

## Klasifikasi Karakteristik *Turnover* Karyawan Menggunakan Algoritma C4.5

Rany Andini<sup>1</sup>, Dinna Yunika Hardiyanti<sup>2\*</sup>

<sup>1</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya

<sup>2</sup>Komputerisasi Akuntansi, Fakultas Ilmu Komputer, Universitas Sriwijaya

\*email: [dinna.yunika@unsri.ac.id](mailto:dinna.yunika@unsri.ac.id)

DOI: <https://doi.org/10.31603/komtika.v8i2.12504>

Received: 20-10-2024, Revised: 09-11-2024, Accepted: 11-11-2024

### ABSTRACT

*Turnover is an act or behavior of leaving or leaving employees from an organization or company. Turnover can have a negative impact on the company such as delayed projects and targets, team dissolution, shortage of human resources (HR), difficulty recruiting to find candidates with various criteria in a short time, to disruption of workplace productivity. The research objective is to determine the main factors and characteristics that have a significant effect on employee turnover. The research was conducted by managing employee turnover data using classification data mining techniques and applying the C4.5 algorithm. The results of the model produce decision trees and rules in determining employees who will turnover. The results showed that there are 8 attributes that significantly affect employee turnover, namely Age, Monthly Billing, Monthly Rate, OverTime, YearsAtCompany, YearsInCurrentRole, YearsLastPromotion, and DistanceFromHome. The model produced in the study has an accuracy of 81.63%.*

**Keywords:** *Datamining, Decission tree, Clasifikasi, C4.5 algorithm, Turnover*

### ABSTRAK

*Turnover merupakan suatu Tindakan atau perilaku meninggalkan atau keluar karyawan dari sebuah organisasi atau perusahaan. Turnover dapat membawa dampak negatif bagi perusahaan seperti proyek dan target yang tertunda, pembubaran tim, kekurangan sumber daya manusia (SDM), kesulitan melakukan perekrutan untuk mencari kandidat dengan berbagai kriteria dalam waktu singkat, hingga gangguan produktivitas tempat kerja. Tujuan penelitian yaitu mengetahui faktor utama dan karakteristik yang berpengaruh signifikan dalam turnover karyawan. penelitian dilakukan dengan mengelola data turnover karyawan menggunakan teknik data mining klasifikasi dan menerapkan algoritma C4.5. Hasil modelling menghasilkan pohon keputusan dan rule dalam menentukan karyawan yang akan melakukan turnover. Hasil penelitian menunjukkan bahwa terdapat 8 atribut yang berpengaruh secara signifikan terhadap turnover karyawan yaitu Age, Monthly Billing, Monthly Rate, OverTime, YearsAtCompany, YearsInCurrentRole, YearsLastPromotion, dan DistanceFromHome. Model yang dihasilkan dalam penelitian memiliki akurasi sebesar 81,63%.*

**Kata kunci:** *Datamining, Decission tree, Klasifikasi, Algoritma C4.5, Turnover*

### PENDAHULUAN

Sumber daya manusia (SDM) merupakan elemen penting dalam keberhasilan setiap perusahaan. Dalam era globalisasi dan perkembangan yang pesat, kemampuan perusahaan untuk mengelola SDM secara efektif menjadi kunci dalam menjaga daya saing dan mencapai tujuan jangka panjang [1]. Pengelolaan SDM tidak hanya berfokus pada perekrutan dan pengembangan karyawan, tetapi juga pada upaya untuk mempertahankan karyawan yang berkualitas agar tetap bekerja dalam perusahaan. Salah satu indikator keberhasilan pengelolaan SDM adalah rendahnya tingkat *turnover* karyawan, yang sering menjadi masalah signifikan bagi perusahaan di berbagai sektor [2].

*Turnover* merupakan suatu tindakan atau perilaku meninggalkan atau keluar karyawan dari sebuah organisasi atau perusahaan. *Turnover* atau pemberhentian merupakan pemutusan hubungan kerja karyawan dari sebuah organisasi atau perusahaan, dengan adanya pemberhentian tersebut maka berakhir hubungan keterikatan kerja karyawan dengan perusahaan [3]. *Turnover* menjadi salah satu tantangan dan pertimbangan tersendiri bagi manajerial perusahaan, karena mengindikasikan ketidakpuasan pekerjaan [4]. *Turnover* merupakan faktor yang mempengaruhi perilaku organisasi, dimana dengan dapat mengendalikan dan mengurangi hal tersebut secara efektif meningkatkan kinerja karyawan, penghematan waktu dan biaya perekrutan, serta pencapaian target dan tujuan yang tepat waktu dan sesuai [5]

*Turnover* dapat membawa dampak negatif bagi perusahaan seperti proyek dan target yang tertunda, pembubaran tim, kekurangan SDM, kesulitan melakukan perekrutan untuk mencari kandidat dengan berbagai kriteria dalam waktu singkat, hingga gangguan produktivitas tempat kerja [5]. Tingginya tingkat *turnover* juga akan mempengaruhi motivasi dan semangat kerja karyawan yang tetap ada di perusahaan, karyawan yang sebelumnya tidak sedang berusaha mencari pekerjaan lain akan mulai mencari kerja yang kemudian memungkinkan akan melakukan *turnover* [6].

