

ANALISIS PENERIMAAN KARYAWAN POSISI FIELD COLLECTOR MENGUNAKAN ALGORITMA C4.5 PADA PT. PRISMAS JAMINTARA JAKARTA

Ayuni Asistyasari¹, Taufik Baidawi²
Program Studi Sistem Informatika

¹Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jl. Kramat 18 Jakarta Pusat
Email : Ayunioasis26@gmail.com

²Program Studi Manajemen Informatika
AMIK BSI Sukabumi
Jl. Cemerlang No. 8 Sukakarya Sukabumi
Email: taufiq.tfb@bsi.ac.id

Abstract — *PT Prismas Jamintara one the service human resources company every month selected employee on Field Collector position for customer needed. The process of human resources need have criterias such as Old, Education, Experience and vehicle. The purpose of processed employee is customer need. Algorithm C4.5 is one of the method for analysisit datas used entrophy and gain concept. With that Algorithm. made the decision tree to measure the best criteria and get accuracy value 71,54%+/-%+/-9,13%, precision 80,47% +/-8,16 and recall 83,78% +/-9,29% with ROC value 0,721.*

Intisari — PT. Prismas Jamintara sebagai perusahaan penyedia jasa tenaga kerja setiap bulannya melakukan penerimaan karyawan posisi *Field Collector* untuk memenuhi kebutuhan pelanggannya. Proses penerimaan karyawan tersebut melalui tahap seleksi penilaian kriteria-kriteria seperti usia, pendidikan, pengalaman dan kepemilikan kendaraan. Hal itu dilakukan untuk mendapatkan kriteria yang dibutuhkan oleh pelanggannya. Algoritma C4.5 merupakan salah satu metode yang dapat menganalisis data dalam jumlah yang banyak menggunakan konsep *entropy* dan *gain*. Dengan algoritma tersebut dibuatkan analisis pohon keputusan untuk mengukur kriteria terbaik dan menghasilkan angka akurasi 71,54%+/-9,13%, *precision* 80,47% +/-8,16 dan *recall* 83,78% +/-9,29% serta dengan angka curva ROC 0,721.

Kata Kunci: *penerimaan karyawan, decision tree, algoritma c4.5.*

I. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi yang saat ini berkembang dengan pesat tidak selalu berpengaruh positif terhadap kualitas para pencari kerja. Ditambah lagi dengan pertumbuhan jumlah para pencari kerja yang tinggi membuat perusahaan harus

lebih selektif dalam menyeleksi calon pegawai agar mendapatkan pegawai yang sesuai dengan kriteria..

Menurut Purnomo, dkk (2010). “Untuk menentukan calon pegawai yang memenuhi kriteria maka dibutuhkan rekomendasi yang tepat. Perekomendasi yang tepat membutuhkan jumlah data yang banyak, Algoritma C4.5 merupakan salah satu metode yang dapat menganalisis data dalam jumlah yang banyak atau biasa disebut data mining “.

PT Prismas Jamintara yang beralamat di Jl. Jend. Sudirman Kav.26 adalah salah satu penyedia tenaga kerja outsourcing di Indonesia. Salah satu permasalahan yang dihadapi PT Prismas Jamintara yaitu menentukan calon pegawai sesuai dengan kriteria yang dibutuhkan oleh klien pada posisi field collector.

Menurut Purnomo, dkk (2010). “Algoritma C4.5 dapat diimplementasikan dalam pembuatan aplikasi sistem penunjang keputusan penerimaan pegawai”. Atas dasar hal tersebut penulis menganalisa data yang ada menggunakan algoritma c4.5 dan menghasilkan rekomendasi kepada PT Prismas Jamintara Jakarta dalam menentukan calon pegawai sesuai dengan kriteria yang dibutuhkan oleh klien.

