

KLASIFIKASI POTENSI TURNOVER KARYAWAN BERDASARKAN DATA KINERJA MENGGUNAKAN ALGORITMA C4.5 STUDI KASUS PT. REPUBLIC TECHNOLOGY NUSANTARA

Aditama ¹, Achmad Praditya ², Moh Raffles Zein ³, Muzfirah Nadriani ⁴,
Omar Ghifari ⁵, Trisnawati ⁶, Baginda Oloan Lubis ⁷

^{1,2,3,4,5,6} Sistem Informasi, Bina Sarana Informatika

⁷ Rekayasa Perangkat Lunak, Bina Sarana Informatika

Jl. Kramat Raya No. 98 Jakarta, Indonesia

²achmadpraditya@gmail.com

ABSTRAK

Pergantian karyawan merupakan tantangan utama dalam pengelolaan sumber daya manusia yang berdampak besar pada stabilitas organisasi dan biaya operasional. Penelitian ini bertujuan untuk menganalisis potensi pergantian karyawan di PT. Republic Technology Nusantara dengan menerapkan algoritma pohon keputusan C4.5. Data yang digunakan mencakup informasi anggota satuan pengamanan, seperti riwayat pelatihan, pengalaman kerja, hasil evaluasi, dan absensi. Melalui proses pengumpulan dan prapemrosesan data yang sistematis, model klasifikasi dikembangkan untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap pergantian karyawan. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu secara efektif memprediksi turnover karyawan, sehingga memberikan wawasan berharga bagi manajemen dalam menyusun strategi retensi yang lebih optimal. Selain berkontribusi pada metode prediksi turnover, penelitian ini juga menawarkan panduan praktis dalam pengelolaan sumber daya manusia di perusahaan. Dengan menerapkan algoritma C4.5, perusahaan dapat lebih akurat dalam mengidentifikasi karyawan strategis dan berpotensi, sehingga dapat menangani tantangan pergantian karyawan dengan lebih efisien serta meningkatkan kualitas pengelolaan SDM. Hasil penelitian juga mengungkapkan bahwa kesesuaian gaji dengan UMR memiliki pengaruh terhadap tingkat keluar-masuknya karyawan, baik secara sukarela maupun tidak.

Kata kunci : Klasifikasi, Turnover Karyawan, Data Kinerja, Algoritma C4.5

1. PENDAHULUAN

Turnover karyawan merupakan isu yang sangat penting bagi perusahaan, karena dapat berdampak langsung pada keberhasilan dan perkembangan organisasi. Tingkat turnover yang tinggi dapat menurunkan produktivitas serta meningkatkan biaya operasional, termasuk biaya perekrutan dan pelatihan karyawan baru. Selain itu, tingginya turnover juga dapat mengganggu keharmonisan tim dan menurunkan kualitas kerja secara keseluruhan. Fenomena ini sering kali dipicu oleh berbagai faktor, salah satunya adalah ketidakpuasan karyawan terhadap kondisi lingkungan kerja, yang mendorong mereka untuk mencari peluang di tempat lain. [1]

Sumber daya manusia (SDM) memiliki peran yang sangat penting dalam sebuah organisasi, karena SDM merupakan aset utama yang mendukung pencapaian tujuan perusahaan. Dengan pengelolaan SDM yang efektif, karyawan dapat bekerja secara maksimal dan memberikan kontribusi terbaik untuk kemajuan perusahaan. [2] Oleh karena itu, perusahaan perlu mengelola kinerja karyawan secara tepat dan objektif, salah satunya dengan menerapkan proses evaluasi kinerja yang sistematis dan terorganisir. [3]

Kinerja karyawan dipengaruhi oleh berbagai faktor, salah satunya adalah pemberian kompensasi yang tepat, seperti kenaikan gaji. Gaji yang sesuai dapat meningkatkan motivasi dan komitmen karyawan terhadap perusahaan. Oleh karena itu, perusahaan

perlu memiliki sistem yang efektif untuk menilai dan mengelola kinerja karyawan. [4] Namun, untuk menentukan karyawan yang berprestasi, diperlukan proses penilaian yang objektif dan tepat. Tanpa adanya sistem yang jelas, keputusan yang diambil dapat menyebabkan kesalahan dalam penempatan jabatan, yang pada gilirannya dapat memengaruhi produktivitas perusahaan. [3]