Sebuah survei yang dilakukan oleh JobsDB Indonesia dengan melibatkan 2.324 responden dari berbagai bidang pekerjaan yang berbeda di Indonesia menunjukkan hasil bahwa sebanyak 80% responden memiliki keinginan untuk berganti pekerjaan dalam 12 bulan ke depan. Praditya dalam [7]. Keluarnya karyawan pada perusahaan berarti terdapat posisi tertentu yang kosong dan harus segera diisi, sehingga perusahaan harus mengeluarkan biaya untuk proses pelatihan dan perekrutan karyawan hingga mendapat karyawan yang sesuai. Diketahui pada sektor industri di Amerika Serikat kerugian akibat *turnover* karyawan menyebabkan perusahaan harus mengeluarkan biaya sebesar \$60,000 untuk perekrutan karyawan baru, padahal seharusnya hanya perlu menghabiskan biaya sebesar \$40,000 untuk menggaji karyawan yang ada, dan setiap tahunnya karyawan yang melakukan *turnover* terdapat sekitar 16,8% [8]. Berdasarkan dampak-dampak tersebut, penting bagi perusahaan untuk dapat memahami faktor-faktor yang mempengaruhi keputusan karyawan apakah untuk tetap bekerja di perusahaan atau meninggalkan perusahaan.

Untuk mengatasi masalah tersebut, perlu dilakukan analisis terhadap data *turnover* pada perusahaan untuk mengidentifikasi faktor-faktor utama pendorong *turnover* agar dapat memprediksi karyawan yang akan keluar. Dengan demikian, karyawan yang terprediksi akan keluar dapat diberikan perlakuan tertentu dan diharapkan dapat mengendalikan kejadian *turnover* dan mengurangi dampak negatif yang akan terjadi. Salah satu cara yang dapat digunakan untuk mengelola (mengendalikan dan mengurangi) *turnover* karyawan adalah dengan memanfaatkan fungsi *data mining* yaitu klasifikasi dan prediksi [5]. *Data mining* merupakan proses mengekstrak informasi penting dari data dalam jumlah besar yang mungkin tersembunyi [9]. Dalam pengelolaan *turnover*, data mining dapat digunakan untuk mengidentifikasi karakteristik dalam data *turnover* karyawan yang dapat mempengaruhi keputusan karyawan untuk dapat tetap bekerja atau keluar.

Terdapat beberapa metode yang dioperasikan pada *data mining*, adapun metode tersebut yaitu estimasi, prediksi, klusterisasi, klasifikasi, dan asosiasi. Pada penelitian ini metode yang digunakan yaitu klasifikasi. Klasifikasi merupakan metode *data mining* yang digunakan untuk

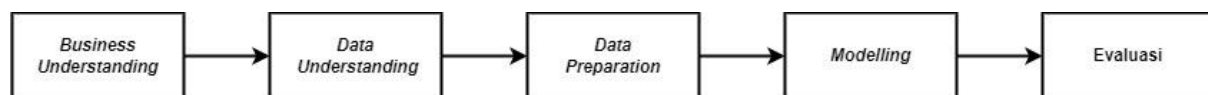
mengekstrak dan memprediksi label kategori pada setiap data. Klasifikasi dapat diartikan juga sebagai proses menemukan model yang mampu membedakan kelas data, tujuannya adalah supaya model tersebut dapat digunakan untuk memprediksi kelas dari objek yang belum diketahui kelasnya [10]. Pada penelitian ini metode klasifikasi digunakan untuk memprediksi karyawan melakukan *turnover* atau tidak dengan mengidentifikasi pola dari karakteristik karyawan yang sudah diketahui. Melalui prediksi yang akurat, perusahaan dapat mengambil langkah proaktif untuk mengelola (mengendalikan dan mengurangi) *turnover* yang tidak diinginkan.

Salah satu algoritma yang dapat digunakan dalam klasifikasi adalah algoritma C4.5. Secara konsep, algoritma C4.5 berfungsi untuk mengubah data menjadi pohon keputusan serta menghasilkan aturan-aturan keputusan (*rule*) [11]. Model ini dipilih karena algoritma C4.5 umumnya memiliki tingkat akurasi yang tinggi. Algoritma C4.5 juga dapat menghasilkan model yang mudah dipahami, dapat menangani dataset yang kompleks, dan mengidentifikasi fitur-fitur penting yang ada pada *dataset*[9].

Beberapa penelitian sebelumnya telah melakukan klasifikasi dengan menerapkan algoritma C4.5 diantaranya yaitu penelitian yang dilakukan Ramadhani[12] untuk mengklasifikasi mahasiswa yang berpotensi *Drop out* menggunakan algoritma C4.5. Hasil evaluasi penelitian menghasilkan akurasi sebesar 98,50%. Penelitian terdahulu lainnya [13] menggunakan lima algoritma Decision Tree yaitu algoritma C4.5, CART, Random Forest, Gradient Boost dan Adaboost untuk memprediksi potensi atrisi, hasil penelitian menunjukkan bahwa algoritma C4,5 dan Random Forest memiliki hasil klasifikasi yang lebih baik dengan nilai akurasi sebesar 82,35%, presisi 86,76%, *recall* 83,35%, dan *f1-score* 81,57%. Berdasarkan penelitian terdahulu, penelitian ini akan menerapkan algoritma C4 dan akan mengelola data *turnover* karyawan untuk mendapatkan pengetahuan yang dapat dimanfaatkan oleh manajemen perusahaan dalam mempertahankan karyawannya. Dalam hal ini diharapkan akan diketahui faktor utama dan karakteristik karyawan yang melakukan *turnover*.

## METODE

Penelitian ini menerapkan metode CRIPS-DM (*Cross Industry Standard for Data Mining*) sebagai tahapan penelitian. Tahapan CRIPS-DM yang dilakukan dalam penelitian ini terdapat pada gambar 1.



