaran terhadap penilaian karyawan akan menimbulkan masalah dalam pengembangan sumber daya manusia di dalam perusahaan. Dalam melakukan proses penilaian kompetensi karyawan, banyak sekali kriteria yang telah ditentukan oleh perusahaan. Masing-masing perusahaan memiliki kriteri-kriteria saat melakukan penilaian kompetensinya. Banyaknya kriteria inilah yang menyulitkan pihak manajemen untuk memberi bobot pada setiap kriteria, selain itu proses penilaian yang masih manual dan standar operasional prosedur yang belum maksimal membuat penilaian semakin sulit dan tidak tepat sasaran. Hal – hal tersebut mengakibatkan tidak mudahnya bagian personalia untuk mengambil keputusan karyawan mana saja yang memang benar-benar memenuhi standar kompetensi yang telah ditentukan.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dan berbagai database besar^[1]. Menurut Gartner Group *data mining* adalah suatu proses

menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika^[2].

Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik *Decision tree* (pohon keputusan) yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihannya adalah dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, dapat menghasilkan aturan aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain^[3].

Decision tree adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain^[4].

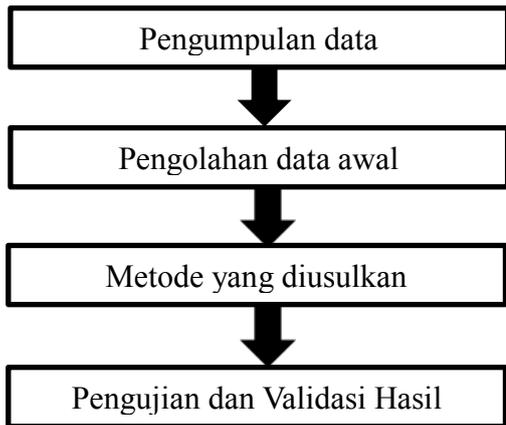
Banyak algoritma yang dapat dipakai dalam pembentukan *decision tree*, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dan algoritma ID3. Data dalam *decision tree* biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut cuaca mempunyai *instance* berupa cerah,

berawan, dan hujan.

Proses pada *decision tree* adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*^[5].

2. Metode Penelitian

Pada penelitian ini, data yang digunakan adalah data nilai hasil tes general dan teknikal Karyawan di PT. Multistrada Arah Sarana. Data nilai tersebut akan diolah menggunakan beberapa metode *data mining* sehingga diperoleh satu metode yang paling akurat dan dapat digunakan sebagai *rules* dalam memprediksi kompetensi karyawan. Berikut tahapan dalam melakukan penelitian *data mining*.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data dari hasil tes kompetensi karyawan PT Multistrada Arah Sarana yang dilakukan pada tahun 2016. Sumber data berupa file excel dengan jumlah populasi sebanyak 257 data karyawan. Data yang di dapat kemudian dibagi menjadi dua untuk data *training* (80%) dan data *testing* (20%).

2.2 Pengolahan data awal

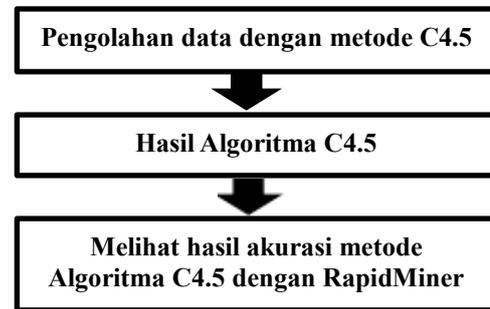
Pada data karyawan yang telah dikumpulkan dilakukan *cleaning* data, yaitu dengan menghilangkan record-record yang noise atau tidak lengkap dan record yang berulang, atribut yang tidak diperlukan juga bisa dihilangkan seperti atribut no, nama, nik. Hal ini dilakukan karena atribut tersebut tidak berpengaruh terhadap pengolahan data pada proses selanjutnya. Berikut adalah atribut data yang digunakan:

- A1 = Mengetahui dasar-dasar pengetahuan ban
- A2 = Mengetahui proses manufaktur ban
- A3 = Mengetahui secara umum ISO/TS 16949:2009 dan control dokumen
- A4 = Memahami konsep dasar perbaikan berkelanjutan
- A5 = Memahami dasar K3 (Kesehatan dan Keselamatan Kerja)
- B1 = Memahami cara membaca spec
- B2 = Memahami Pengisian Checksheet Operator TBM
- B3 = Memahami sistem FIFO

- B4 = Memahami Punching Sistem
- B5 = Mampu Melakukan Proses Assembly sesuai standard
- B6 = Memahami Cara Ganti Size
- B7 = Memahami UPH & Cycle Time
- B8 = Memahami Barcode Sistem
- B9 = Memahami Cara Venting GT
- B10 = Memahami Macam-Macam Defect
- B11 = Memahami Offset Parameter Mesin
- B12 = Mengerti Pemilahan Recycle

2.3 Metode yang diusulkan

Metode yang akan digunakan dalam penelitian ini adalah metode *Algoritma C4.5*. Metode *Algoritma C4.5* dipilih karena salah satu kelebihan adalah dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, dapat menghasilkan aturan aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Data dihitung dengan menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasi. Dalam tahapan ini akan dilakukan beberapa langkah pengujian data yaitu seperti berikut:



Gambar 2. Langkah Pengujian Metode

Data kompetensi karyawan akan diolah dengan metode *Algoritma C4.5*, kemudian dilihat akurasi pada metode tersebut. Hasil pengujian dengan akurasi yang tinggi adalah metode yang akan digunakan untuk prediksi ahli atau tidaknya skill karyawan.

2.4 Pengujian dan Validasi Hasil

a) Data Cleaning

Langkah awal dalam tahapan data mining adalah tahap data *cleaning*. Tahap ini merupakan suatu proses untuk pembersihan dari data yang kotor. Maksud data yang kotor adalah data yang tidak memiliki nilai (*missing value*), tidak konsisten, tidak relevan atau terdapat kesalahan penulisan saat penginputan. Pembersihan data juga akan mempengaruhi kemampuan dari teknik data mining, karena data yang di tangani akan berkurang jumlah dan kompleksitasnya.

b) Data Selection

Data selection merupakan tahap pemilihan data yang relevan yang berada pada tabel, tidak semua data dipakai, hanya data yang sesuai untuk dianalisis yang akan diambil dari tabel. Pada penelitian ini setelah tahap

data cleansing dan data integration, dilakukan penyeleksian atribut untuk memilih data yang relevan sesuai dengan analisis yang ingin dicapai.

c) Data Transformation

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

d) Proses Mining

Proses *Mining* merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan tersembunyi dan berharga dari data. Metode yang digunakan adalah klasifikasi dengan algoritma C4.5. Berikut adalah penerapan perhitungan algoritma *naïve bayes* secara manual dan memakai tool *Rapidminer*.

e) Evaluasi

Dalam tahap ini dilakukan evaluasi dan menafsirkan pola yang didapatkan dari hasil algoritma yang dipakai untuk mengetahui *accuracy, Precision, Recall* dan lain-lain.

3. Hasil dan Pembahasan

3.1 Pengolahan Data awal

Hasil dari proses *data cleaning* dan *data selection* kemudian data diubah kebentuk yang lebih sederhana sesuai dengan kebutuhan proses selanjutnya.

