

# Analisis Penerapan Data Mining Dengan Metode Algoritma C4.5 Untuk Pendataan Karyawan Tetap Di Koni Sumatera Utara

*Analysis Of The Application Of Data Mining Using The C4.5 Algorithm Method For  
Data Collection On Permanent Employees In North Sumatera Coni*

Gina Sonia<sup>1</sup>, Agil Indriyani<sup>2</sup>

<sup>1,2,3</sup>Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera  
Utara

E-mail: <sup>1</sup>soniadomo28@gmail.com, <sup>2</sup>agilindriyani123@gmail.com

## Abstrak

Koni Sumatera Utara merupakan jenis perusahaan yang melakukan pergerakannya dalam dunia olahraga. Koni Sumatera Utara memiliki karyawan yang bisa dibilang banyak, untuk mendata karyawan tetap membutuhkan waktu yang relative lama jika mengecek dengan manual. Perusahaan mereka memiliki waktu yang sulit untuk memutuskan karyawan mana yang lebih disukai. Kendalanya adalah pengelola yang tidak menerapkan metode yang mampu menangani masalah dalam multi kriteria prioritas. Tidak hanya itu, manajemen sumber daya manusia sering mengalami kesulitan dalam pemilihan pegawai tetap, karena banyak pegawai yang harus dievaluasi. Hal inilah yang menjadi kekurangan dalam memutuskan apakah seseorang akan dipilih menjadi pegawai tetap atau tidak. Dengan bantuan data mining yang diterapkan dalam penggunaan metode algoritma C4.5 maka kecepatan dan ketepatan pengambilan keputusan pada evaluasi personel dapat ditingkatkan. Hasil pengujian yang diperoleh algoritma C4.5 mempunyai nilai presisi, akurasi dan recall yang baik.

Kata kunci: Koni Sumut, Data Mining, Algoritma C4.5

## Abstract

*Koni North Sumatra is a type of company that makes its moves in the world of sports. Koni North Sumatra also has a lot of employees, registering permanent employees requires relatively long time if you run it manually. Their company struggled to decide which employee to prefer. The problem is that managers do not apply methods capable of dealing with problems in multi-priority priorities. In addition, HR management often finds it difficult to select permanent employees, as there are many employees to evaluate. It is the failure to decide whether someone will be chosen as a permanent employee or not. With the help of data mining with the applied use of the C4.5 algorithm method, the speed and accuracy of decision making on personnel evaluation can be increased. The test results obtained by the C4.5 algorithm have good values of precision, accuracy and shrinkage.*

Keywords: Koni Sumut, Data Mining, C4.5 . Algorithm

## 1. PENDAHULUAN

Sumber daya manusia memiliki peranan penting karena berfungsi menjadi faktor utama pada setiap tindakan maupun aktivitas perusahaan dalam upaya pencapaian tujuan, seperti pada upaya mencari keuntungan atau bahkan untuk mempertahankan progres perusahaan tersebut menjadi lebih baik dan mencapai kejayaan yang diharapkan.[1] Sumber daya manusia menjadi salah satu sinergi oaling besar yang dibutuhkan dalam proses pencapaian tujuan. Maka saat ini,

---

diperlukan pengelola SDM yang handal yang didasari dengan sistem penilaian kinerja alias *performance appraisal*. [2]

Perkembangan perusahaan juga setelah itu kualitas SDM harus ditingkatkan yang terlibat. Tenaga kerja yang berkualitas dicirikan oleh kemampuan setiap individu untuk melakukan tugas yang diberikan kepada mereka. Perusahaan juga memberikan pengakuan kepada karyawan yang layak untuk mendorong kinerja karyawan lainnya. Evaluasi pegawai bukanlah hal yang mudah ketika pegawai yang ada sangat banyak, sehingga tim evaluasi harus mempertimbangkan kriteria dengan cermat dan hati-hati. Tanpa otomatisasi penambahan data, manipulasi data dapat terjadi karena sulit untuk membuat keputusan dengan kriteria dan data skala besar. Akibatnya, perusahaan tidak dapat memperoleh bakat kompetitif. Salah satu cara untuk melakukannya adalah dengan mengembangkan metode penelitian baru yang dapat meningkatkan akurasi hasil pengambilan keputusan: penggunaan data mining.