Gambar 1. Tahapan Penelitian

### ***Business Understanding***

*Business understanding* merupakan tahapan pertama yang dilakukan dalam penelitian ini, tahapan ini bertujuan untuk meningkatkan pemahaman terkait masalah, tujuan, dan hasil dari penerapan data mining. Dilakukan analisis masalah untuk menentukan data yang sesuai. Permasalahan dalam penelitian ini sendiri telah dijelaskan pada latar belakang bahwa diketahui *turnover* yang tinggi pada perusahaan dapat menyebabkan berbagai dampak negatif

bagi perusahaan, selain itu *turnover* yang tinggi juga merupakan indikasi adanya permasalahan dalam perusahaan. Salah satu strategi untuk dapat mengelola (mengendalikan dan mengurangi) *turnover* adalah dengan mengidentifikasi berbagai faktor yang menyebabkan *turnover* karyawan.

### Data Understanding

Pada penelitian ini objek yang diteliti yaitu data karyawan yang melakukan *turnover*. Pada tahap ini dilakukan pengumpulan data dari *website open data*. Data yang digunakan merupakan data sekunder yang diambil dari *website Kaggle[14]* yang berjudul *employee turnover dataset*. *Dataset* mencakup berbagai informasi yang berkaitan dengan karyawan dan merupakan faktor yang mempengaruhi *turnover* karyawan. *Dataset* terdiri dari 1470 data dengan 23 atribut dan memiliki 1 label yaitu *turnover* yang memiliki 2 kelas yaitu “Yes” dan “No”. Pada tabel 1 dapat dilihat atribut yang terdapat pada *dataset* dan keterangan atribut tersebut.

Tabel 1. Atribut pada *Dataset*

Atribut	Keterangan
<i>EmployeeId</i>	ID Karyawan
<i>Age</i>	Usia karyawan
<i>Gender</i>	Apakah karyawan laki-laki atau perempuan
<i>MaritalStatus</i>	Status pernikahan karyawan ( <i>Single, Married, Divorce</i> )
<i>Turnover</i>	Variabel yang mengindikasikan apakah karyawan tetap atau keluar dari perusahaan (Yes, No)
<i>Travellin</i>	Menunjukkan seberapa sering karyawan bepergian untuk pekerjaan ( <i>No, Sometimes, Mostly</i> )
<i>Vertical</i>	Divisi atau departemen tempat karyawan bekerja
<i>EducationField</i>	Bidang studi karyawan
<i>Role</i>	Jenis pekerjaan atau peran spesifik
<i>MonthlyBilling</i>	Tagihan bulanan
<i>MonthlyRate</i>	Gaji bulanan
<i>Work Experience</i>	Jumlah tahun pengalaman kerja
<i>OverTime</i>	Apakah karyawan sering lembur
<i>PercentSalaryHike</i>	Kenaikan gaji tahunan
<i>LastRating</i>	Penilaian Terakhir Terhadap karyawan
<i>RelationshipSatisfaction</i>	Kepuasan Hubungan Kerja
<i>Hours</i>	Jam Kerja
<i>TrainingTimesLastYear</i>	Seberapa sering karyawan mendapatkan pelatihan
<i>YearsAtCompany</i>	Lama bekerja di perusahaan
<i>YearsInCurrentRole</i>	Lama bekerja di posisi saat ini
<i>YearsSinceLastPromotion</i>	Lama bekerja sejak promosi terakhir
<i>YearsWithCurrentManager</i>	Lama bekerja dengan manajer sekarang
<i>DistanceFromHome</i>	Jarak antara rumah dan tempat kerja

### Data Preparation

Pada tahap *data preparation* dilakukan proses perbaikan pada *dataset* sebelum data digunakan pada tahap pemodelan. Pada tahap ini dilakukan *data cleaning* untuk membersihkan dan mengecek apakah terdapat *missing value* pada *dataset* yang akan digunakan. Hasil dari *data cleaning* menunjukkan bahwa tidak ada *missing value* pada *dataset*. Gambar 2 menunjukkan hasil *data cleaning*.

Name	Type	Missing	Min	Max	Average
PercentSalaryHike	Integer	0	11	25	15.122
Last Rating	Integer	0	3	4	3.163
RelationshipSatisfaction	Integer	0	1	4	2.728
TrainingTimesLastYear	Integer	0	0	6	2.769
YearsAtCompany	Integer	0	0	33	6.772
YearsInCurrentRole	Integer	0	0	17	4.067
YearsSinceLastPromotion	Integer	0	0	15	2.095
YearsWithCurrentManager	Integer	0	0	17	3.946
DistanceFromHome	Integer	0	1	29	9.278

Gambar 2. Data Cleaning

Setelah proses *data cleaning*, proses yang dilakukan yaitu penghapusan atribut yang tidak diperlukan. Dalam proses *modelling* menggunakan algoritma C4.5 tidak semua atribut yang ada pada *dataset* harus digunakan dalam membangun model. Jika terdapat atribut atau kolom yang sekiranya tidak memiliki pengaruh sebaiknya dihapus. Pada penelitian ini dilakukan penghapusan atribut yang tidak diperlukan yaitu atribut *employeeId* dan *hours*, atribut *employeeid* merupakan nomor id karyawan yang tidak akan mempengaruhi terhadap *turnover* karyawan dan untuk atribut *hours* juga tidak berpengaruh karena semua karyawan mempunyai jam kerja yang sama. *Dataset* yang telah dilakukan *preparation* dapat dilihat pada tabel 3.

