

Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritma C4.5

Tulus Harry Lamramot ¹, Asep Id Hadiana ², Irma Santikarama ³

¹²³ Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani, Cimahi 40513, Indonesia
e-mail: tulus23789@gmail.com^{*1}, ahadiana@gmail.com², irma.santikarama@lecture.unjani.ac.id³

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 31 Januari 2022

Revisi Akhir: 13 Juni 2022

Diterbitkan Online: 02 Juli 2022

Kata Kunci :

Atrisi Karyawan, Prediksi, Algoritma C4.5,
Decision Tree, Classification, Data Mining

Korespondensi :

Telepon / Hp : +62 812 2052 8694

E-mail : tulus23789@gmail.com

A B S T R A K

Karyawan yang berbakat menjadi bagian penting bagi sebuah perusahaan. Menjadi masalah yang krusial bagi perusahaan ketika terdapat karyawan berbakat yang meninggalkan perusahaan untuk pindah ke perusahaan yang menawarkan peluang yang lebih baik. Kesiapan perusahaan dalam menghadapi masalah tenaga kerja dapat terlaksana dengan diberlakukannya sistem yang dapat memprediksi kasus atrisi karyawan. Penelitian ini bertujuan untuk merancang sistem prediksi awal terhadap atrisi karyawan. Dalam sudut pandang informatika untuk mencegah terjadinya atrisi karyawan, digunakan sistem Data Mining yang mampu memprediksi terjadinya atrisi karyawan tersebut dengan metode prediksi yaitu metode decision tree menggunakan algoritma C4.5. Data yang digunakan adalah data primer. Populasi yang digunakan adalah data yang diambil dari perusahaan PT.Indorama Petrochemicals yang berjumlah 1010 sampel data karyawan. Penelitian ini mendapatkan akurasi data sebesar 92,9% dengan algoritma C4.5 sehingga mendapatkan hasil yang baik pada prediksi atrisi karyawan tersebut.

1. PENDAHULUAN

Setiap perusahaan pasti menginginkan kesuksesan. Untuk mencapai kesuksesan tersebut, perusahaan harus mempunyai strategi dan rencana jangka panjang dalam menyeimbangkan kekuatan dan kelemahan internal maupun ancaman eksternal agar mempertahankan keunggulan dalam bersaing. Permasalahan pada suatu perusahaan adalah salah satunya tenaga kerja yang cukup kompleks dari segi pengadaan Sumber Daya Manusia (SDM) maupun pada penempatan dan pengembangannya[1]. Agar tujuan perusahaan dapat tercapai dengan baik, perlu didukung oleh Sumber Daya Manusia yang memenuhi syarat-syarat dan kriteria pada organisasi tersebut. Sumber daya manusia merupakan faktor utama yang terdapat dalam organisasi. Kebutuhan akan profesionalisme untuk membantu organisasi dalam melaksanakan tujuannya, menunjukkan semakin berperannya sumber daya manusia dalam mencapai keberhasilan suatu organisasi[1].

Dalam mendapatkan Sumber Daya Manusia yang baik harus dilakukan manajemen yang baik dalam segi kemampuan untuk melakukan program-program di suatu organisasi tersebut. Manajemen dalam sebuah perusahaan perlu melakukan seleksi terhadap karyawan yang ada sehingga seluruh Sumber Daya Manusia dalam perusahaan memenuhi kualifikasi dan diharapkan mendukung kemajuan perusahaan[2]. Manajemen yang lebih tinggi akan lebih tertarik untuk mengetahui calon karyawan yang mungkin benar-benar keluar sehingga mereka dapat mengatur Tindakan karyawan yang ingin keluar dari perusahaan[2]. Proses seleksi tersebut melibatkan investasi waktu & uang, sehingga menyebabkan kerugian besar jika salah satu dari sumber daya ini keluar dari organisasi[2]. Kinerja karyawan

adalah salah satu faktor penting yang mempengaruhi dalam proses produktif sebuah perusahaan. Dalam studi peneliti terdahulu, faktor pendorong terdapat efek negatif pada kepuasan kerja karyawan dan akhirnya mengarah pada keputusan untuk meninggalkan lingkungan kerja[3]. Faktor paling umum yaitu keseimbangan pekerjaan, keluarga, gaya hidup, hubungan yang buruk dengan rekan kerja, stres dalam pekerjaan dan tidak terpenuhinya kebutuhan pribadi karyawan dalam perusahaan seperti jam kerja yang panjang, persyaratan relokasi dan hubungan pimpinan yang tidak memuaskan dan ketidakadilan yang dirasakan dalam penugasan kerja. Semua faktor tersebut berpotensi mengganggu prioritas dalam rumah dan gaya hidup karyawan[4]. Faktor inilah yang disebut sebagai Atrisi Karyawan.