	B5	B6	B7	B8	B9	B10	B11	B12	Keputusan
1	3	4	3	4	3	4	3	4	Kompeten
2	3	4	3	4	3	4	3	4	Kompeten
3	3	4	3	4	3	4	3	4	Kompeten
4	3	4	3	4	3	4	3	4	Kompeten
5	3	4	3	4	3	4	3	4	Kompeten
6	3	4	3	4	3	4	3	4	Kompeten
7	3	4	3	4	3	4	3	4	Kompeten
8	3	4	3	4	3	4	3	4	Kompeten
9	3	4	3	4	3	4	3	4	Kompeten
10	3	4	3	4	3	4	3	4	Kompeten
11	3	4	3	4	3	4	3	4	Kompeten
12	3	4	3	4	3	4	3	4	Kompeten
13	3	4	3	4	3	4	3	4	Kompeten
14	3	4	3	4	3	4	3	4	Kompeten
15	3	4	3	4	3	4	3	4	Kompeten
16	3	4	3	4	3	4	3	4	Kompeten
17	3	4	3	4	3	4	3	4	Kompeten
18	3	4	3	4	3	4	3	4	Kompeten
19	3	4	3	4	3	4	3	4	Kompeten
20	3	4	3	4	3	4	3	4	Kompeten
21	3	4	3	4	3	4	3	4	Kompeten
22	3	4	3	4	3	4	3	4	Kompeten
23	3	4	3	4	3	4	3	4	Kompeten
24	3	4	3	4	3	4	3	4	Kompeten
25	3	4	3	4	3	4	3	4	Kompeten
26	3	4	3	4	3	4	3	4	Kompeten
27	3	4	3	4	3	4	3	4	Kompeten
28	3	4	3	4	3	4	3	4	Kompeten
29	3	4	3	4	3	4	3	4	Kompeten
30	3	4	3	4	3	4	3	4	Kompeten
31	3	4	3	4	3	4	3	4	Kompeten
32	3	4	3	4	3	4	3	4	Kompeten
33	3	4	3	4	3	4	3	4	Kompeten
34	3	4	3	4	3	4	3	4	Kompeten
35	3	4	3	4	3	4	3	4	Kompeten
36	3	4	3	4	3	4	3	4	Kompeten
37	3	4	3	4	3	4	3	4	Kompeten
38	3	4	3	4	3	4	3	4	Kompeten
39	3	4	3	4	3	4	3	4	Kompeten
40	3	4	3	4	3	4	3	4	Kompeten
41	3	4	3	4	3	4	3	4	Kompeten
42	3	4	3	4	3	4	3	4	Kompeten
43	3	4	3	4	3	4	3	4	Kompeten
44	3	4	3	4	3	4	3	4	Kompeten
45	3	4	3	4	3	4	3	4	Kompeten
46	3	4	3	4	3	4	3	4	Kompeten
47	3	4	3	4	3	4	3	4	Kompeten
48	3	4	3	4	3	4	3	4	Kompeten
49	3	4	3	4	3	4	3	4	Kompeten
50	3	4	3	4	3	4	3	4	Kompeten
51	3	4	3	4	3	4	3	4	Kompeten
52	3	4	3	4	3	4	3	4	Kompeten
53	3	4	3	4	3	4	3	4	Kompeten
54	3	4	3	4	3	4	3	4	Kompeten
55	3	4	3	4	3	4	3	4	Kompeten
56	3	4	3	4	3	4	3	4	Kompeten
57	3	4	3	4	3	4	3	4	Kompeten
58	3	4	3	4	3	4	3	4	Kompeten
59	3	4	3	4	3	4	3	4	Kompeten
60	3	4	3	4	3	4	3	4	Kompeten
61	3	4	3	4	3	4	3	4	Kompeten
62	3	4	3	4	3	4	3	4	Kompeten
63	3	4	3	4	3	4	3	4	Kompeten
64	3	4	3	4	3	4	3	4	Kompeten
65	3	4	3	4	3	4	3	4	Kompeten
66	3	4	3	4	3	4	3	4	Kompeten
67	3	4	3	4	3	4	3	4	Kompeten
68	3	4	3	4	3	4	3	4	Kompeten
69	3	4	3	4	3	4	3	4	Kompeten
70	3	4	3	4	3	4	3	4	Kompeten
71	3	4	3	4	3	4	3	4	Kompeten
72	3	4	3	4	3	4	3	4	Kompeten
73	3	4	3	4	3	4	3	4	Kompeten
74	3	4	3	4	3	4	3	4	Kompeten
75	3	4	3	4	3	4	3	4	Kompeten
76	3	4	3	4	3	4	3	4	Kompeten
77	3	4	3	4	3	4	3	4	Kompeten
78	3	4	3	4	3	4	3	4	Kompeten
79	3	4	3	4	3	4	3	4	Kompeten
80	3	4	3	4	3	4	3	4	Kompeten

Gambar 3. Hasil proses *Data Transformation*

3.2 Proses Mining

Pada tahap ini metode *data mining* diterapkan untuk menemukan pengetahuan tersembunyi dan berharga dari data. Metode yang digunakan adalah klasifikasi dengan algoritma C4.5. Berikut adalah penerapan algoritma C4.5 memakai tool *Rapidminer*.