Tabel 2. Dataset setelah preparation

<i>Age</i>	<i>Gender</i>	<i>Marital Status</i>	...	<i>YearsWith Current Manager</i>	<i>Distance FromHome</i>
29	M	<i>Divorced</i>	...	2	2
23	M	<i>Single</i>	...	0	6
36	M	<i>Single</i>	...	0	10
30	M	<i>Divorced</i>	...	9	12
29	M	<i>Single</i>	...	2	14
30	M	<i>Divorced</i>	...	2	2
...	...	<i>Single</i>	...	...	...
25	M	<i>Married</i>	...	2	2
38	F	<i>Single</i>	...	3	5
31	M	<i>Divorced</i>	...	0	1

Pada tabel 3 merupakan *dataset* yang telah dilakukan *preparation* dengan menghilangkan beberapa atribut. *Dataset* ini yang akan digunakan untuk proses selanjutnya dalam penelitian yaitu *modelling* dengan menerapkan C4.5 yang dilakukan di Rapidminer.

Proses klasifikasi pada penelitian ini dilakukan dengan menerapkan algoritma C4.5. Algoritma C4.5 merupakan metode yang dapat digunakan untuk membentuk sebuah pohon keputusan (*Decission Tree*). Algoritma ini merupakan hasil pengembangan dari algoritma ID3 yang juga berfungsi membentuk *Decission Tree* (Febriani & Sulistiani, 2021). Algoritma C4.5 juga dianggap sangat membantu dalam klasifikasi data, algoritma ini memudahkan pengguna

untuk mengamati informasi yang terklasifikasi dengan jelas, baik dalam bentuk pohon keputusan (*rule*) maupun ke pada aturan *if then* (Kusrini & Hidayat, 2024).

Langkah awal dalam membangun sebuah pohon keputusan adalah dengan menentukan atribut atau variabel yang akan dijadikan sebagai akar dari pohon keputusan tersebut. Atribut utama ini ditentukan berdasarkan perhitungan total *entropy*, *entropy* setiap atribut, dan nilai *gain*. Rumus *entropy* pada C4.5 adalah sebagai berikut: [15]

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

Keterangan :

S = Himpunan kasus

N = jumlah partisi s

$p_i$  = Proporsi  $S_i$  terhadap s

Setelah proses menghitung nilai *entropy total* dan *entropy* pada setiap atribut, Langkah selanjutnya yaitu menghitung nilai *gain* pada tiap – tiap atribut. *Gain* merupakan ukuran efektivitas data yang diklasifikasikan dari atribut, rumus *gain* pada algoritma C4.5 adalah sebagai berikut: Widayu dalam [15]

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

Keterangan :

S = Himpunan kasus

A = Atribut atau fitur

$|S_i|$  = Proporsi  $S_i$  terhadap S

$|S|$  = jumlah kasus dalam S

N = jumlah partisi tribute A

## Evaluasi

Evaluasi kinerja model dalam penelitian ini dapat dihitung dengan *confusion matrix* yaitu sebuah visualisasi dalam melakukan evaluasi dari kinerja suatu model klasifikasi. *Confusion matrix* menyajikan informasi mengenai kelas sebenarnya dan kelas prediksi yang dihasilkan model. Informasi ini disajikan dalam bentuk matriks, dimana pada setiap kolom mewakili kelas yang diprediksi dan pada setiap baris menunjukkan kejadian kelas sebenarnya [16].

Terdapat beberapa istilah pada *confusion matrix* yang perlu diketahui, yaitu:

- True Positive* (TP): jika data bernilai *class positif* dan pada hasil klasifikasi data juga bernilai *class positif* maka dikatakan *true positive*.
- True Negative* (TN): jika data bernilai *class negative* dan pada hasil klasifikasi data juga bernilai *class negative* maka dikatakan *true negative*.
- False Positive* (FP): jika data bernilai *class negative* tetapi pada hasil klasifikasi data bernilai *class positif* maka dikatakan *false negative*.

- d. *False Negative* (FN): jika data bernilai *class positif* tetapi pada hasil klasifikasi data bernilai *class negative* maka dikatakan *false negative*.

Berikut merupakan perhitungan untuk berbagai ukuran evaluasi model:

- a. *Accuracy* menunjukkan hubungan anatar nilai prediksi dengan nilai aktual

$$Accuracy = \frac{TP+TN}{Total\ Data} \times 100\%$$

- b. *Precision* merupakan tingkat ketepatan terhadap hasil yang diharapkan dan hasil sistem.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

- c. *Recall* menunjukkan seberapa baik model berhasil untuk mendapatkan informasi kembali

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

- d. *Specificity* merupakan keefektifan model dalam melakukan klasifikasi untuk kelas *negative*.

$$Specificity = \frac{TN}{TN+FP} \times 100\%$$

- e. *F-Measure* atau *F1-score* menggambarkan keseimbangan pada nilai *recall* dan *precision*.

$$F\text{-Measure} = 2 \times \frac{precision \times recall}{precision+recall}$$

## HASIL DAN PEMBAHASAN

### *Business Understanding*

Berdasarkan analisis, *turnover* yang tinggi dapat berdampak negatif pada perusahaan seperti meningkatnya biaya perekrutan karyawan dan menurunnya produktivitas karyawan. Dengan demikian, penelitian ini bertujuan untuk mengetahui faktor-faktor penting yang mempengaruhi *turnover* karyawan. Pengetahuan terhadap faktor-faktor yang menyebabkan *turnover* karyawan dapat membantu perusahaan dalam mengidentifikasi karyawan yang berpotensi meninggalkan perusahaan.