Atrisi Karyawan adalah atau pemberhentian tetap dengan pengurangan yang bertahap pada jumlah staf yang terjadi saat karyawan pensiun, mengundurkan diri atau tidak diganti[4]. Bagi Karyawan perubahan seperti itu menyebabkan dampak negatif yang menimbulkan rasa cemas, stress, dan rasa tidak nyaman terhadap lingkungan kerja dan rasa tidak aman terhadap kelangsungan dari pekerjaan mereka. Begitupun dengan perusahaan tersebut dapat membuat pendapatan suatu perusahaan akan berkurang. Berdasarkan fenomena yang ditemukan terdapat 73% Karyawan tidak puas dengan pekerjaan mereka (Sumber situs Jobstreet.com) karena karyawan tersebut tidak bekerja sesuai latar belakang pendidikan, tidak memiliki jenjang karir, dan tidak memiliki *work life balance* dan karakter atasan yang acuh. Agar dapat mengetahui terjadinya dari attrisi karyawan, perlu dilakukannya analisis pekerjaan. Ketika karyawan meninggalkan sebuah organisasi, mereka akan membawa segala pengetahuan yang sudah dikerjakan selama menjadi seorang karyawan di organisasi tersebut.

Agar perusahaan terus memiliki keunggulan yang lebih tinggi, maka tugas perusahaan harus meminimalkan pengurangan karyawan seperti kasus atrisi karyawan yang tinggi sehingga menyebabkan karyawan pindah ke perusahaan lain yang lebih layak. Adapun konsekuensi jika atrisi karyawan tidak ditangani, seperti biaya pergantian karyawan adalah: 20% dari gaji tahunan karyawan untuk posisi menengah (menghasilkan \$ 30.000 hingga \$ 50.000 setahun). Misalnya, biaya untuk mengganti karyawan yang dibayar \$ 30.000 adalah \$ 6.000. Sebuah studi oleh "*Center for American Progress*" menemukan bahwa biaya turnover sering kali diperkirakan 100% hingga 300% dari gaji pokok karyawan yang diganti, tergantung pada gaji dan peran seorang karyawan. Hal ini mengakibatkan kerugian finansial untuk menggantikan karyawan tersebut. Sekarang Prediksi Atrisi Karyawan menjadi masalah utama dalam Perusahaan.

Berdasarkan pemaparan tentang hubungan perusahaan dan atrisi karyawan maka penulis bermaksud untuk melakukan penerapan sistem prediksi awal di perusahaan. Dalam penelitian ini data yang akan digunakan adalah data perusahaan PT. Indorama Petrochemicals (PTIP) Indonesia beroperasi sejak tahun 2013, yang bergerak di bidang pengelolaan Polyster terbesar di Indonesia. PT. Indorama Petrochemicals yang berlokasi di Cilegon dan data yang didapat yaitu data *Performance Employee* tahun 2019. Perusahaan tersebut memiliki keinginan dalam meningkatkan kualitas sumber daya manusia. Dalam menentukan keputusan atrisi pada perusahaan tersebut, masih dilakukan secara manual. Proses prediksi atrisi karyawan bisa dilakukan secara manual dengan melihat pola nilai performa karyawan, tetapi sulit jika dilakukan untuk mengolah data karyawan yang lumayan banyak, karena memakan waktu lama dan rawan kesalahan prediksi. Penggunaan sistem untuk melakukan prediksi atrisi karyawan bisa digunakan untuk menangani masalah kerumitan dan ketepatan hasil prediksi, dengan menggunakan metode pendekatan yang cocok untuk prediksi dengan teknik Data Mining sehingga diharapkan dapat membantu perusahaan PT. Indorama Petrochemicals memutuskan atrisi.