3.3 Menentukan Data Training dan Data Testing

Data karyawan yang ada dibagi menjadi 2 bagian. Dari seluruh populasi, data dipisah menjadi 80 % dan 20 %. data yang akan di jadikan *data training* berjumlah 80 %. Data akan dipakai untuk membuat model atau *rule, sedangkan* 20% data akan

dipakai sebagai *data testing*.

Jumlah populasi (N) = 257

Jumlah data testing = 20% x 257 = 51

Jumlah sampel (n) = 51

Interval Sampling (k) = N/n = 257/ 51 = 5

Unsur pertama yang diambil untuk data testing (s) = 1

Unsur kedua = s + k

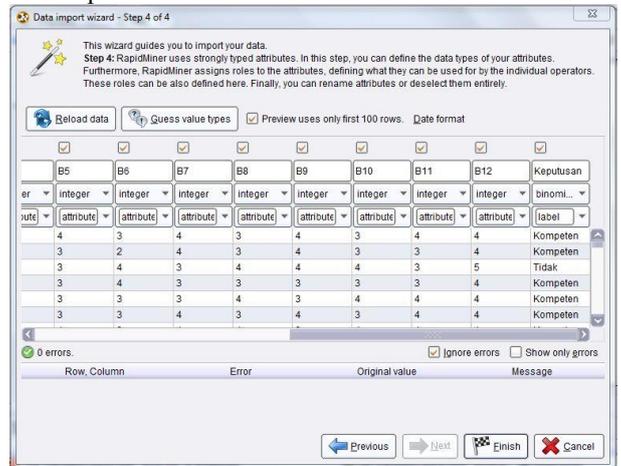
Unsur ketiga = s + 2k

Unsur keempat = s + 3k, dan seterusnya hingga unsur ke-n.

Dari hasil perhitungan diatas, diperoleh data *testing* yang akan di ambil mulai dari urutan data ke 5, 10, 15, 20, dst sampai jumlah data mencapai 51 data karyawan, maka sisanya dijadikan data *training* sebanyak 206 data karyawan.

3.4 Proses Import Data

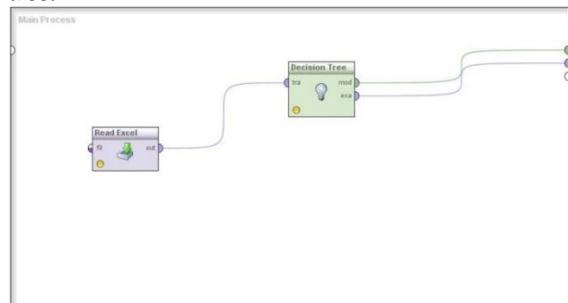
Pada tahapan ini data karyawan yang sudah dipisahkan antara *data training* dan *data testing* di import kedalam tool *Rapidminer 5.3*. Data yang digunakan dalam aplikasi adalah *data training*. Di dalam tahap ini juga dilakukan pemilihan jenis atribut pada data.



Gambar 4. Proses *Import*

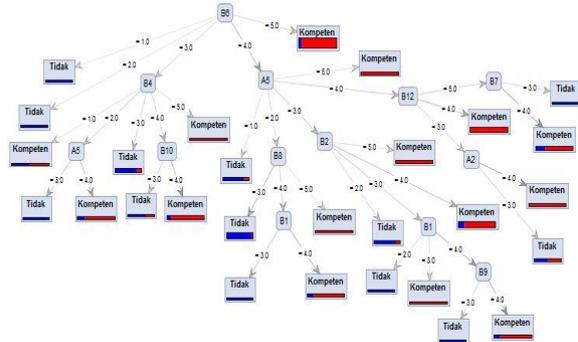
3.5 Proses Training

Proses selanjutnya yaitu melakukan proses pelatihan data pada model C4.5 untuk menghasilkan *decision tree* yang selanjutnya akan digunakan untuk proses berikutnya yaitu proses *testing*. Skema yang dipakai dalam proses *training* adalah seperti gambar berikut. Blok *Data training* di hubungkan dengan operator *decision tree* untuk menghasilkan *decision tree*.