### *Data Understanding*

Pada penelitian ini, dilakukan pengumpulan data dimana *dataset* yang digunakan terdiri dari 1.470 data dengan 23 atribut. Yang berisi informasi informasi yang berkaitan dengan karyawan. Tabel 4 merupakan *dataset* *turnover* karyawan.

Tabel 3. *Dataset Turnover* Karyawan

<i>EmployeeId</i>	<i>Age</i>	<i>Gender</i>	...	<i>YearsWith Curent Manager</i>	<i>Distance FromHome</i>
63	29	M	...	2	2
723	23	M	...	0	6
1297	36	M	...	0	10
51	30	M	...	9	12
1498	29	M	...	2	14
81	30	M	...	2	2
...	...	...	...	...	...
1395	25	M	...	2	2
919	38	F	...	3	5
97	31	M	...	0	1

### Data Preparation

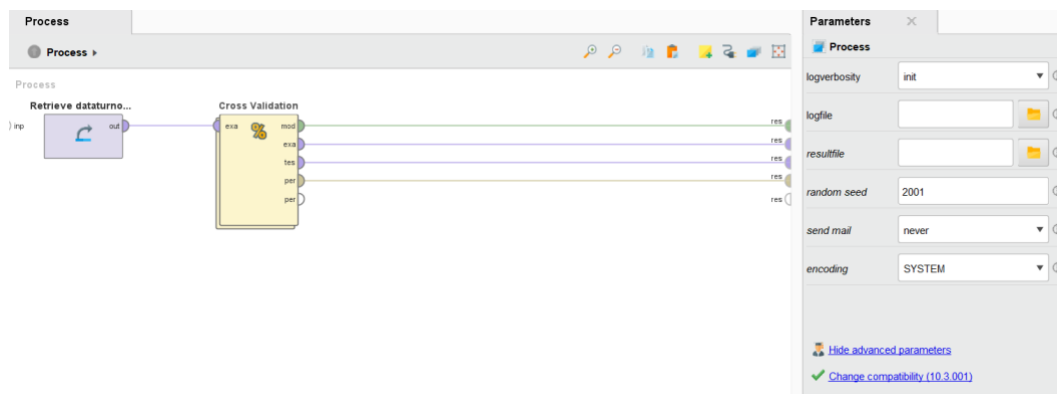
Pada proses ini dilakukan *cleaning data* dan penghapusan atribut yang tidak diperlukan dalam proses *modelling*. Tabel 5 merupakan Hasil data yang telah dilakukan *preparation*.

Tabel 4. Dataset setelah *preparation*

Age	Gender	Marital Status	...	YearsWith Cureent Manager	Distance FromHome
29	M	Divorced	...	2	2
23	M	Single	...	0	6
36	M	Single	...	0	10
30	M	Divorced	...	9	12
29	M	Single	...	2	14
30	M	Divorced	...	2	2
...	...	Single	...	...	...
25	M	Married	...	2	2
38	F	Single	...	3	5
31	M	Divorced	...	0	1

### Modelling

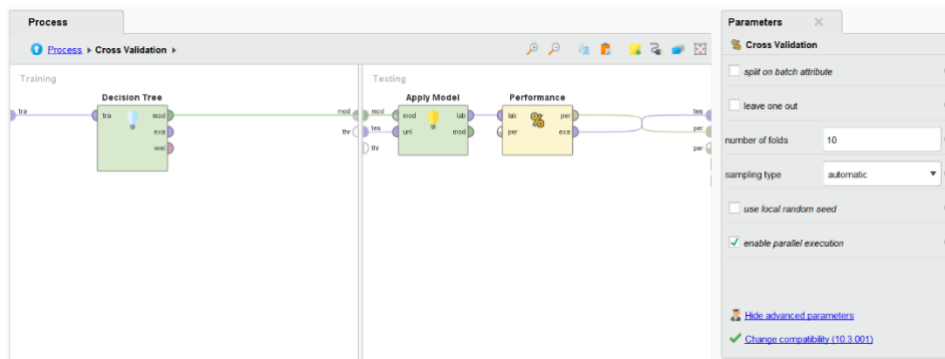
Proses *modelling* dilakukan dengan menggunakan aplikasi rapidminer dapat dilihat pada gambar 3 dan gambar 4.



Gambar 3. Proses *modeling*

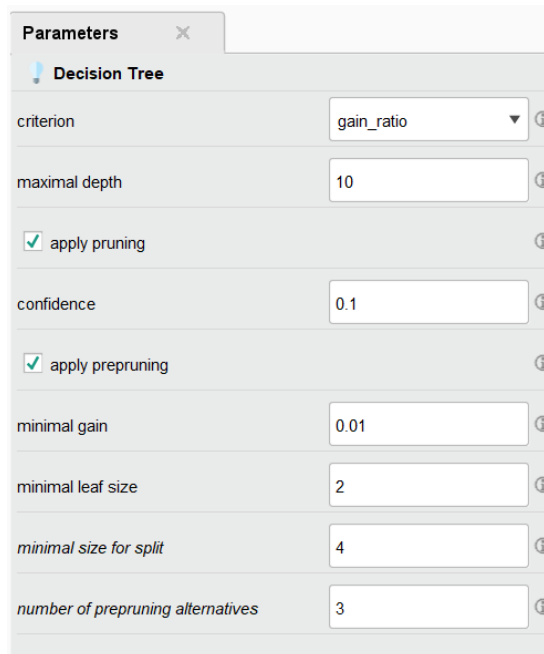
Pada gambar 3 terdapat operator *Retrive* dan operator *Cross Validation*. Operator *Retrive* merupakan operator yang digunakan untuk melakukan proses mengambil *dataset* yang akan digunakan dan membawa data tersebut pada rapidminer. *Cross validation* digunakan dalam pembagian data dimana operator *Cross Validation* secara otomatis akan membagi data menjadi data *training* dan data *testing*. *Cross validation* menerapkan *K-Fold Validation* dengan jumlah *fold* 10, sehingga perbandingan pada data *training* dan data *testing* sebesar 9 : 1 atau 90% : 10%.