Data Mining adalah proses analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya yang dipahami dan berguna bagi pemilik data tersebut. Data mining memiliki beberapa bagian yang diantaranya *association rule*, *prediction*, [5] *classification*, [6] *clustering* dan *regression*. Data Mining dibagi menjadi 2 bagian yaitu *Descriptive mining* dan *predictive*. *Descriptive mining* suatu proses untuk memperoleh pola (*correlation*, *trend*, *cluster*, *trajectory*, *anomaly*) untuk menyimpulkan hubungan di dalam data, sedangkan *predictive* suatu proses prediksi nilai dari atribut tertentu berdasarkan nilai dari atribut lainnya. Klasifikasi merupakan salah satu teknik data mining yang termasuk kedalam *predictive*, klasifikasi bertujuan untuk mengelompokkan objek menjadi kelas tertentu berdasarkan nilai atribut yang berkaitan dengan objek yang diamati tersebut [7].

Salah satu algoritma yang digunakan dalam klasifikasi adalah algoritma C 4.5.

Algoritma 4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau pengelompokan pada dataset. Keuntungan menerapkan algoritma C4.5 adalah dapat mengolah data numerik dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain [8]. Dasar dari algoritma C 4.5 adalah pembentukan pohon keputusan (*Decision Tree*). *Decision Tree* yaitu pohon keputusan yang mempunyai indikator atribut dan metode tersebut sederhana untuk dipahami namun dapat divisualisasikan dengan menggunakan rumus *Gain Ratio* [9]. Beberapa penelitian terdahulu telah banyak menggunakan algoritma C4.5 salah satunya adalah digunakan untuk prediksi kompetensi [10], prediksi kelulusan mahasiswa menggunakan metode *Decision Tree* [11].

Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [12]. Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda. Dengan mengambil *Entropy* dan *Gain* akan mendapatkan sebuah informasi output yang akurat.

Penelitian ini bertujuan untuk mengetahui kebenaran apakah karyawan di dalam perusahaan tersebut terkena atrisi/atau tidak dengan menerapkan teknik data mining dan prediksi menggunakan algoritma C4.5 dengan metode *Decision Tree* dan agar dataset yang telah dibagi menjadi data latih dan data uji dapat terprediksi dengan baik sehingga perusahaan dapat mendapatkan informasi yang benar.

2. METODE PENELITIAN

A. Perolehan Data

Metode pengumpulan data yang dilakukan pada penelitian ini melalui beberapa cara, diantaranya :

A. Studi Pustaka

Pada tahap ini dilakukan studi pustaka untuk mendapatkan informasi *Variable* pendukung yang berkaitan dengan penelitian seperti jurnal-jurnal terdahulu.

B. Wawancara

Pada tahap ini dilakukan studi pustaka untuk mendapatkan informasi *Variable* pendukung yang berkaitan dengan penelitian seperti jurnal-jurnal terdahulu.

B. Proses KDD

Knowledge Discovery in Database (KDD) merupakan proses pencarian pengetahuan

tersembunyi dalam sejumlah data besar yang mampu menghasilkan dan menyimpan suatu informasi. Serangkaian proses tahapan dalam proses KDD adalah sebagai berikut :

1. Pemilihan Data : melakukan pemilihan data, data yang akan dilakukan adalah atribut yang akan dipilih dalam proses selanjutnya untuk data mining, dan disimpan dalam database yang terpisah dari data operasional[6].
2. Pembersihan Data : pada tahap ini, data hasil seleksi akan dilakukan pemrosesan pendahuluan dan pembersihan data. Sebelum dilakukannya proses data mining, harus dilakukan pembersihan data agar tidak terjadinya duplikasi data[6].
3. Transformasi Data : merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Hasil data yang telah di bersihkan , disajikan Kembali sebagai dataset untuk pemrosesan data mining[6].
4. Data Mining : Proses ini merupakan proses mencari pola atau informasi yang sangat menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik Data mining yang digunakan adalah algoritma C4.5[6]
5. Interpretation/Evaluation : Pada tahap akhir ini dilakukan identifikasi terkait hasil uji evaluasi mengenai model yang sudah digunakan. Tahap ini menggunakan uji akurasi dengan menggunakan *Confusion Matrix*.