Sistem pengambilan keputusan merupakan bentuk sistem yang proses perancangannya dilakukan guna membantu tindakan pengambilan keputusan. Sebagian besar bisnis menerapkan teknik pendukung keputusan ini dalam setiap penyelesaian tugasnya. Pada data mining, jenis metode yang cukup sering digunakan adalah dengan menerapkan metode pohon keputusan yang sangat cocok digunakan dalam menyelesaikan permasalahan dengan jumlah data yang relatif besar. Jenis metode ini adalah jenis yang banyak dipergunakan dalam berbagai masalah karena proses pembangkitan data yang tergolong cepat, dan disertai dengan hasil model yang cukup mudah dipahami. Hasil prediksi yang terlihat pun sangat diperlukan dalam proses pengambilan keputusan.

Pada sistem yang diterapkan oleh KONI SUMUT saat ini masih mempunyai berbagai kesulitan dengan durasi yang relatif lama dalam perekrutan karyawan baru. Ini karena pengambil keputusan di KONI Sumatera Utara masih menggunakan metode manual untuk memilih dan mengevaluasi calon karyawan baru untuk hasil pengambilan keputusan yang berdampak. masih subjektif. Maka tujuan dilakukannya penelitian ini yaitu agar dapat memberikan kemudahan bagi KONI SUMUT pada tahapan pengambilan keputusan, satu diantaranya adalah pada pemilihan karyawan baru yang dianggap layak dalam penempatan menjadi karyawan tetap.

## 2. METODE PENELITIAN

Dengan melakukan analisis keputusan oleh karyawan baru menjadi karyawan tetap yang menerapkan algoritma C4.5

### 2.1 Algoritma C4.5

Algoritma C4.5 menggunakan struktur pengumpulan data dengan jumlah yang cukup besar yang dijadikan sebagai bentuk sebuah himpunan rekaman yang lebih efisien dan menggunakan rangkain aturan keputusan yang sesuai. [3]

Algoritma C4.5 yang diterapkan dalam upaya pembangunan pohon harus memperhatikan beberapa hal berikut, diantaranya:

- a. Memilih atribut yang menjadi akar
- b. Membuat cabang pada setiap nilai
- c. Membagi kasus pada cabang tersebut
- d. Mengulangi tahapan pada seluruh cabang hingga setiap kasus mempunyai kelas yang selaras. [4]

Sedangkan dalam pembangunan *decision tree* dengan menerapkan algoritma C4.5, beberapa tahapan yang harus dipenuhi yaitu:

- a. Menyiapkan data training.
- b. Menentukan akar dari pohon. [5]
- c. Konsep Entropy diterapkan pada proses megekstrak sebuah kelas. Entropy digunakan pada ketidakaslisan s, dengan rumus yang digunakan, diantaranya:

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

dengan:

- S = Himpunan Kasus  
A = Fitur  
n = Total partisipasi  
p<sub>i</sub> = proporsi dalam S<sub>i</sub> oleh S [6]

- d. Menerapkan metode gain dengan rumus berikut:

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

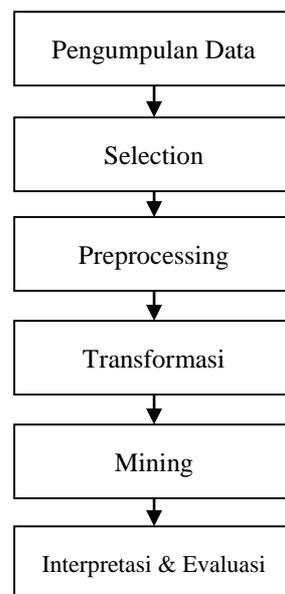
Keterangan:

- A = Atribut
- S = Sampel
- n = Total partisis himpunan atribut A
- $|S_i|$  = Total sampel dalam pertisi ke  $-i$
- $|S|$  = Total sampel pada S[6]

- e. Ulangi tahap ke 2 hingga seluruh record terpartisi.
- f. Progres partisi *decision tree* akan berhenti ketika setiap record mempunyai kelas N yang sama, dan rekaman tersebut tidak mempunyai atribut yang harus dipartisi, serta tidak terdapat rekaman pada cabang yang kosong.