Gambar 4. Sub-proses modelling

Pada gambar 4 merupakan *sub-proses* dari operator *Cross Validation* yang terdiri dari *training* dan *testing*. Dalam proses *training* terdapat operator *Decision Tree* menerapkan algoritma C4.5. Pada *sub-proses training* ini dibentuk suatu model berdasarkan data *training* dan pada *sub-proses testing*, terdapat *apply model* dan *performance*, dengan menggunakan data *testing* model yang dibuat diuji kinerja dan akurasinya.



Gambar 5. Parameter C4.5

Untuk mendapatkan model yang baik dilakukan penyesuaian parameter dari operator *Decision Tree*. Pada gambar 5 merupakan parameter algoritma C4.5 yang diterapkan pada model ini. Pada bagian *parameter criterion*, digunakan *gain\_ratio* sebagai ukuran pemilihan atribut yang menjadi *node*. Adapun beberapa parameter yang mempengaruhi tingkat akurasi dari model yang dihasilkan menggunakan algoritma C4.5 adalah *maximal depth*, *pruning*, dan *prepruning*

Pada penelitian ini, *parameter maximal depth* yang digunakan ditetapkan sebesar 10, Selain itu, *apply pruning* juga diterapkan dimana nilai *confidence* diatur sebesar 0.1, dengan menerapkan *apply pruning* tersebut akan dilakukan pemangkasan cabang jika nilai pada



### Tree

```

Age > 18.500
| MonthlyBilling > 19785: No {No=15, Yes=0}
| MonthlyBilling ≤ 19785
| | MonthlyRate > 26905.500
| | | OverTime = No: No {No=7, Yes=1}
| | | OverTime = Yes: Yes {No=1, Yes=5}
| | MonthlyRate ≤ 26905.500
| | | YearsAtCompany > 1.500
| | | | OverTime = No
| | | | | MonthlyBilling > 1121.500
| | | | | | MonthlyBilling > 1401: No {No=782, Yes=93}
| | | | | | MonthlyBilling ≤ 1401
| | | | | | | YearsInCurrentRole > 1: No {No=7, Yes=1}
| | | | | | | YearsInCurrentRole ≤ 1: Yes {No=0, Yes=2}
| | | | | MonthlyBilling ≤ 1121.500
| | | | | | YearsAtCompany > 4.500: No {No=5, Yes=0}
| | | | | | YearsAtCompany ≤ 4.500: Yes {No=0, Yes=3}
| | | | | OverTime = Yes
| | | | | | MonthlyBilling > 19703.500: Yes {No=0, Yes=2}
| | | | | | MonthlyBilling ≤ 19703.500
| | | | | | | YearsInCurrentRole > 12.500: No {No=13, Yes=0}
| | | | | | | YearsInCurrentRole ≤ 12.500
| | | | | | | | MonthlyRate > 2235
| | | | | | | | | YearsSinceLastPromotion > 14.500: Yes
| | | | | | | | | | YearsSinceLastPromotion ≤ 14.500: No
| | | | | | | | | | | {No=0, Yes=2}
| | | | | | | | | | | {No=211, Yes=79}
| | | | | | | | | | | | MonthlyRate ≤ 2235: No {No=9, Yes=0}
| | | | | | | | | | | | YearsAtCompany ≤ 1.500
| | | | | | | | | | | | Age > 19.500
| | | | | | | | | | | | | MonthlyRate > 26895.500: Yes {No=0, Yes=2}
| | | | | | | | | | | | | MonthlyRate ≤ 26895.500
| | | | | | | | | | | | | | YearsSinceLastPromotion > 4.500: No {No=5, Yes=0}
| | | | | | | | | | | | | | YearsSinceLastPromotion ≤ 4.500
| | | | | | | | | | | | | | | DistanceFromHome > 28.500: No {No=5, Yes=0}
| | | | | | | | | | | | | | | DistanceFromHome ≤ 28.500
| | | | | | | | | | | | | | | | DistanceFromHome > 27.500: Yes {No=0,
| | | | | | | | | | | | | | | | | Yes=2}
| | | | | | | | | | | | | | | | | DistanceFromHome ≤ 27.500: No {No=137,
| | | | | | | | | | | | | | | | | | Yes=66}
| | | | | | | | | | | | | | | | | | Age ≤ 19.500: Yes {No=0, Yes=3}
| | | | | | | | | | | | | | | | | | Age ≤ 18.500
| | | | | | | | | | | | | | | | | | | MonthlyRate > 7722.500: No {No=6, Yes=3}
| | | | | | | | | | | | | | | | | | | MonthlyRate ≤ 7722.500: Yes {No=0, Yes=3}

```

Gambar 7. Rule Base

Berikut merupakan serangkaian aturan dari pohon keputusan:

1. IF Age > 18.5 AND MonthlyBilling ≤ 19785 THEN Turnover = No.
2. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = No THEN Turnover = No.
3. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany ≤ 1.5 THEN Turnover = Yes.
4. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany > 1.5 THEN Turnover = No.
5. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = No AND MonthlyBilling ≤ 1121.5 THEN Turnover = No.
6. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = No AND MonthlyBilling > 1121.5 AND MonthlyBilling ≤ 1401 AND YearsInCurrentRole ≤ 1 THEN Turnover = Yes.
7. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = No AND MonthlyBilling > 1121.5 AND MonthlyBilling > 1401 THEN Turnover = No.
8. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = Yes AND MonthlyBilling ≤ 17003.5 THEN Turnover = Yes.
9. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = Yes AND MonthlyBilling > 17003.5 AND YearsInCurrentRole ≤ 12.5 THEN Turnover = No.
10. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = Yes AND MonthlyBilling > 17003.5 AND YearsInCurrentRole > 12.5 AND YearsSinceLastPromotion ≤ 14.5 THEN Turnover = Yes.

11. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany ≤ 1.5 AND OverTime = Yes AND MonthlyBilling > 17003.5 AND YearsInCurrentRole > 12.5 AND YearsSinceLastPromotion > 14.5 THEN Turnover = No.
12. IF Age ≤ 18.5 AND MonthlyRate ≤ 77722.5 THEN Turnover = No.
13. IF Age ≤ 18.5 AND MonthlyRate > 77722.5 THEN Turnover = Yes.
14. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany > 1.5 AND MonthlyBilling ≤ 1121.5 THEN Turnover = No.
15. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany > 1.5 AND MonthlyBilling > 1121.5 AND MonthlyBilling ≤ 17003.5 THEN Turnover = Yes.
16. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany > 1.5 AND MonthlyBilling > 17003.5 AND YearsSinceLastPromotion ≤ 1.5 THEN Turnover = No.
17. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany > 1.5 AND MonthlyBilling > 17003.5 AND YearsSinceLastPromotion > 1.5 AND YearsSinceLastPromotion ≤ 14.5 THEN Turnover = Yes.
18. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = Yes AND YearsAtCompany > 1.5 AND MonthlyBilling > 17003.5 AND YearsSinceLastPromotion > 14.5 THEN Turnover = No.
19. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate ≤ 26905.5 AND YearsAtCompany > 1.5 AND OverTime = No THEN Turnover = No.
20. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = No AND MonthlyRate ≤ 160500 THEN Turnover = No.
21. IF Age > 18.5 AND MonthlyBilling > 19785 AND MonthlyRate > 26905.5 AND OverTime = No AND MonthlyRate > 160500 THEN Turnover = Yes.

## Evaluasi

*Confusion Matrix* merupakan metode yang digunakan untuk mengukur dan menganalisis kinerja model dalam klasifikasi. Berikut adalah tabel *confusion matrix* dapat dilihat pada tabel 6. Accuracy: 81.63% +/- 1.24% (*micro average*: 81.63%)

Tabel 5. *Confusion Matrix*

	<i>True No</i>	<i>True Yes</i>	<i>Class precision</i>
<i>Pred. No</i>	1185	252	82.46%
<i>Pred. Yes</i>	18	15	45.45%
<i>class recall</i>	98.50%	5.62%	

Pada tabel 6 merupakan hasil *confusion matrix* yang berisi informasi terkait klasifikasi. Pada penelitian ini atribut yang dipilih sebagai label berisi dua kelompok data yaitu “Yes” dan “No”, sehingga pada tabel *confusion matrix* terdapat empat kemungkinan *output* yang didapatkan yaitu sebanyak 1.185 data diprediksi tidak melakukan *turnover* dan benar tidak melakukan *turnover* (*true positive*), terdapat 15 data diprediksi melakukan *turnover* dan benar melakukan *turnover* (*true negative*), lalu terdapat 253 data diprediksi tidak melakukan *turnover* sedangkan sebenarnya melakukan *turnover* (*false positive*), dan sebanyak 18 data diprediksi melakukan *turnover* padahal sebenarnya tidak melakukan *turnover* (*false negative*).

Evaluasi pada model dilakukan dengan menghitung nilai *Accuracy*, *Precision*, *Specificity*, *Recall*, dan *F-Measure* atau *F1-score*. Perhitungan evaluasi tersebut dapat dilihat pada tabel 7.

Tabel 6. Hasil Perhitungan Matriks Evaluasi

<i>Accuracy</i>	<i>Precision</i>	<i>Specificity</i>	<i>Recall</i>	<i>F-Measure</i>
81,63%	82,46%	5,62%	98,50%	89,77%

Hasil perhitungan menunjukkan keakuratan klasifikasi sebesar 81,63%. *Precision* menunjukkan bahwa 82,46% prediksi benar. *Specificity* atau *True Negative rate* menunjukkan bahwa dari semua karyawan yang tidak melakukan *turnover*, model memprediksi 5,62% dengan benar. *Recall* atau *true positive rate* memperlihatkan dari semua karyawan yang benar melakukan *turnover*, model mampu memprediksi 98,50% dengan benar. Dan *F1-score* diperoleh sebesar 89,77% menunjukkan bahwa model memiliki *Precision* dan *recall* yang baik. Berdasarkan evaluasi tersebut, model klasifikasi C4.5 yang dilakukan pada penelitian ini dapat dikatakan baik dengan nilai *accuracy* sebesar 81,62%. Model ini dapat membantu perusahaan dalam memprediksi karyawan yang berpotensi melakukan *turnover*.