C. Algoritma C4.5

Tahapan algoritma C4.5 ini merupakan tahapan melakukan perhitungan untuk prediksi atrisi karyawan berdasarkan data *performance employee* yang telah melalui proses perubahan menjadi data uji dan akan di olah menjadi pohon keputusan atau *decission tree* yang dimana akan dibagi menjadi satu label yaitu “Tidak Atrisi” dan “Atrisi”. Secara umum *Algoritma C4.5* untuk membangun pohon keputusan adalah sebagai berikut:

1. Siapkan data sample, data sample disini untuk menghitung *Gain* dan *Entropy*.
2. Hitung Jumlah Karyawan yang terkena atrisi dan tidak atrisi.
3. Hitung Nilai *Entropy Total*
4. Hitung Nilai *Gain* untuk masing-masing atribut. Kemudian tentukan nilai *Gain* tertinggi. Atribut dengan nilai tertinggi maka atribut tersebut dijadikan sebagai akar.
5. Membuat Pohon Keputusan dari perhitungan *Gain* Tertinggi sebagai Node 1 dan seterusnya pada atribut lainnya.

D. Hitung Nilai Entropy

Pada proses pertama perhitungan algoritma C4.5 yaitu menentukan nilai entropy[13]. Untuk Langkah pertama yang harus dilakukan ialah menentukan *entropy* total, yaitu rumus *entropy* seperti persamaan berikut :

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

(1)

Keterangan :

S : Himpunan Kasus

A : Fitur

n : jumlah partisi S

pi : proporsi dari Si terhadap S

$$Entropy\ Total = ((17/20)*\log_2(17/20)*\log_2(3/20)) \\ = 0,609840305$$

E. Hitung Nilai Gain

Pada proses kedua perhitungan algoritma C4.5 yaitu menentukan nilai *gain* masing-masing atribut[13].Kemudian tentukan nilai *gain* tertinggi. Atribut dengan nilai tertinggi dijadikan sebagai akar. Sebagai contoh hitung nilai *gain* untuk atribut Kemampuan Pemecahan Masalah yaitu :

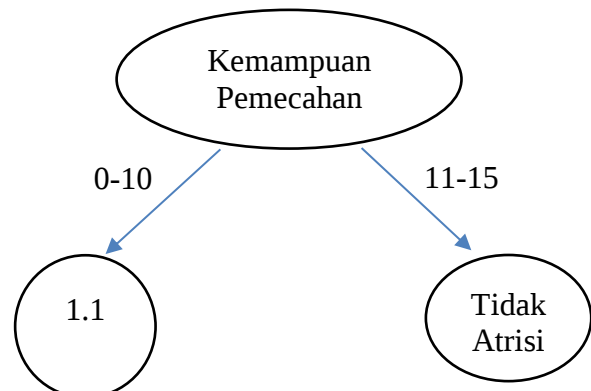
$$Gain(S, A) = Entropy\ Total - \sum_{i=0}^n \frac{S_i}{S} * Entropy(S_i) \quad (2) \\ Gain = 0,609840305 - ((6/20*0,918295834) + (14/20*0)) \\ = 0,334351555$$

Perhitungan nilai *entropy* dan *gain* memakai beberapa atribut sample agar dilakukan untuk mendapatkan nilai *gain* tertinggi yang akan dijadikan sebagai akar.

