

Seleksi Karyawan Menggunakan Algoritme C4.5 Pada Perusahaan Tekstil

Wilda Amalia
Universitas Buana Perjuangan Karawang
if15.wildaamalia@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar
Universitas Buana Perjuangan Karawang
amril.mutoi@ubpkarawang.ac.id

Santi Arum Puspita Lestari
Universitas Buana Perjuangan Karawang
santi.arum@ubpkarawang.ac.id

Abstract—

Seleksi karyawan baru pada perusahaan adalah hal yang sangat penting, karena karyawan yang terpilih harus mempunyai kemampuan-kemampuan yang sesuai dengan pekerjaannya. Banyaknya kriteria yang diperlukan oleh perusahaan menyebabkan pihak rekrutmen kurang efektif dalam menyeleksi. Agar dapat memudahkan dalam proses penyeleksian karyawan baru perusahaan perlu mengetahui pola data seleksi karyawan yang lulus dan gagal. Data seleksi karyawan diolah menggunakan teknik klasifikasi, algoritme C4.5 menghasilkan atribut yang berpengaruh dalam seleksi karyawan yaitu nilai tes, jahit dan sehat. Berdasarkan atribut tersebut membentuk aturan-aturan yang digunakan untuk seleksi karyawan. Penelitian ini menghasilkan tingkat akurasi yang sangat baik yaitu 100% dan nilai sensitivitas 100%. Pohon keputusan dan aturan yang dihasilkan ini mampu membantu pihak perusahaan dalam menentukan calon karyawan baru yang lulus atau gagal.

Kata kunci — C4.5, klasifikasi, pohon keputusan, seleksi karyawan

I. PENDAHULUAN

Seleksi karyawan baru pada perusahaan merupakan hal yang sangat penting, karena karyawan yang terpilih harus mempunyai kemampuan-kemampuan yang sesuai dengan kriteria perusahaan [1]. Seleksi karyawan adalah proses pengambilan keputusan memilih seseorang untuk mengisi suatu posisi pekerjaan atas dasar kesesuaian karakteristik individu dengan kebutuhan posisi pekerjaan tersebut [2]. Schuler dan Jackson [3] dan Marwansyah [4] menjelaskan pengertian dari seleksi karyawan yaitu proses mendapatkan karyawan yang sesuai dengan kebutuhan perusahaan. Tenaga kerja merupakan penduduk yang berada dalam usia kerja. Menurut UU No. 13 tahun 2003 Bab I pasal 1 ayat 2 disebutkan bahwa tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang atau jasa baik untuk memenuhi kebutuhan sendiri maupun untuk masyarakat. Secara garis besar penduduk suatu negara dibedakan menjadi dua kelompok, yaitu tenaga kerja dan bukan tenaga kerja [5]. Proses pemilihan karyawan memiliki kendala karena banyaknya kriteria yang diperlukan oleh perusahaan untuk menentukan calon karyawan terbaik. Masalah tersebut dapat diselesaikan dengan teknik *data mining*, karena teknik *data mining* mampu mengatasi masalah terkait seleksi data [6,7,8,9]

Penelitian yang dilakukan Pramadhani dan Setiadi [6] pada penyakit Infeksi Saluran Pernapasan (ISPA) dengan algoritme C4.5, menghasilkan klasifikasi ISPA Akut dan Tidak Akut. Berdasarkan penelitian tersebut faktor yang paling berpengaruh dalam penentuan klasifikasi penyakit ISPA adalah gizi, imunisasi dan usia, sedangkan atribut yang paling tidak berpengaruh adalah jenis kelamin. Kemudian, Elisa [7] menggunakan algoritme C4.5 untuk memperoleh faktor-faktor kecelakaan kerja konstruksi. Faktor-faktor yang berpengaruh berdasarkan penelitian tersebut adalah pekerja dan cara kerja, lingkungan tempat kerja dan alat pelindung diri. Sambani dan Nuraeni [8] juga menggunakan algoritme C4.5 pada penelitiannya tentang klasifikasi pola penjurusan SMK. Penelitian tersebut memiliki nilai akurasi 97,22% yang diperoleh berdasarkan aturan yang dihasilkan dari atribut yang berpengaruh. Atribut-atribut dalam penelitian tersebut terdiri dari bahasa inggris, bahasa indonesia, matematika, IPA, tes kesehatan dan tes olahraga. Selanjutnya Supriyanti [9] juga meneliti perbandingan kinerja algoritme C4.5 dan naive bayes untuk ketepatan pemilihan konsentrasi mahasiswa. Dalam penelitian tersebut kinerja algoritme C4.5 lebih unggul dibandingkan naive bayes karena tingkat akurasi algoritme C4.5 lebih tinggi. Nilai akurasi algoritme C4.5 sebesar 84.98% sedangkan algoritme naive bayes sebesar 82.01%.