II. KAJIAN LITERATUR

Dalam penelitian ini diperlukan kajian literature dari berbagi sumber yang sudah terangkum dibawah ini :

A. KDD (*Knowledge discovery in database*)

Menurut Kusriani dan Luthfi (2009). “Istilah *data mining* dan KDD seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki

konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu hapan dalam keseluruhan proses KDD adalah *data mining*”.

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. *Data Selection*
2. *Pre-processing/ Cleaning*
3. *Transformation*
4. *Data mining*
5. *Interpretation/ Evalution*

B. Pohon Keputusan

Pohon keputusan menurut Kusri dan Luthfi (2009). “merupakan Metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekpresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu”.

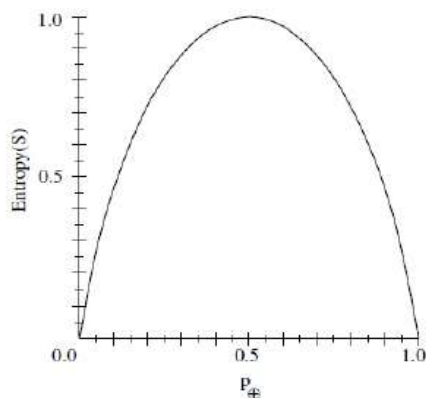
C. Algoritma C4.5

Menurut Dennis, dkk. (2013). “Algoritma c4.5 dapat menangani data numeric dan diskret. Algoritma C4.5 menggunakan rasio perolehan (*gain ratio*)”.

Namun Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam atuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep :

1. Konsep Entropy

Menurut Dennis, dkk. (2013). *Entropy* (S) “merupakan jumlah bit yang dibutuhkan untuk menyatakan suatu kelas”. semakin kecil nilai Entropy maka akan semakin *Entropy* digunakan dalam mengekstrak suatu kelas. Entropi dgunakan untuk mengukur ketidakaasian S.



Sumber : Denis (2013:50)

Gambar II.1. Grafik Entropy

Adapun untuk penghitungan Entropy dengan perhitungan :

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

Keterangan :

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- p_i : proporsi dari S_i terhadap S

2. Konsep Gain

Menurut Dennis, dkk. (2013). Gain (S,A) merupakan “Perolehan informasi dari atribut A relative terhadap output data S”. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

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

Keterangan :

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- $|S_i|$: Jumlah sampel pada pertisi ke -i
- $|S|$: Jumlah sampel dalam S

D. Rapid Miner

Menurut Dennis, dkk. (2013). “Perangkat lunak yang bersifat terbuka (*open source*). *Rapid Miner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining, text mining* dan analisis prediksi *Rapid Miner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik”.

III. BAHAN DAN METODE

Adapun metode penelitian yang penulis gunakan adalah sebagai berikut :

- A. Observasi
Penulis melakukan observasi langsung yang di laksanakan mulai tanggal 18 april – 18 mei 2016 di PT. Prisma Jamintara, Harmoni, Jakarta Pusat. Kegiatan dari observasi yang penulis lalukan adalah mengumpulkan data daftar pelamar di PT. Prisma Jamintara.
- B. Wawancara
Wawancara di lakukan di bagian HRD (*Human Resource Development*) di PT. Prisma Jamintara, Harmoni, Jakarta Pusat. Dengan menanyakan tentang syarat-syarat apa saja

yang di perlukan untuk dapat lulus dalam proses interview.

- C. Studi Pustaka
Studi Pustaka dilakukan untuk memperkuat penelitian. Dengan membaca beberapa jurnal ilmiah, buku-buku yang di baca di perpustakaan, artikel di internet yang berhubungan dengan tema penulisan penelitian ini.