F. Pohon Keputusan

1. Pohon Keputusan Node 1 :

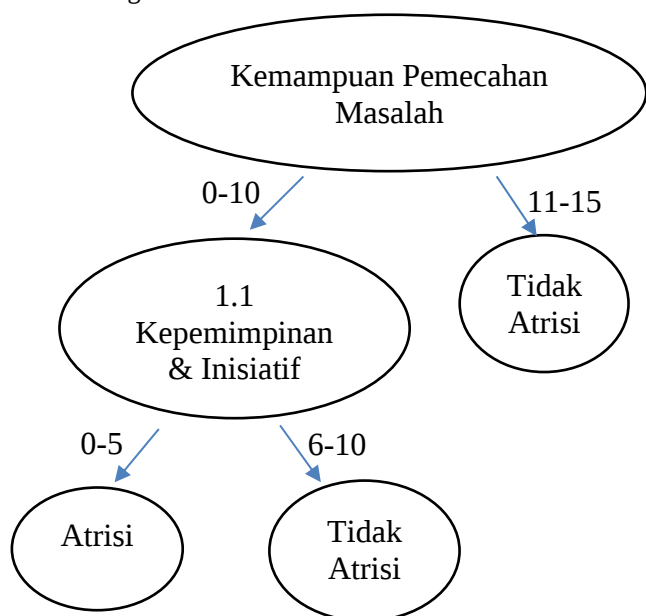
Pada proses ketiga yaitu membuat sebuah pohon keputusan, dimana hasil perhitungan *Gain* tertinggi akan digunakan sebagai Akar. Pada perhitungan di atas, nilai tertinggi adalah atribut Kemampuan Pemecahan Masalah, terdapat dua nilai dari atribut Kemampuan Pemecahan Masalah, yaitu 0-10 dan 11-15. Dari kedua nilai tersebut, nilai atribut 11-15 sudah mengklasifikasikan kasus menjadi 1, yaitu keputusannya tidak atrisi, sehingga tidak perlu dilakukan perhitungan lagi. Untuk nilai 0-10 masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara seperti pada Gambar 1 :



Gambar 1 Pohon Keputusan Hasil Perhitungan Node 1

2. Pohon Keputusan Node 1.1 :

Selanjutnya menghitung nilai dari *Atribut* Kepemimpinan masalah untuk dijadikan node percabangan dari nilai 0-10. Perhitungan dilakukan sama dengan seperti pada rumus entropy dan gain di atas, dan mendapatkan pohon keputusan seperti gambar 2:



Gambar 2 Pohon Keputusan Hasil Perhitungan Node 1.1

Berdasarkan Pohon Keputusan Node1.1 Pada atribut Kepemimpinan dan Inisiatif mendapatkan hasil Atrisi dan tidak atrisi

G. Confusion Matrix

Confusion Matrix merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada Gambar 3.

		True Values	
		True	False
Prediction	True	TP Correct Result	FP Unexpected result
	False	FN Missing Result	TN Correct absence of result

Gambar 3 Confusion Matrix

Pada gambar I dapat kita lihat terdapat 4 kombinasi berbeda dari nilai prediksi dan nilai actual. Ada 4 istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu *True Positive (TP)*, *True Negative (TN)*, *False Negative(FN)*, dan *False Positive (FP)*. Nilai *True Negative(TN)* merupakan jumlah data negative yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negative terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan

kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negative[14].

Jika sudah mendapatkan proses akurasi *Confusion Matrix* maka selanjutnya menghitung tiap nilai dengan *accuracy*, *precision*, *recall*, dan *F-1 Score*. Untuk Rumus dalam *accuracy*, *precision*, *recall* dan *F-1 Score* terdapat keterangan dibawah ini

$$ACCURACY = \frac{TN+TP}{TP+FP+FN+TN} \quad (1)$$

Accuracy : menggambarkan seberapa akurat model yang digunakan dalam klasifikasi dan prediksi.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Precision : menggambarkan akurasi antara data yang diminta dengan hasil prediksi diberikan oleh model.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Recall : menggambarkan keberhasilan model dalam menemukan Kembali informasi

$$F - 1 \text{ Score} = \frac{2 \cdot (Recall \cdot Precision)}{Recall + Precision} \quad (4)$$

F-1 Score: ini menggambarkan perbandingan rata-rata *precision* dan *recall* yang dibobotkan. *Accuracy* tepat kita gunakan sebagai acuan performansi algoritma jika dataset kita memiliki jumlah data *False Negatif* dan *False Positif* yang sangat mendekati (*symmetric*). Namun jika jumlahnya tidak mendekati, maka sebaiknya kita menggunakan *F1 Score* sebagai acuan. Ada penelitian yang emguji hasil akurasi menggunakan *Confusion Matrix* dilakukan oleh P. . Usha dan N. . Balaji, "Analysing employee attrition using machine learning.". Dengan melakukan pengujian model menggunakan *Confusion Matrix* sangat efektif dalam mengevaluasi data yang diprediksi[12]

3. PERANCANGAN SISTEM

Perancangan sistem ini membahas bagian isi dari tahapan yang dilakukan pada metode penelitian.