Berdasarkan penelitian yang telah dilakukan tentang algoritme C4.5 yang terbukti mampu menyelesaikan berbagai masalah. Maka pada penelitian ini menerapkan algoritme C4.5 untuk seleksi karyawan dan menghasilkan pohon keputusan beserta aturan yang dapat melakukan seleksi karyawan. Sehingga dapat mempermudah pihak rekrutmen untuk proses penyeleksian karyawan dalam jumlah yang banyak.

Secara lengkap penelitian ini dijelaskan pada bagian II tentang algoritme C4.5, bagian III mengenai penerapan metode algoritme C4.5 untuk seleksi karyawan, pada bagian IV dijelaskan tentang hasil penelitian. Dan terakhir bagian V menjelaskan tentang kesimpulan dari penelitian.

II. ALGORITME C.45

Algoritme C4.5 pertama kali digunakan tahun 1993 dan merupakan pengembangan algoritme ID3 [10]. Algoritme ID3 tidak pernah melakukan backtracking untuk merevisi keputusan pemilihan attribute yang telah dilakukan sebelumnya. ID3 hanya menangani nilai-nilai attribute yang sedikit dan diskret. Di sisi lain keunggulan algoritme C4.5 adalah lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian [7]. Algoritme C4.5 mampu menangani atribut dengan tipe diskrit atau kontinu, mampu menangani atribut yang kosong (*Missing Value*), pemangkasan pohon keputusan, dan mampu melakukan pemilihan atribut yang dilakukan menggunakan *Gain Ratio* [10].

Dalam melakukan perhitungan algoritme C4.5 langkah pertama yaitu menghitung nilai entropi dan nilai *gain* dengan persamaan 1 dan 2.

$$\text{Entropi}(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (1)$$

Keterangan :

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

Setelah menghitung nilai entropi, selanjutnya ialah menghitung nilai *gain* untuk menemukan nilai tertinggi yang akan dijadikan sebagai node akar menggunakan persamaan 2

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

| Si | = Jumlah Kasus pada partisi ke-i

| S | = Jumlah Kasus dalam S

III. IMPLEMENTASI ALGORITME C4.5 UNTUK KARYAWAN

3.1 Data Penelitian

Data yang digunakan adalah data seleksi karyawan pada PT. Naigai Shirts Indonesia dengan rentang tahun 2016-2017, dan memiliki jumlah data sebanyak 210 data. Data seleksi karyawan memiliki tujuh atribut dan satu label yaitu Usia, Status, Nilai Tes, Tinggi Badan, Pendidikan, Sehat, Jahit dan Keterangan sebagai Label.

3.2 Aturan Dalam Pohon Keputusan

Cara untuk membangun pohon keputusan dimulai dengan menghitung nilai entropi menggunakan persamaan 1 dan menghitung nilai *gain* dengan persamaan 2. Agar pohon keputusan dapat dibangun contoh perhitungan entropi ditunjukkan pada persamaan 3 dengan mengimplementasikan persamaan 1 dan menggunakan atribut nilai tes. Dapat dilihat pada persamaan 3 hasil dari perhitungan nilai entropi untuk atribut tes adalah 0.365. Selanjutnya, cara menghitung nilai *gain* dengan mengimplementasikan persamaan 2 dan menggunakan atribut nilai tes diperoleh nilai *gain* sebesar 0.756.