## 2.2 Tahapan Data Mining

Data Mining adalah hal dengan integrasi dalam *Knowledge Discovery in Databases* (KDD), yang mempunyai tujuan dalam mentransformasi data asli menjadi sebuah data informasi baru yang dibutuhkan oleh beberapa pihak. Apabila dilakukan penggambaran terhadap KDD, maka gambaran tahapan yang harus dilalui adalah:



Gambar 1. Tahapan KDD

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Dengan menerapkan data karyawan pada bulan November - September 2022 sebanyak 30 nama Karyawan, adapun teknik pengumpulan yang dilakukan yaitu:

- a. Observasi, penulis melakukan langsung observasi selama kegiatan KP berlangsung di bulan November – September 2022, kegiatan yang dilakukan adalah mengumpulkan karyawan tidak tetap.
- b. Wawancara, penulis melakukan beberapa wawancara kepada HRD atau Ketua

KONI Sumut, Tentang Indikator apa yang dibutuhkan menjadi karyawan tetap.

Berikut adalah Contoh Data yang sudah dikumpulkan:

Tabel 1. Data Karyawan Tidak Tetap

No	Nama	Jenis Kelamin	Agama	Umur	Masa Kerja	Terlambat Kerja	Absensi	Loyalitas	Disiplin	Kinerja
1	N01	L	Islam	22	<2 Bulan	6	30	Sangat Baik	Baik	Baik
2	N02	P	Islam	23	<1 Bulan	3	30	Baik	Sangat Baik	Sangat Baik
3	N03	P	Islam	21	<3 Bulan	3	30	Baik	Sangat Baik	Baik
4	N04	P	Islam	22	<1 Bulan	1	30	Cukup Baik	Sangat Baik	Baik
5	N05	P	Islam	20	<2 Bulan	4	28	Sangat Baik	Baik	Sangat Baik
6	N06	P	Kristen	23	<2 Bulan	3	30	Baik	Sangat Baik	Sangat Baik
7	N07	P	Islam	24	<3 Bulan	5	30	Sangat Baik	Baik	Cukup Baik
8	N08	L	Islam	25	<3 Bulan	6	30	Baik	Baik	Cukup Baik
9	N09	L	Islam	23	<3 Bulan	6	25	Sangat Baik	Baik	Baik
10	N10	P	Islam	22	<3 Bulan	7	30	Cukup Baik	Kurang	Kurang
11	N11	L	Islam	21	<2 Bulan	3	30	Baik	Baik	Sangat Baik

### 3.1. Selection dan Preprocessing Data

Dalam tahap ini, akan diterapkan sebuah seleksi terhadap data yang mempunyai pengaruh dalam penelitian ini dan dijadikan sebagai atribut data set penelitian. Dataset yang telah diterima tersebut akan melewati proses pemilihan sesuai dengan parameter yang diteliti.[7] Parameter yang digunakan berupa atribut pada seluruh data nama karyawan yang belum dinyatakan sebagai karyawan tetap dan dijadikan sebagai variabel input pada penelitian ini. Berikut adalah data karyawan dengan indikator yang dijadikan sebagai data training dalam pembuatan *decision tree*. [8]

Tabel 2. Pemilihan Indicator

Nama	Masa Kerja	Absensi	Loyalitas	Disiplin	Kinerja
N01	<2 Bulan	30 Hari	Sangat Baik	Baik	Baik
N02	<1 Bulan	30Hari	Baik	Sangat Baik	Sangat Baik
N03	<3 Bulan	30Hari	Baik	Sangat Baik	Baik
N04	<1 Bulan	30Hari	Cukup Baik	Sangat Baik	Baik
N05	<2 Bulan	28Hari	Sangat Baik	Baik	Sangat Baik
N06	<2 Bulan	30Hari	Baik	Sangat Baik	Sangat Baik
N07	<3 Bulan	30Hari	Sangat Baik	Baik	Cukup Baik
N08	<3 Bulan	30Hari	Baik	Baik	Cukup Baik
N09	<3 Bulan	25Hari	Sangat Baik	Baik	Baik
N10	<3 Bulan	30Hari	Cukup Baik	Kurang	Kurang
N11	<2 Bulan	30Hari	Baik	Baik	Sangat Baik

### 3.2 Hasil Transformasi

Dalam proses ini, data yang telah diterima akan diolah menjadi bentuk yang sesuai dengan penelitian data mining, yakni dengan merubah masa kerja dan absensi atribut sehingga tabel menjadi seperti dibawah ini:

Tabel 3. Klasifikasi Masa Kerja

Masa Kerja	Kriteria
>3 Bulan	Sangat Baik
<3 Bulan	Baik
<2 Bulan	Cukup Baik
<1 Bulan	Kurang

Tabel 4. Klasifikasi Absensi

Absensi	Kriteria
30 hari	Sangat Baik
<28 Hari	Baik
<25Hari	Cukup Baik
<23Hari	Kurang

Berikut adalah hasil proses dari transformasi, diantaranya:

Tabel 5. Hasil Transformasi

Nama	Masa Kerja	Absensi	Loyalitas	Disiplin	Kinerja
N01	Cukup Baik	Sangat Baik	Sangat Baik	Baik	Baik
N02	Kurang	Sangat Baik	Baik	Sangat Baik	Sangat Baik
N03	Cukup Baik	Sangat Baik	Baik	Sangat Baik	Baik
N04	Kurang	Sangat Baik	Cukup Baik	Sangat Baik	Baik
N05	Cukup Baik	Baik	Sangat Baik	Baik	Sangat Baik
N06	Cukup Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik
N07	Baik	Sangat Baik	Sangat Baik	Baik	Cukup Baik
N08	Baik	Sangat Baik	Baik	Baik	Cukup Baik
N09	Baik	Cukup Baik	Sangat Baik	Baik	Baik
N10	Baik	Sangat Baik	Cukup Baik	Kurang	Kurang
N11	Cukup Baik	Sangat Baik	Baik	Baik	Sangat Baik

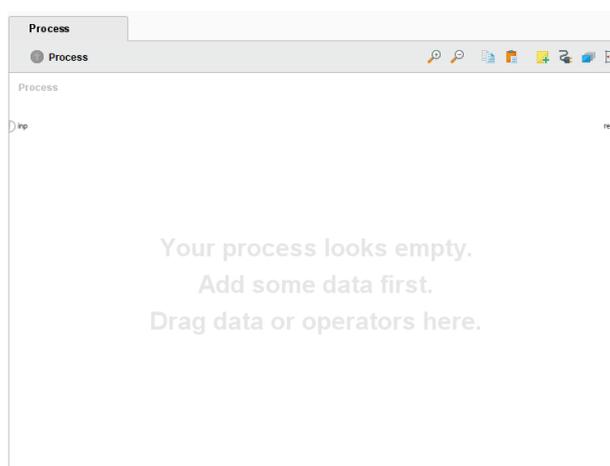
Data tersebut telah melewati proses seleksi dan transformasi yang selanjutnya akan menerapkan konsep pohon keputusan. Dataset yang menjadi atribut klasifikasi selanjutnya melewati proses analisis dengan menerapkan Rapid Miner. Data karyawan tidak tetap KONI Sumut November – September dijadikan sebagai dataset dalam penelitian ini. Dataset mempunyai 5 atribut dengan total 30 data yang menjadi *data training*. Selanjutnya pohon keputusan akan terbentuk apabila perhitungan jumlah nilai entropy dan nilai entropy pada setiap atribut selesai, dan dengan menemukan nilai information gain, yang sesuai dengan tabel berikut:

Tabel 6. Tabel Perhitungan Entropy dan Gain

	Jumlah(S)	Diterima(Si)	Ditolak(Si)	Entropy	
Total Masa Kerja	30	19	21	0.777544506	
	Sangat Baik	4	1	0.811278124	0.39388534
	Baik	9	6	0.918295834	
	Cukup Baik	12	12	0	
	Kurang	5	0	0	
Absensi					0.115789432
	Sangat Baik	22	15	0.902393283	
	Baik	3	3	0	
	Cukup Baik	1	1	0	
	Kurang	4	4	0	
Loyalitas					-0.067483121
	Sangat Baik	13	9	0.89049164	
	Baik	15	10	0.918295834	

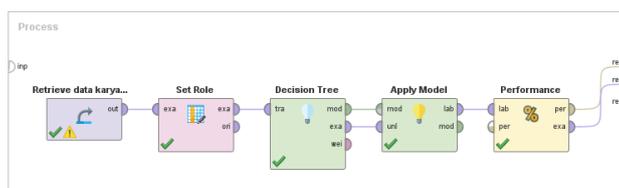
Disiplin	Cukup Baik	2	0	2	0	
	Kurang	0	0	0	0	
Kinerja	Sangat Baik	9	5	4	0.99107606	-0.064401526
	Baik	15	12	3	0.721928095	
	Cukup Baik	0	0	0	0	
	Kurang	6	2	4	0.918295834	
	Sangat Baik	11	9	2	0.684038436	
	Baik	11	7	4	0.945660305	
	Cukup Baik	3	2	1	0.918295834	-0.032162631
	Kurang	5	1	4	0.721928095	