Dari kegiatan tersebut menghasilkan data sampel dalam penelitian ini yang ditarik dengan teknik acak (*stratified random sampling*)

Tabel III.1
Sampel

| Bulan | Jumlah Transaksi |
|----------|------------------|
| Januari | 62 |
| Februari | 48 |
| Maret | 53 |
| Jumlah | 163 |

Sumber : Divisi HRD PT Prismas Jamintara (2016)

Dalam penelitian ini menggunakan model CRISP-DM (*Cross-Industry Standard Proseses for Data Mining*), yang terdiri dari 6 tahap yaitu :

1. *Business/Research Understanding Phase*

Proses perekrutan karyawan dengan posisi *field collector* awalnya dilakukan untuk memenuhi kebutuhan salah satu klien yang ada. Namun permasalahan muncul ketika banyaknya pelamar yang tidak sesuai dengan kriteria dan kemampuan yang dibutuhkan. Penulis menggunakan metode Algoritma C4.5 untuk mengklasifikasikan pelamar yang ada agar sesuai dengan kriteria dan kemampuan yang dibutuhkan.

2. *Data Understanding Phase*

Atribut-atribut yang menjadi patameter terlihat pada tabel III.2 yaitu:

Table III.2
Atribut dan Nilai Katagori

| No | Atribut | Nilai |
|----|------------|-----------|
| 1 | Usia | 18-35 |
| | | >35 |
| 2 | Kendaraan | Ada |
| | | Tidak Ada |
| 3 | Pendidikan | SMP |
| | | SMA |
| | | D3 |
| | | S1/S2 |

| 4 | Pengalaman | Ya |
|---|------------|-------|
| | | Tidak |

Sumber : Pengolahan Data (2016)

3. *Data Preparation Phase*

Setelah dilakukan *preprocesiing* data yang didapat dari PT Prismas Jamintara sebanyak 163 *record*. Untuk data training yang diambil sebanyak 130 data, dan data testing diambil sebanyak 33 data sesuai dengan rumus dibawah ini :

- Rumus Data Training (Data Sampel * 80%)
= 163*80% = 130 data
- Rumus Data Testing (Data Populasi * 20%)
= 163*20%= 33 data

4. *Modeling Phase*

Tahap ini juga disebut tahap *learning* karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan keputusan. Pada penelitian ini, pembuatan model menggunakan algoritma C4.5.

5. *Evaluation Phase*

Pada tahap ini dilakukan pengujian model untuk mendapatkan informasi model yang akurat. Dalam penelitian ini pengujian model menggunakan *confusion matrix* dan kurva ROC/AUC (*Area Under Cover*) Pengujian model menggunakan menghasilkan nilai *accuracy, precision, dan recall* menggunakan rapid miner

6. *Deployment Phase*

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model ke proses seleksi karyawan.

IV. HASIL DAN PEMBAHASAN

Langkah untuk menentukan pohon keputusan dengan menggunakan algoritma c4.5 dengan memakai data training sebanyak 130 data, yaitu :

- a. Menyiapkan *data training* sebanyak 130 data yang digunakan dalam penelitian ini. *Data training* biasanya diambil dari data primer yang sebelumnya sudah dikelompokkan kedalam kelas-kelas masing-masing.
- b. Hitung nilai *entropy* dan *gain*
Setelah dilakukan perhitungan *entropy* dan *gain* didapat hasil nilai *entropy* sebagai berikut :

Tabel VI.1
Nilai *entropy* dan *gain* node 1

| | Pelamar | (s1) | (s2) | Entro | Gain |
|---------|---------|------|------|--------|-------|
| Total | 130 | 99 | 31 | 0.7924 | |
| Usia | | | | | 0.001 |
| ≤27 | 87 | 67 | 20 | 0.7778 | |
| >27 | 43 | 32 | 11 | 0.8203 | |
| Pend | | | | | 0.011 |
| SMA | 101 | 74 | 27 | 0.8376 | |
| D3 | 13 | 11 | 2 | 0.6193 | |
| S1 | 16 | 14 | 2 | 0.5435 | |
| Kend | | | | | 0.050 |
| Ada | 62 | 40 | 22 | 0.9383 | |
| Tidak | 68 | 59 | 9 | 0.5638 | |
| Pengala | | | | | 0.080 |
| Ya | 89 | 60 | 29 | 0.9106 | |
| Tidak | 41 | 39 | 2 | 0.2811 | |