$$\text{Entropi}(\text{Nilai Tes}, \leq 69) = \left(-\frac{7}{100} \times \log_2 \left(\frac{7}{100}\right)\right) + \left(-\frac{93}{100} \times \log_2 \left(\frac{93}{100}\right)\right) \quad (3)$$

$$= 0.365923651$$

$$\text{Entropy}(\text{Nilai Tes}, \geq 70) = \left(-\frac{86}{110} \times \log_2 \left(\frac{86}{110}\right)\right) + \left(-\frac{24}{110} \times \log_2 \left(\frac{24}{110}\right)\right) \quad (4)$$

$$= 0.756833632$$

$$\begin{aligned} \text{Gain (Total, Nilai)} &= 0.9905577 - \left(\left(\frac{100}{210} \times 0.365 \right) + \left(\frac{110}{210} \times 0.756 \right) \right) \\ &= 0.419871678 \end{aligned} \quad (5)$$

3.3 Evaluasi

Evaluasi pada penelitian ini menggunakan akurasi dan sensitivitas. Akurasi adalah tingkat kedekatan pengukuran kuantitas nilai benar dengan seluruh kuantitas, sensitivitas adalah fraksi dari sampel yang relevan yang diambil. Perhitungan akurasi dan sensitivitas berdasarkan persamaan 1 dan 2, yang dihitung menggunakan persamaan 6 dan 7.

Tabel 1 Confusion Matrix

<i>Predict</i>	<i>Actual</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Negative</i>
<i>Negative</i>	<i>False Positive</i>	<i>True Negative</i>

Akurasi dan Sensitivitas ditunjukkan pada persamaan 6 dan 7.

$$\text{Akurasi} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}} \quad (6)$$

$$\text{Sensitivitas} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (7)$$

IV. HASIL DAN PEMBAHASAN

Hasil klasifikasi yang digunakan merupakan data seleksi karyawan pada PT. Naigai Shirts Indonesia. Proses klasifikasi dilakukan dengan persamaan 1 dan persamaan 2. Setelah melakukan perhitungan nilai entropi dan *gain* kemudian menentukan nilai *gain* tertinggi yang akan dijadikan sebagai node akar, seperti yang ditunjukkan pada Algoritme 1. Kemudian melakukan validasi berdasarkan data uji yang sudah ditentukan.

```

Algoritme 1 Algoritme C4.5
Input : Data Training T; Atribut S
Output : Pohon Keputusan Tree
1 if T is NULL then
2   return failure
3 end if
4 if S is NULL then
5   return Tree as a single node with most frequent class label in T
6 end if
7 if |S| = 1 then
8   return Tree as a single node S
9 end if
10 set Tree { }
11 for a ∈ S do
12   set Info (a, T) = 0, and Split Info (a,T) = 0
13   compute entropi(a)
14   for v ∈ values (a,T) do
15     set Ta,v as the subset of T with attribute a=v
16     info (a,T)+=  $\frac{|Ta,v|}{|Ta|} \text{Entropi}(a_v)$ 
17   Split Info (a,T)+=  $\frac{|Ta,v|}{|Ta|} \log \frac{|Ta,v|}{|Ta|}$ 
18 end for
19 Set abest= argmar{Gain rasio (a,T)}
20 attach abest into Tree
21 for v ∈ values (abest, T) do