Gambar 5. Proses *Training*

Hasil dari proses di atas didapatkan sebuah *decision tree* dengan pola sebagai berikut:

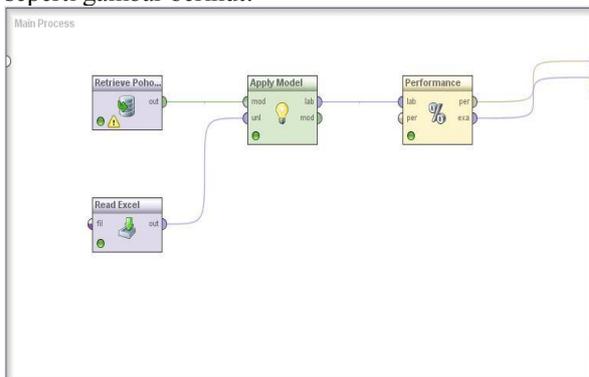


Gambar 6. *Decision tree*

Berdasarkan *Decision Tree* yang telah terbentuk, diperoleh *rule* untuk klasifikasi prediksi kompetensi karyawan sebanyak 28 *rules*, dengan 13 *rules* mengelompokkan sample ke kelas 0 (Tidak Kompeten), dan 15 *rules* yang mengklasifikasikan sampel ke kelas 1 (Kompeten).

3.6 Proses Testing

Decision tree yang telah terbentuk selanjutnya diuji performanya menggunakan *data testing* yang sebelumnya telah disiapkan untuk mengetahui besarnya *accuracy*, *precision*, dan *recall*. Skema yang dibuat pada proses testing adalah seperti gambar berikut.



Gambar 7. Proses *Testing*

Pada proses di atas, blok *decision tree* dioperasikan menggunakan blok *Apply model* dengan mengambil sumber data dari blok *data testing*. Hasil output yang keluar dari *Apply mode* kemudian di *meassure* meggunakan blok *performance* untuk mendapatkan *accuracy*, *precision*, dan *recall*.

3.6.1 Hasil Testing

Setelah melakukan pengujian *data testing* yang berjumlah 51 data menggunakan *Rapidminer* dengan metode *decision tree*, Hasil yang diperoleh adalah sebagai berikut:

Tabel 1. Hasil pengujian *data testing*

accuracy: 80.39%			
	true Kompeten	true Tidak	class precision
pred Kompeten	17	3	85.00%
pred Tidak	7	24	77.42%
class recall	70.83%	88.89%	

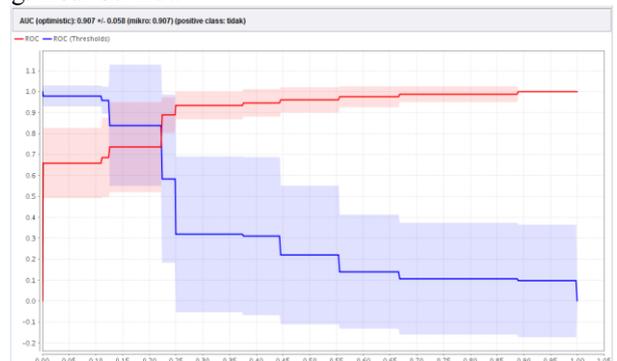
Dari tabel diatas, dapat disimpulkan besarnya Akurasi (*Accuracy*), Presisi (*Precision*), dan *Recall* sebagai berikut:

Tabel 2. *Accuracy*, *Precision*, Dan *Recall*

Parameter	Persentase
Akurasi (<i>Accuracy</i>)	80,39 %
<i>Class Precision</i> Kompeten	85%
<i>Class Precision</i> Tidak	77,42 %
<i>Class Recall</i> Kompeten	70, 83 %
<i>Class Recall</i> Tidak	88,89%

3.6.2 AUC (Area Under Curve)

Selain *confusion matrix*, kurva *Receiver Operating Characteristic (ROC)* dihasilkan oleh *Rapidminer*. Kurva tersebut dapat dilihat pada gambar berikut.