Sumber : Pengolahan Data (2016)

Dari tabel IV.1 dapat dilihat nilai *gain* tertinggi ada pada atribut pengalaman yakni 0,080362 sehingga dapat disimpulkan sebagai akar dari pohon keputusan atau *node* awal. Kemudian dilakukan kembali perhitungan nilai *entropy* dan *gain* untuk menentukan *node* 1.1, nilai yang dihitung berdasarkan atribut pengalaman “ya” dan “tidak”.

Perhitungan nilai *entropy* dan *gain* dapat dilihat pada tabel IV.2:

Tabel IV.2
Nilai *entropy* dan *gain* node 1.1

| | Pelamar | (s1) | (s2) | Entro | Gain |
|-----------|---------|------|------|--------|--------|
| Peng "Ya" | 89 | 60 | 29 | 0.9106 | |
| Usia | | | | | 0.0009 |
| ≤27 | 53 | 35 | 18 | 0.9245 | |
| >27 | 36 | 25 | 11 | 0.8880 | |
| Pend | | | | | 0.0044 |
| SMA | 85 | 58 | 27 | 0.9018 | |
| D3 | 2 | 1 | 1 | 1 | |
| S1 | 2 | 1 | 1 | 1 | |
| Kend | | | | | 0.0756 |
| Ada | 44 | 23 | 21 | 0.9985 | |
| Tidak | 45 | 37 | 8 | 0.6752 | |

Sumber : Pengolahan Data (2016)

Berdasarkan hasil perhitungan tabel IV.2 dapat dilihat nilai *gain* tertinggi ada pada atribut

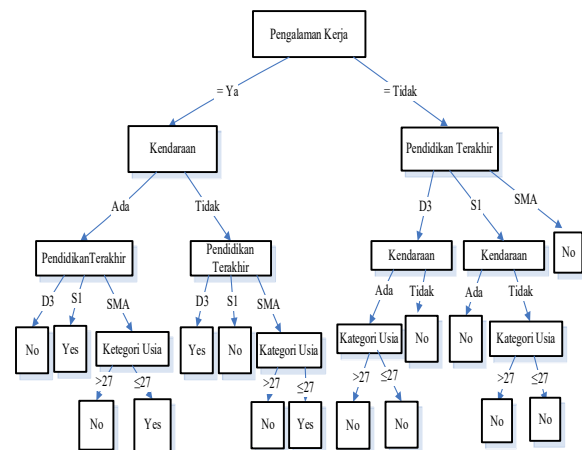
kendaraan yakni 0,075588 sehingga dapat disimpulkan kendaraan dijadikan *node* 1.1.

Tabel IV.3
Nilai *entropy* dan *gain* node 1.2

| | Pelama | (s1) | (s2) | Entro | Gain |
|--------------------|--------|------|------|--------|--------|
| Pengalaman "Tidak" | 41 | 39 | 2 | 0.2812 | |
| Usia | | | | | 0.0135 |
| ≤27 | 34 | 32 | 2 | 0.3228 | |
| >27 | 7 | 7 | 0 | 0 | |
| Pendidikan | | | | | 0.0365 |
| SMA | 16 | 16 | 0 | 0 | |
| D3 | 11 | 10 | 1 | 0.4395 | |
| S1 | 14 | 13 | 1 | 0.3712 | |
| Kendaraan | | | | | 0.0006 |
| Ada | 18 | 17 | 1 | 0.3095 | |
| Tidak | 23 | 22 | 1 | 0.2580 | |

Sumber : Pengolahan Data (2016)