Dalam pengelolaan SDM, teknologi, terutama data mining, dapat dimanfaatkan untuk meningkatkan akurasi dalam proses seleksi dan prediksi kinerja karyawan. Salah satu metode dalam data mining yang dapat diterapkan adalah algoritma C4.5, yang berfungsi untuk mengklasifikasikan data berdasarkan kriteria tertentu. Algoritma ini dapat membantu perusahaan memprediksi potensi turnover karyawan dengan mengklasifikasikan karyawan berdasarkan data kinerja mereka. [5] Teknik clustering dalam data mining, seperti algoritma K-Means, juga dapat diterapkan untuk mengelompokkan karyawan berdasarkan karakteristik kinerja mereka, sehingga mempermudah pengambilan keputusan terkait promosi atau pemutusan hubungan kerja. [4]

Tantangan yang dihadapi perusahaan dalam manajemen SDM tidak hanya berkaitan dengan pengelolaan kinerja, tetapi juga masalah eksternal, seperti serangan dunia maya. Salah satu masalah yang umum dalam dunia digital adalah serangan spam email, yang dapat mengganggu produktivitas

perusahaan. Untuk mengatasi hal ini, penggunaan algoritma seperti Naive Bayes dan C4.5 dapat membantu mendeteksi spam email, sehingga meningkatkan efisiensi komunikasi perusahaan. [6]

Oleh karena itu, perusahaan perlu mengelola SDM dengan baik, mengoptimalkan teknologi, dan mengurangi turnover karyawan untuk memastikan keberhasilan jangka panjang. Penelitian ini bertujuan untuk mengklasifikasikan potensi turnover karyawan di PT. Republic Technology Nusantara menggunakan algoritma C4.5, berdasarkan data kinerja karyawan. Pendekatan ini diharapkan dapat memberikan wawasan yang lebih jelas bagi manajemen perusahaan dalam mengambil keputusan yang lebih tepat terkait pengelolaan karyawan. [7]

Turnover merujuk pada proses ketika karyawan meninggalkan organisasi secara sukarela. Bentuk turnover meliputi perpindahan keluar organisasi, pengunduran diri, pemberhentian, atau kematian anggota organisasi. Terdapat dua jenis turnover, yaitu turnover sukarela (voluntary) dan turnover tidak sukarela (involuntary). Menurut Shaw, baik turnover sukarela maupun tidak sukarela dihitung sebagai turnover, termasuk berhenti, diberhentikan, mengundurkan diri, atau dipecat. Abelson mengemukakan bahwa karyawan yang meninggalkan organisasi secara sukarela sebagian besar dikategorikan sebagai turnover sukarela yang dapat dihindari (avoidable voluntary turnover) dan yang tidak dapat dihindari (unavoidable voluntary turnover). Penyebab avoidable voluntary turnover antara lain kompensasi gaji, hubungan dengan atasan, kondisi kerja, atau tawaran dari organisasi lain yang lebih baik, sedangkan unavoidable voluntary turnover disebabkan oleh faktor keluarga atau jalur karir. Dengan demikian, turnover dapat dipahami sebagai suatu proses di mana karyawan berhenti dari pekerjaannya. [8]

2. TINJAUAN PUSTAKA

2.1. Penyebab Tingginya Turnover Karyawan

Tingginya angka turnover karyawan dipengaruhi oleh beberapa faktor utama. Salah satunya adalah kurangnya motivasi kerja, di mana sebagian karyawan hanya bekerja untuk memenuhi kebutuhan ekonomi tanpa merasa pekerjaannya sesuai dengan passion atau keahlian mereka. Hal ini dapat menyebabkan penurunan motivasi, peningkatan stres, dan terhambatnya pengembangan karier. Selain itu, beban kerja yang berlebihan juga menjadi penyebab utama. Tugas yang terlalu banyak, termasuk tuntutan lembur dan bekerja di hari libur, sering kali mengganggu keseimbangan antara pekerjaan dan kehidupan pribadi, yang pada akhirnya memicu frustrasi dan meningkatkan risiko turnover.