Rapid Miner menjadi *software* yang dipilih dalam pengujian data, dan dengan disertai algoritma C4.5



Gambar 2. Tampilan New Project Rapid Miner

Dengan menerapkan algoritma C4.5, akan ditemukan akurasi serta *decision tree* sesuai dengan dataset yang digunakan. Gambar di bawah ini menampilkan model dalam proses pengujian dengan menerapkan algoritma pohon keputusan disertai dengan validasi data penghubung *building block performance* dengan menerapkan *software* Rapid Miner.



Gambar 3. Tampilan Pengujian Validasi Menggunakan Algoritma C4.5

Berikut tampilan mengenai rule model dari pohon keputusan yang dihasilkan dan akan mempermudah beberapa pihak dalam tindakan pengambilan keputusan karyawan tetap.



Gambar 4. Pohon Keputusan Penentuan Jadwal *Overtime*

Gambar berikut menunjukkan data akurasi, *precision*, dan *recall* dalam proses pengolahan datanya.

accuracy: 86.67%

	true Diterima	true Ditolak	class precision
pred. Diterima	19	4	82.61%
pred. Ditolak	0	7	100.00%
class recall	100.00%	63.64%	

Gambar 5. Hasil Perhitungan Confusion Matrix

Maka data akurasi yang dihasilkan senilai 86%, sedangkan *precision* yang didapatkan senilai 82% dan perhitungan *recall* 100%.

#### 4. KESIMPULAN DAN SARAN

Dengan dilakukannya penelitian ini yang menerapkan algoritma C4.5 pada pendataan karyawan tetap KONI Sumatera Utara, maka klasifikasi yang didapatkan mempunyai nilai akurasi sebesar 86%, sedangkan *precision* yang didapatkan senilai 82% dan perhitungan *recall* 100%.

Sebagai sebuah saran, Peneliti mengharapkan kepada peneliti selanjutnya untuk menggunakan atribut yang lebih bervariasi kedepannya.

#### DAFTAR PUSTAKA

- [1] F. Rachmawati, "Analisis Algoritma C4.5 Untuk Pengangkatan Karyawan Tetap Studi Kasus PT Citra Abadi Sejati Bogor," *J. Ilm. Teknol. dan Inf.*, vol. 6, no. 2, pp. 25–36, 2016.
- [2] M. Ardi, R. W. Sembiring, and D. Hartama, "Analisis Regresi Linier Berganda Dalam Pengaruh Tingkatan Absensi Dan Keterlambatan Jam Kerja Terhadap Prestasi Kerja Karyawan Pada Kantor Kementrian Agama," *J. SmartEDU*, vol. 1, no. 3, pp. 93–100, 2022.
- [3] Saefudin and S. Lestari, "Sistem Pendukung Keputusan Untuk Penilaian Kinerja Karyawan PT. Mulya Spindo Mills Menggunakan Metode Algoritma C4.5," *J. ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform.*, vol. 2, no. September, pp. 40–43, 2015, doi: 10.30656/protekinfor.v2i0.48.
- [4] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [5] A. N. Z. Hidayah and A. F. Rozi, "Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 ( Studi Kasus : Universitas Mercu Buana Yogyakarta )," *J. Inf. Syst. Artif. Intell.*, vol. 1, no. 2, pp. 73–80, 2021.
- [6] A. Asistiyasari and T. Baidawi, "Analisis Penerimaan Karyawan Posisi Field Collector Menggunakan Algoritma C4.5 Pada Pt. Prismas Jamintara Jakarta," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 2, no. 2, pp. 1–6, 2017.
- [7] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform.*, vol. 4, no. September, pp. 694–702, 2020, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti>.
- [8] Febryantahanuji, I. Sembiring, and H. D. Purnomo, "Pengambilan Keputusan Pegawai Tidak tetap menjadi Pegawai Tetap Dengan Decission Tree," *Joined J. (Journal Informatics Educ.*, vol. 1, no. 2, p. 99, 2018, doi: 10.31331/joined.v1i2.736.