Berdasarkan hasil perhitungan tabel IV.3 dapat dilihat nilai *gain* tertinggi ada pada atribut pendidikan yakni 0,036518 sehingga dapat disimpulkan pendidikan dijadikan *node* 1.2. Dan Setelah dilakukan perhitungan *entropy* dan *gain* hingga *node* terakhir maka akan menghasilkan pohon keputusan sebagai berikut :

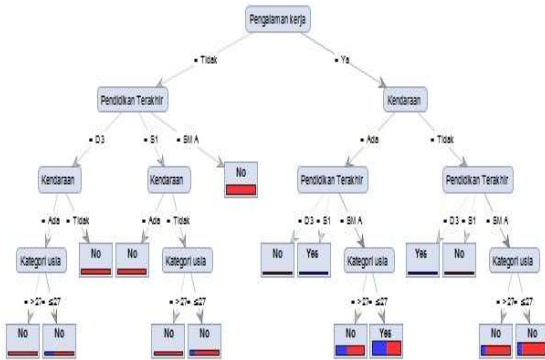


Sumber : Pengolahan Data (2016)

Gambar IV.1
Pohon keputusan perhitungan manual

Selain itu dilakukan perhitungan menggunakan aplikasi *Rapid Miner* untuk membandingkan

perhitungan manual menggunakan aplikasi. Dari data testing yang diolah menggunakan *Rapid Miner*, maka akan diketahui pohon keputusan sebagai gambar berikut.



Sumber : Pengolahan Data (2016)

Gambar IV.2

Pohon keputusan menggunakan *Rapid Miner*

Dari hasil yang ada dapat dievaluasi dengan kriteria seperti akurasi, kecepatan dan kehandalan. Nilai *accuracy*, *precision*, dan *recall* dari data training dapat di evaluasi menggunakan *rapid miner*. Dan hasil evaluasi data yang ada adalah sebagai berikut

accuracy: 71.54% +/- 9.13% (mikro: 71.54%)

| | true Yes | true No | class precision |
|--------------|----------|---------|-----------------|
| pred. Yes | 10 | 16 | 38.46% |
| pred. No | 21 | 83 | 79.81% |
| class recall | 32.26% | 83.84% | |

Sumber : Pengolahan Data (2016)

Gambar IV.3

Confusion Matrix (accuracy) data training

precision: 80.47% +/- 8.16% (mikro: 79.81%) (positive class: No)

| | true Yes | true No | class precision |
|--------------|----------|---------|-----------------|
| pred. Yes | 10 | 16 | 38.46% |
| pred. No | 21 | 83 | 79.81% |
| class recall | 32.26% | 83.84% | |

Sumber : Pengolahan Data (2016)

Gambar IV.4

Confusion Matrix (precision) data training

recall: 83.78% +/- 9.29% (mikro: 83.84%) (positive class: No)

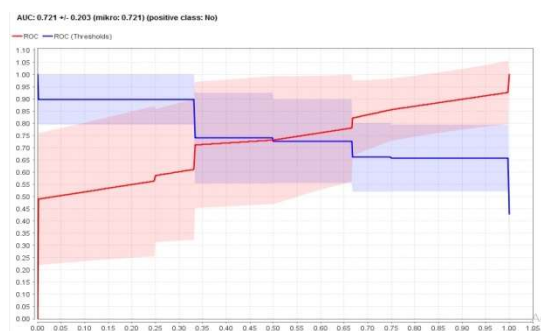
| | true Yes | true No | class precision |
|--------------|----------|---------|-----------------|
| pred. Yes | 10 | 16 | 38.46% |
| pred. No | 21 | 83 | 79.81% |
| class recall | 32.26% | 83.84% | |

Sumber : Pengolahan Data (2016)

Gambar IV.5

Confusion Matrix (recall) data training

Sedangkan untuk untuk validasi dapat menggunakan *curva ROC*. Dan hasil validasi penelitian ini adalah



Sumber : Pengolahan Data (2016)