```

```

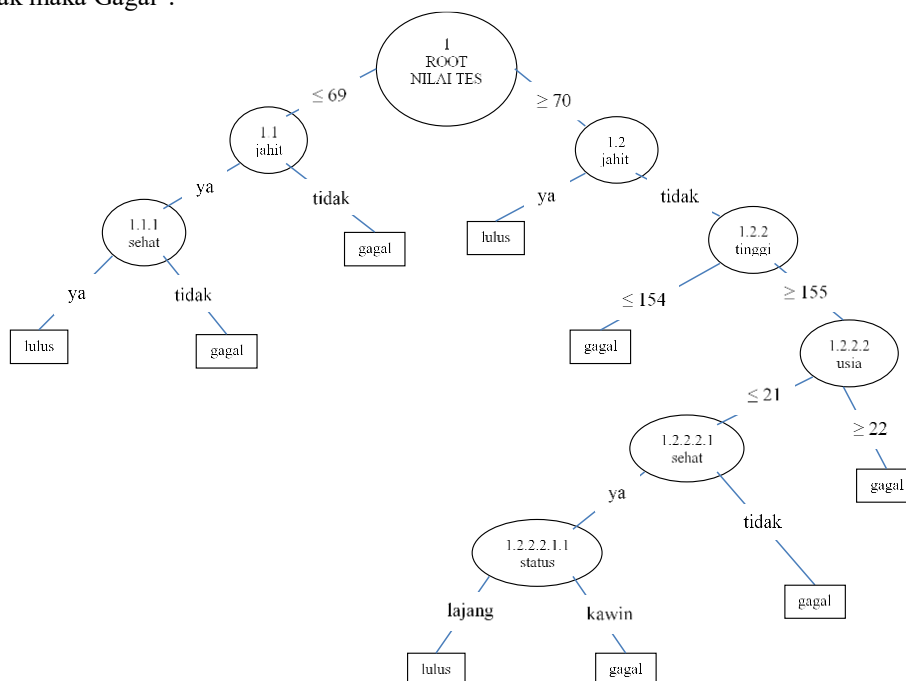
22 call C4.5
23 end for (Ta,v)
24 return Tree
    
```

Hasil komputasi dengan algoritme 1 ditunjukkan pada Tabel 2 berdasarkan atribut nilai tes. Atribut nilai tes menjadi node akar karena pada perhitungan atribut nilai tes memiliki nilai *gain* tertinggi.

Tabel 2 Hasil Perhitungan Entropy dan Gain Informasi Untuk Node 1

Node	Atribut	Nilai Atribut	S (jumlah kasus)	Y (Lulus)	N (Gagal)	Entropy	Gain
1	Total		210	93	117	0,99056	
	Usia						0,009371437
		≤ 21	200	91	109	0,99415	
		≥ 22	10	2	8	0,72193	
	Status						0,022395282
		Lajang	183	87	96	0,99825	
		Kawin	27	6	21	0,7642	
	Nilai Tes						0,419871678
		≤ 69	100	7	93	0,36592	
		≥ 70	110	86	24	0,75683	
	Tinggi Badan						0,062410125
		≤ 154	36	5	31	0,58132	
		≥ 155	174	88	86	0,9999	
	Pendidikan						0,132073595
		SMA	163	54	109	0,91624	
		SMK	47	39	8	0,65819	
	Sehat						0,149273179
		Ya	177	93	84	0,99813	
		Tidak	33	0	33	0	
	Jahit						0,24666017
		Ya	55	50	5	0,4395	
		Tidak	155	43	112	0,85191	

Berdasarkan Tabel 2 hasil komputasi algoritme 1 menghasilkan pohon keputusan yang ditunjukkan pada Gambar 1. Bersumber pada pohon keputusan yang telah terbentuk (Gambar 1) aturan pertama yang dihasilkan adalah “Jika Nilai Tes ≤ 69 dan Jahit Ya dan Sehat Ya maka Lulus” dan aturan kedua adalah “Jika Nilai Tes ≤ 69 dan Jahit Tidak maka Gagal”.



Gambar 1 Pohon Keputusan Klasifikasi Seleksi Karyawan

Pohon keputusan yang telah terbentuk selanjutnya diuji performanya menggunakan data *testing* yang sebelumnya telah disiapkan untuk mengetahui besarnya *akurasi* dan *sensitivitas*. Data *testing* sebanyak 42 ditunjukkan pada Tabel 3.