A. Perolehan Data

Perolehan data yang diambil dari PT.Indorama Petrochemicals, data didapat oleh wawancara pada pihak perusahaan tersebut. Data yang digunakan dalam penelitian ini yaitu data *PERFORMANCE EMPLOYEE* data yang di peroleh adalah data 2019 dengan jumlah 1010 record. Data tersebut dibagi menjadi dua data untuk data latih 80% dan data Uji 20%, Data *PERFORMANCE EMPLOYEE* ini berisi catatan performa karyawan dalam setiap kegiatannya yang berupa dokumen excel yang kemudian disimpan terpisah dari data lainnya. Data *PERFORMANCE EMPLOYEE* yang digunakan memiliki 29 variable, pada proses *selections* dilakukan juga pemilihan variable data didukung oleh studi Pustaka yang menguatkan pada seleksi variable yang akan digunakan untuk proses *prediction*. Data *Performance Employee* dapat dilihat pada table 1 .

Tabel 1. Perolehan Data

No	Nama	Umur	Kepemimpinan & Inisiatif
1	Randy Reza	41	9
2	Zulkarnain	49	8
3	Nefson Ramadhan	37	0
4	Regita Puspa Melati	33	0
5	Dewi Harmadiani	27	8

B. Preprocessing Data

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Pada penelitian ini terdapat 4 langkah dalam Preprocessing data diantaranya:

1. Selections Data

Selections Data dilakukan untuk memilih variable yang akan digunakan pada proses selanjutnya, pada penelitian ini, variable yang digunakan yaitu sebanyak 16 variable yaitu Variable Atrisi menjadi Label serta Variable lainnya adalah Nama Karyawan, Umur, Komunikasi, Kepemimpinan, Kemampuan, Tingkat Pekerjaan, Skala Kepuasan, Status Pernikahan, Pendatan Bulanan, Kualitas Kerja, Kerja Tim, Total Tahun Lama Bekerja dan Total Tahun Lama dengan Manajer Saat ini, digunakan dalam proses *Prediction* ditunjukkan pada contoh tabel 2.

Tabel 2. Data Hasil Selections

No	Nama	Atrisi	Komunikasi	Kepemimpinan & Inisiatif	Skala Kepuasan
1	Randy Reza	Tidak	10	9	4
2	Zulkarnain	Tidak	10	8	2
3	Nefson Ramadhan	Ya	14	2	3
4	Regita Puspa Melati	Tidak	14	0	3
5	Dewi Harmadiani	Tidak	10	8	2

2. Transformation Data

Tahap selanjutnya yaitu data transformasi yang berkaitan dengan proses konversi data yaitu mengubah data dari data sumber menjadi bentuk data yang disesuaikan dengan kebutuhan pemrosesan metode. Tujuan transformasi data adalah untuk mempermudah proses penggalan data. Data transformation terlihat di contoh tabel 3.

Tabel 3. Transformation data

No	Nama	Atrisi		Overtime
1	Randy Reza	Tidak		No
2	Zulkarnain	Tidak		No
3	Nefson Ramadhan	Ya		Yes
4	Regita Puspa Melati	Tidak		Yes
5	Dewi Harmadiani	Tidak		No

Tabel 4. Data setelah di Transformasi

No	Nama	Atrisi		Overtime
1	Randy Reza	0		0
2	Zulkarnain	0		0
3	Nefson Ramadhan	1		1
4	Regita Puspa Melati	0		1
5	Dewi Harmadiani	0		0

Dari tabel di atas, pada proses transformasi, label Atrisi menjadi numerik, begitupun untuk atribut overtime menjadi numerik yaitu 0-1, dan status pernikahan menjadi numerik yaitu 0-2.