Gambar IV.6

Grafik ROC

Model yang dihasilkan dengan metode *c4.5* terlihat pada nilai *accuracy*, *peceision* dan *recall*. Untuk klasifikasi nilai *AUC* dapat dibagi menjadi beberapa kelompok[3] :

- a) 0.90-1.00 = klasifikasi sangat baik
- b) 0.80-0.90 = klasifikasi baik
- c) 0.70-0.80 = klasifikasi cukup
- d) 0.60-0.70 = klasifikasi buruk
- e) 0.50-0.60 = klasifikasi salah

V. KESIMPULAN

Berdasarkan hasil perhitungan klasifikasi data training dengan model *decision tree* menggunakan algoritma *c4.5* yang dievaluasi dengan *confusion matrix* menghasikan angka akurasi 71,54% +/- 9,13%, *precision* 80,47% +/- 8,16 dan *recall* 83,78% +/- 9,29% serta dengan angka *curva ROC* 0,721 dan disimpulkan hasil penelitian ini dapat memberikan rekomendasi kepada PT Prisma Jamintara Jakarta dalam menentukan calon pegawai sesuai dengan kriteria yang dibutuhkan oleh klien.

VI. REFERENSI

- Dennis, Aprilia, Donny Aji Baskoro, Lia Ambarwati dan I Wayan Simri Wicaksana. 2013. Belajar Data Mining dengan Rapid Miner. Jakarta : Gramedia Pustaka Utama
- Divisi HRD. 2016. Data Seleksi Karyawan. Jakarta: PT. Prisma Jamintara.
- Gorunescu, Florin. 2011. *Data mining :concepts, models, and techniques* Verlag Berlin Heidelberg : Springer.
- Kusrini dan Emha Taufiq Luthfi. 2009. Algoritma Data Mining. Yogyakarta: Andi Offset.
- Purnomo, Joko, Wawan Laksito YS, dan Yustina Retno Wahyu U. 2010. Implementasi Algoritma C 4.5 dalam pembuatan aplikasi penunjang keputusan penerimaan pegawai CV. Dinamika Ilmu. ISSN: 2338-40. Jakarta: (p3m.sinus.ac.id/jurnal/index.php/TIKomSiN/article/download/158/26, diakses 20 Juni 2016).

BIODATA PENULIS



Ayuni Asistiyasari, S. Kom. Tahun 2016 Lulus Program Strata Satu (S1) Program Studi Sistem Informatika STMIK Nusa Mandiri.



Taufik Baidawi. Tahun 2004 lulus S1 Program Studi Sistem Informasi STMIK Bina Mulya Jakarta. Tahun 2011 lulus S2 Jurusan Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Beberapa tulisan yang dihasilkan diantaranya Jurnal Paradigma AMIK BSI Jakarta berjudul: Pemanfaatan Teknologi Informasi, Eksistensi Hak Kekayaan Intelektual Dalam Persepektif Hukum Siber (Cyber Law) Vol. IX. No.3, Agustus 2007. Prosiding Seminar Nasional Nasional Inovasi dan Tren (SNIT) BSI 2013 dengan Judul: 1. Sistem Pakar Identifikasi Penanggulangan Hama dan Penyakit Pada Anggrek Phalaenopsis Berbasis Web. 2. Implementasi E-Learning Untuk Menunjang Sistem Pembelajaran Menggunakan Server Cloud Desktop Computing Dengan Metode EYEOS. Jurnal Sistem Informasi STMIK Antar Bangsa Vol. IV No. 1 Februari 2015 berjudul: Perancangan Animasi Interaktif Pengenalan Bahasa Inggris Dasar Bagi Anak Kelas 1 dan 2 Pada SDN Teluk Pucung VII Bekasi. Jurnal Swabumi AMIK BSI Sukabumi Vol.III No. 1 September 2015 berjudul: Sistem Pendukung Keputusan Penilaian Tanaman Anggrek Pada Ud. Sanjiwani Orchid Menggunakan Metode Simple Additive Weighting.