Tabel 3 Data *Testing*

Usia	Status	Nilai tes	Tinggi Badan	Pendidikan	Sehat	Jahit	Keterangan
1	Lajang	75	165	SMA	Ya	Tidak	?
2	Lajang	55	157	SMA	Ya	Tidak	?
3	Lajang	54	158	SMA	Ya	Tidak	?
4	Kawin	63	157	SMA	Tidak	Tidak	?
5	Kawin	68	144	SMA	Ya	Tidak	?
6	Lajang	88	156	SMK	Ya	Ya	?
7	Lajang	59	153	SMA	Tidak	Tidak	?
.							
.							
.							
23	Lajang	78	155	SMA	Ya	Tidak	?

Langkah selanjutnya dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yaitu, *Confusion Matrix* untuk pengukuran tingkat akurasi model, dan *split validation*. Berikutnya pengujian dilakukan dengan menghitung *akurasi* pada persamaan 7 dan *sensitivitas* pada persamaan 8. Berdasarkan hasil dari proses pengujian didapatkan nilai akurasi dan *confusion matrix* yang dapat dilihat pada Tabel 4.

Tabel 4 Hasil akurasi, presisi dan sensitivitas

Parameter	Persentase
Akurasi (Akurasi)	100.00 %
Class Sensitivitas Lulus	100.00 %
Class Sensitivitas Gagal	100.00 %

IV. KESIMPULAN DAN SARAN

Bersumber pada hasil penelitian menggunakan algoritme C4.5, dapat diambil kesimpulan bahwa algoritme C4.5 mampu melakukan klasifikasi seleksi karyawan dengan sangat baik. Penelitian klasifikasi seleksi karyawan dikatakan baik berdasarkan tingkat akurasi yang dihasilkan 100% dan sensitivitas 100%. Penerapan algoritme C4.5 ini menghasilkan pohon keputusan dan 11 aturan yang membantu pihak rekrutmen dalam menentukan calon karyawan yang sesuai dengan kebutuhan perusahaan.

Penelitian selanjutnya disarankan dapat mengembangkan penelitian menjadi aplikasi web yang dapat digunakan sebagai sistem pengambilan keputusan.

PENGAKUAN

Naskah ilmiah ini adalah sebagian penelitian Tugas Akhir milik Wilda Amalia dengan judul Penerapan Algoritme C4.5 untuk seleksi karyawan pada PT. Naigai Shirts Indonesia yang dibimbing oleh Amril Mutoi Siregar dan Santi Arum Puspita Lestari.

DAFTAR PUSTAKA

- [1] W. Amalia, "Penerapan Algoritme C4.5 untuk Seleksi Karyawan Pada PT. Naigai Shirts Indonesia" 2019.
- [2] Castetter and William B. 1996. *The Human Resource Function in Educational Administrastiofi* Columbus, Ohio: Merril, Englewood Cliffs, New Jersey.
- [3] Schuler, Randal S and S. E. Jackson. 1997. *Strategic Theory Research*. Oxlord Blacwell
- [4] Marwansyah and Mukaram. 2000. *Manajemen Sumber Daya Manusia*, Bandung, Politeknik Bandung Press.
- [5] UU No. 13 tahun 2003 tentang Tenaga Kerja.
- [6] A.E. Pramadhani and T. Setiadi, "Penerapan *Data mining* untuk Klasifikasi Prediksi Penyakit Ispa dengan Algoritme Decision Tree (ID3)" vol. 02, no. 01, Februari 2014.
- [7] Elisa, "Analisa dan Penerapan Algoritme C4.5 Dalam *Data mining* Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Konstruksi PT. Arupadhatu Adisesanti", vol. 02, no. 01, 2017.
- [8] E. B. Sambani and F. Nuraeni, "Penerapan Algoritme C4.5 untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya", vol. 09, no. 03, pp 149-157, Oktober 2017.

- [9] W. Supriyanti., Kusrini, and A. Amborowati, “Perbandingan Kinerja Algoritme C4.5 dan Naive Bayes untuk Ketepatan Pemilihan Konsentrasi Mahasiswa”. vol. 01, no. 03, 46–52, 2016.
- [10] Informatikalogi.com. Algoritme C4.5. <https://informatikalogi.com/algoritme-c4-5/>