3. Data Mining C4.5

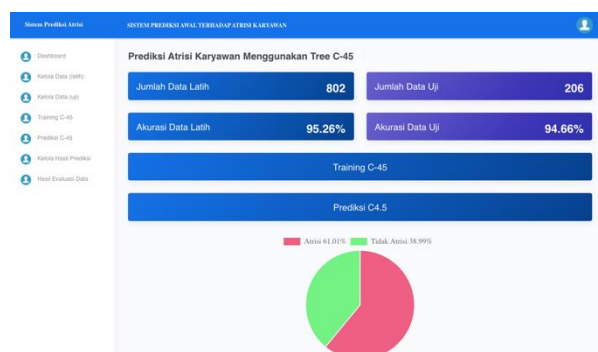
Tahapan algoritma C4.5 ini merupakan tahapan melakukan perhitungan untuk prediksi atrisi karyawan berdasarkan data *Performance Employee* yang telah melalui proses perubahan menjadi data uji dan akan di olah menjadi pohon keputusan atau *decision tree* yang dimana akan dibagi menjadi satu label yaitu "Tidak Atrisi" dan "Atrisi". Secara umum *Algoritma C4.5* untuk membangun pohon keputusan adalah sebagai berikut :

1. Siapkan data sample, data sample disini untuk menghitung *Gain* dan *Entropy* berjumlah data.
2. Hitung Jumlah Karyawan yang atrisi dan tidak atrisi.
3. Hitung Nilai *Entropy Total* dimana diketahui karyawan yang terkena atrisi sebanyak 6 dan tidak atrisi sebanyak 27.
4. Hitung Nilai *Gain* untuk masing-masing atribut. Kemudian tentukan nilai *Gain* tertinggi. Atribut dengan nilai tertinggi maka atribut tersebut dijadikan sebagai akar.
5. Membuat Pohon Keputusan dari perhitungan *Gain* tertinggi sebagai Node 1 dan seterusnya pada atribut lainnya.

4. HASIL DAN PEMBAHASAN

A. Hasil

Pada tahap ini dilakukannya implementasi dari perancangan sistem yaitu menampilkan jumlah data karyawan yang terkena atrisi dan tidak di atrisi beserta akurasi yang diperoleh.



Gambar 4. Visualisasi Data

Dari hasil visualisasi diatas, dapat dilihat bahwa tingkat tertinggi yang didapat oleh sistem prediksi karyawan, yaitu karyawan yang terkena tidak atrisi lebih besar dari pada karyawan yang terkena atrisi.

B. Evaluasi Hasil / Pengujian Akurasi

Pengujian akurasi merupakan tahapan untuk menguji hasil akurasi yang didapatkan pada model yang digunakan. Pengujian tersebut menggunakan metode *confusion matrix* yang dapat dilihat pada tabel 5.

Tabel 5. Pengujian Akurasi

True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)
43	10	1	152

Dari table diatas menunjukkan variable-variable dari hasil uji evaluasi *Confusion Matrix* dengan algoritma C4.5, meliputi :

- True Positive (TP) : Jumlah yang ada pada atribut TP menandakan data karyawan yang tidak atrisi dan benar tidak di atrisi
- False Positive (FP) : Jumlah yang ada pada atribut FP menandakan data karyawan yang dianggap tidak atrisi namun di prediksi salah ternyata data karyawan Atrisi
- True Negative (TN) : Jumlah yang ada pada atribut TN menandakan data karyawan yang atrisi dan benar di atrisi
- False Negative (FN) : Jumlah yang ada pada atribut FN menandakan data karyawan yang dianggap atrisi namun prediksi salah ternyata data karyawan tidak atrisi.

Lalu nilai yang telah didapatkan dihitung dengan rumus dibawah ini,

$$ACCURACY = \frac{TN+TP}{TP+FP+FN+TN} \times 100\%$$

Hasil :

$$ACCURACY = \frac{152+43}{43+10+1+152} \times 100\% \quad (1)$$

$$= 94,66\%$$

Hasil menunjukan bahwa akurasi dari model yang digunakan pada penelitian ini dengan pengujian *confusion matrix* sebesar 94,66%. Hasil tersebut menandakan bahwa model yang digunakan dapat memprediksi data dengan baik.