## KESIMPULAN

Berdasarkan hasil penelitian yang diperoleh dengan menerapkan klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa terdapat 8 atribut yang secara signifikan mempengaruhi *turnover* karyawan yaitu *Age*, *Monthly Billing*, *Monthly Rate*, *OverTime*, *YearsAtCompany*, *YearsInCurrentRole*, *YearsLastPromotion*, dan *DistanceFromHome*. Menggunakan *cross validation*, akurasi yang diperoleh sebesar 81,63%, diharapkan dengan adanya model ini dapat membantu perusahaan untuk memprediksi *turnover* karyawan dan dapat mengelola dan mengurangi tingkat *turnover* di perusahaan. Dengan tingkat akurasi yang diperoleh pada penelitian ini yaitu sebesar 81,63%, penelitian ini dapat dikembangkan lagi dengan menerapkan algoritma *data mining* lainnya sehingga dapat membandingkan tingkat akurasinya dan mengetahui prediksi *turnover* karyawan dengan lebih baik lagi.

## DAFTAR PUSTAKA

- [1] D. Saruksuk, M. Magito, and D. H. Perkasa, "Pengaruh Disiplin, Beban Kerja, dan Kompensasi terhadap Turn Over Karyawan," *GLOBAL: Jurnal Lentera BITEP*, vol. 1, no. 01, pp. 43–53, 2023.
- [2] A. P. Dinata and S. Suhana, "Pengaruh Beban Kerja, Stres Kerja, dan Job Insecurity Terhadap Turnover Karyawan," *Journal of Management and Bussines (JOMB)*, vol. 5, no. 1, pp. 722–731, 2023.
- [3] A. P. A. NingTyas, S. H. Purnomo, and A. Aswar, "Pengaruh Kepuasan Kerja Terhadap Turnover Intention Dengan Komitmen Organisasi Sebagai Variabel Intervening," *E-Jurnal Manajemen Universitas Udayana*, vol. 9, no. 4, p. 1634, Apr. 2020, doi: 10.24843/EJMUNUD.2020.v09.i04.p20.
- [4] L. Tamengkel and W. Rumawas, "Analysis of factors that influence employees turnover intention (Study of starred hotel employees in North Sulawesi)," *Jurnal Administrasi Bisnis (Jab)*, vol. 10, no. 1, pp. 32–40, 2020.

- [5] Luthfiyyah Wasiilah Maahiroh, “Klasifikasi Turnover Karyawan Menggunakan Algoritma XGBoost (Studi kasus: Divisi Engineering, Perusahaan Jasa Pertambangan),” Universitas Islam Indonesia, 2024.
- [6] D. A. Harvida and C. Wijaya, “Faktor Yang Mempengaruhi Turnover Karyawan dan Strategi Retensi Sebagai Pencegahan Turnover Karyawan: Sebuah Tinjauan Literatur,” *JIANA (Jurnal Ilmu Administrasi Negara)*, vol. 18, no. 2, pp. 13–23, 2020.
- [7] Lestari, “Pengaruh Lingkungan Kerja, Kompensasi, dan Komitmen Organisasional Terhadap Turnover Intention Pada PT Somnotec Indonesia dan Tinjauannya Dari Sudut Pandang Islam.” Universitas YASRI, 2018.
- [8] Larasati, “Hubungan Komitmen Organisasi Dengan Intensi Turnover (Keinginan Berpindah Pekerjaan) Pada Karyawan Outsourcing Di Pt.Telkom Regional 1 Sumatera,” Universitas Medan Area, 2019.
- [9] Z. Zafitri and M. I. Jambak, “Karakteristik Pembatalan Reservasi Kamar Hotel Pada Online Travel Agent Menggunakan Algoritma C4. 5,” *Indonesian Journal of Computer Science*, vol. 12, no. 4, 2023.
- [10] A. C. Imanda, N. Hidayat, and M. T. Furqon, “Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified K-Nearest Neighbor,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 8, pp. 2392–2399, 2018.
- [11] R. A. Siallagan, “Prediksi penyakit diabetes mellitus menggunakan algoritma c4. 5,” *Jurnal Responsif: Riset Sains dan Informatika*, vol. 3, no. 1, pp. 44–52, 2021.
- [12] A. Ramadhani, R. F. Noor, D. Vernanda, and T. Herdiawan, “Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma C4. 5 di Politeknik Negeri Subang,” *Jurnal Tekno Kompak*, vol. 18, no. 1, pp. 101–112, 2024.
- [13] I. Jayanto, “Analisis Perbandingan Algoritma Decision Tree untuk Prediksi Karyawan dengan Potensi Atrisi di PT. XYZ,” *FAHMA: Jurnal Informatika Komputer, Bisnis dan Manajemen*, vol. 22, no. 1, pp. 49–59, 2024.
- [14] <https://www.kaggle.com/dataset/a246/employee-turnover-dataset> “Employee Turnover Dataset.”
- [15] Eko Afrianto, “Sistem Informasi Klasifikasi Siswa Penerima Kartu Indonesia pintar (KIP) Menggunakan Metode Decision Tree Dengan Algoritma C4.5,” Universitas Diponegoro, 2020.
- [16] N. A’ayunnisa, Y. Salim, and H. Azis, “Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab,” *Indonesian Journal of Data and Science*, vol. 3, no. 3, pp. 115–121, 2022.
- [17] A. Sifaunajah and R. D. Wahyuningtyas, “Penggunaan Algoritma ID3 Untuk Klasifikasi Data Calon Peserta Didik,” *CSRID (Computer Science Research and Its Development Journal)*, vol. 14, no. 2, p. 103, 2022.