Gambar 8. Pengukuran *AUC Optimistic*

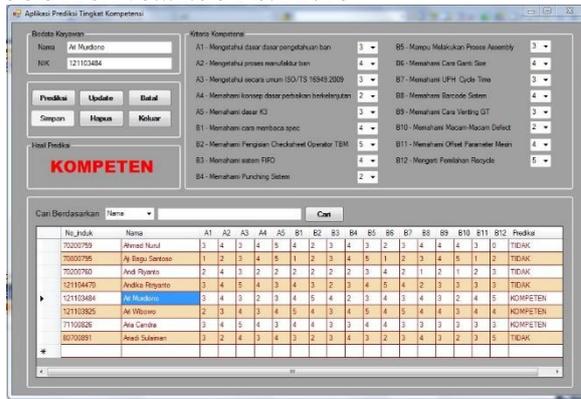
Hasil perhitungan divisualisasikan dengan kurva *ROC (Receiver Operating Characteristic)* atau *AUC (Area Under Curve)*. *ROC* memiliki tingkat nilai diagnosa yaitu:

1. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
2. Akurasi bernilai 0.80 – 0.90 = *good classification*
3. Akurasi bernilai 0.70 – 0.80 = *fair classification*
4. Akurasi bernilai 0.60 – 0.70 = *poor classification*
5. Akurasi bernilai 0.50 – 0.60 = *failure*

Hasil yang didapat dari *ROC* untuk *Algoritma C4.5* dengan menggunakan *data testing* sebesar 0.907 (Pengukuran *AUC Optimistic*), itu berarti hasil klasifikasi penelitian ini masuk kedalam tingkat diagnosa *excellent classification*.

3.7 Implementasi

Rule hasil klasifikasi dari *Algoritma C4.5* dapat diterapkan kedalam pembuatan aplikasi untuk prediksi tingkat kompetensi karyawan dengan menggunakan *Microsoft Visual Studio* seperti pada gambar berikut.



Gambar 9. Aplikasi Prediksi Tingkat Kompetensi Karyawan

Pada aplikasi prediksi tingkat kompetensi karyawan pada Gambar 9 dihasilkan prediksi kompeten atau tidaknya karyawan. Cara penggunaannya adalah dengan memasukan data karyawan pada program tersebut sesuai dengan atribut yang dibutuhkan, kemudian klik tombol Prediksi, maka secara otomatis akan tampil hasil prediksi kompetensi karyawan. Hasil prediksi akan berupa label kompeten dan tidak kompeten.

4. Kesimpulan

Berdasarkan hasil penelitian prediksi tingkat kompetensi karyawan menggunakan Algoritma C4.5, dapat diambil beberapa kesimpulan sebagai berikut:

- Algoritma C4.5 dapat implementasikan dalam sistem penilaian kompetensi karyawan PT Multistrada Arah Sarana dengan cara mengolah data – data hasil tes kompetensi di masa lalu untuk di jadikan *data training* dalam pembentukan *decision tree*. Dari *decision tree*

ini akan diperoleh *rule* yang dapat digunakan untuk memprediksi tingkat kompetensi karyawan.

- Pada penelitian ini penerapan algoritma *data mining* C4.5 terhadap kompetensi karyawan memiliki tingkat akurasi yang cukup tinggi yaitu dari 51 data yang di prediksi, terdapat 41 data yang diprediksi dengan tepat, berarti akurasi algoritma C4.5 pada penelitian ini mencapai 80,39%. Penerapan *algoritma C4.5* pada penelitian ini juga menghasilkan nilai *precision*, dan *recall* masing-masing 85 % dan 70,83%. Selain itu, dengan di dapatkan nilai *AUC Optimistic* sebesar 0,907 menjadikan penelitian ini masuk dalam kategori klasifikasi sangat baik (*Excellent classification*) karena masuk ke dalam *range* nilai antara 0.90 - 1.00.

Referensi

- Turban, E., dkk. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset
- Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
- Kusrini and Emha Taufiq Lutfhfi. 2009. *Algoritma Data Mining*: Andi, Yogyakarta.
- Berry, Michael J.A. dan Gordon S. Linoff. 2004. Data Mining Techniques for Marketing, Sales, Customer Relationship Management. Second Edition. Wiley Publishing, Inc.
- Basuki A dan Syarif I. 2003. Decision Tree. Politeknik Elektronika Negeri Surabaya (PENS) – ITS www.eepis-its.edu/~basuki/lecture/Decision tree.pdf. diakses tanggal 12 Oktober 2017