5. KESIMPULAN

Berdasarkan hasil penelitian tugas akhir dapat disimpulkan bahwa karyawan yang ada di PT.Indorama Petrochemicals lebih banyak terkena atrisi. Dalam penelitian ini 80% dari data digunakan sebagai latih dan 20% dari data digunakan sebagai data uji dan hasil akurasi yang diperoleh cukup tinggi dalam memprediksi karyawan yang benar terkena atrisi, namun ada beberapa karyawan yang seharusnya tidak di atrisi, namun mereka terkena atrisi. Kesimpulan dari penelitian ini adalah algoritma C4.5 merupakan algoritma yang dapat

memprediksi kebenaran karyawan lama yang terkena atrisi di PT. Indorama Petrochemicals yang mempunyai masalah dalam proses atrisi manual dan sulit di prediksi, hasil yang diperoleh menggunakan metode *Confussion Matrix* sebesar 0,9466 apabila dihitung dengan persen hasil yang didapatkan adalah 94,6%, dari hasil evaluasi tersebut bahwa algoritma C4.5 memiliki peforma yang cukup baik dalam memprediksi data.

DAFTAR PUSTAKA

- [1] G. E. Franco Peña *et al.*, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI," *Ayan*, vol. 8, no. 2, hal. 2019, 2010, [Daring]. Tersedia pada: https://barnard.edu/sites/default/files/inline/student_user_guide_for_spss.pdf%0Ahttp://www.ibm.com/support%0Ahttp://www.spss.com/sites/dm-book/legacy/ProgDataMgmt_SPSS17.pdf%0Ahttps://www.neps-data.de/Portals/0/WorkingPapers/WP_XLV.pdf%0Ahttp://www2.psy.
- [2] J. Ho, S.-P. Loke, dan A. Downe, "Employee attrition in the Malaysian service industry: Push and pull factors," *IUP J. Organ. Behav.*, vol. 9, no. 1&2, hal. 16–31, 2010.
- [3] K. Sehgal, H. Bindra, A. Batra, dan R. Jain, *Prediction of Employee Attrition Using GWO and PSO Optimised Models of C5.0 Used with Association Rules and Analysis of Optimisers*, vol. 74. Springer Singapore, 2019.
- [4] R. Punnoose dan P. Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms," *Int. J. Adv. Res. Artif. Intell.*, vol. 5, no. 9, hal. 22–26, 2016, doi: 10.14569/ijarai.2016.050904.
- [5] A. Wibowo, D. Manongga, dan H. D. Purnomo, "The Utilization of Naive Bayes and C.45 in Predicting The Timeliness of Students' Graduation," *Sci. J. Informatics*, vol. 7, no. 1, hal. 99–112, 2020, doi: 10.15294/sji.v7i1.24241.
- [6] H. KAWANO, "Knowledge Discovery and Data Mining," *J. Japan Soc. Fuzzy Theory Syst.*, vol. 9, no. 6, hal. 851–860, 1997, doi: 10.3156/jfuzzy.9.6.851.
- [7] S. D. GHEWARE, A. S. KEJKAR, dan S. M. TONDARE, "Data Mining Task Tools Techniques and Applications," *Ijarccce*, vol. 3, no. 10, hal. 8095–8098, 2014, doi: 10.17148/ijarccce.2014.31003.
- [8] F. F. Harryanto dan S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *J. Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, hal. 95–103, 2017, [Daring]. Tersedia pada: <http://jurnal.mdp.ac.id/index.php/jatisi/article/view/71>.
- [9] R. S. Shankar, J. Rajanikanth, V. V. Sivaramaraju, dan K. Vssr Murthy, "PREDICTION of EMPLOYEE ATTRITION

- USING DATAMINING,” *2018 IEEE Int. Conf. Syst. Comput. Autom. Networking, ICSCA 2018*, hal. 1–8, 2018, doi: 10.1109/ICSCAN.2018.8541242.
- [10] A. S. Sunge, “Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 (Studi Kasus : PT Hankook Tire Indonesia),” *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, hal. 23–24, 2018.
- [11] SUSI MASHLAHAH, *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*. 2013.
- [12] P. . Usha dan N. . Balaji, “Analysing employee attrition using machine learning,” hal. 277–282, 2011.
- [13] Y. Kustiyahningsih dan E. Rahmanita, “Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5. untuk Penjurusan SMA,” *J. Semantec*, vol. 5, no. 2, hal. 101–108, 2016.
- [14] X. Deng, Q. Liu, Y. Deng, dan S. Mahadevan, “An improved method to construct basic probability assignment based on the confusion matrix for classification problem,” *Inf. Sci. (Ny)*., vol. 340–341, hal. 250–261, 2016, doi: 10.1016/j.ins.2016.01.033.