Faktor lain yang berkontribusi adalah ketidaksesuaian antara gaji dan beban kerja. Ketika karyawan merasa bahwa kompensasi yang diterima tidak sebanding dengan tanggung jawab yang mereka emban, mereka cenderung mencari peluang lain yang

lebih menguntungkan. Oleh karena itu, pemberian gaji yang kompetitif serta penghargaan berbasis kinerja dapat menjadi strategi efektif dalam mempertahankan karyawan. Selain itu, kurangnya feedback dan pengakuan terhadap kinerja karyawan juga dapat menurunkan semangat kerja. Umpan balik yang konstruktif dan apresiasi atas kontribusi mereka sangat penting untuk meningkatkan motivasi dan produktivitas.

Terakhir, sistem seleksi yang kurang tepat dapat menjadi pemicu turnover yang cukup signifikan. Mempekerjakan individu yang tidak sesuai dengan budaya atau nilai-nilai perusahaan dapat menyebabkan ketidaknyamanan bagi karyawan tersebut, sekaligus merugikan organisasi secara keseluruhan. Oleh karena itu, proses rekrutmen yang selektif dan berbasis kompetensi sangat penting untuk memastikan bahwa karyawan yang direkrut memiliki kecocokan dengan perusahaan, sehingga dapat bertahan lebih lama dan berkontribusi secara optimal.

2.2. Data Mining

Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan trend dengan menambang sejumlah repository data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in [9]

2.3. Algoritma C4.5

Algoritma C4.5 adalah salah satu metode dalam data mining yang digunakan untuk membangun model pohon keputusan. Algoritma ini dikembangkan oleh Ross Quinlan pada tahun 1993 sebagai penyempurnaan dari algoritma pendahulunya, yaitu ID3. [10]

Algoritma C4.5 memanfaatkan konsep pembelajaran mesin untuk membangun model prediksi berdasarkan data yang tersedia. Inti dari algoritma ini adalah menganalisis struktur data dan mengelompokkan data ke dalam kategori yang sesuai. Proses ini membantu pengguna dalam mengeksplorasi hubungan antar variabel dan mendapatkan wawasan yang lebih mendalam tentang dataset yang dianalisis. [11]

2.4. Fungsi Algoritma C4.5

Algoritma C4.5 memiliki fungsi utama untuk membangun model pohon keputusan yang dapat digunakan dalam proses klasifikasi atau prediksi. Model ini memanfaatkan atribut-atribut dalam data untuk membantu pengambilan keputusan. Selain itu, algoritma C4.5 dilengkapi dengan fitur pemangkasan (pruning) pohon keputusan guna mencegah terjadinya overfitting. [12]

2.5. Pohon Keputusan (Decision Tree)

Pohon keputusan (Decision Tree) adalah salah satu metode kecerdasan buatan berbentuk struktur pohon, di mana setiap cabang merepresentasikan pilihan atau alternatif, sedangkan daun menunjukkan keputusan yang diambil. Metode ini populer karena mampu mengintegrasikan pola atau informasi ke dalam bentuk pohon keputusan. Decision Tree terdiri dari sejumlah node (simpul) yang saling terhubung melalui cabang. Secara umum, terdapat tiga jenis node dalam struktur Decision Tree. [13]

- Root node adalah simpul awal dalam pohon keputusan yang tidak menerima input namun memiliki lebih dari satu output.
- Internal node adalah simpul yang menerima satu input dan menghasilkan lebih dari dua output.
- Leaf atau terminal node adalah simpul yang hanya menerima satu input tanpa menghasilkan output.

Dalam setiap pohon keputusan, setiap leaf mewakili satu nama kelas. Sementara itu, root node dan internal node memuat aturan yang digunakan untuk membedakan data berdasarkan karakteristiknya.

3. METODE PENELITIAN

Data penelitian merupakan kumpulan informasi atau fakta yang dikumpulkan atau dihasilkan selama proses penelitian. Data ini biasanya berbentuk fakta, angka, atau informasi lain yang dapat diukur dan diuji secara empiris. Data penelitian dapat diperoleh melalui berbagai metode, seperti survei, wawancara, observasi, atau pengukuran. Adapun rincian data yang digunakan dalam penelitian ini adalah sebagai berikut:

- Sumber Data**
Data penelitian dalam studi ini diperoleh dari PT. Republic Technology Nusantara. Data tersebut mencakup informasi tentang anggota satuan pengamanan perusahaan, termasuk atribut-atribut yang relevan untuk mengevaluasi kinerja dan kualitas mereka.
- Jenis Data**
Data yang dikumpulkan mencakup variabel seperti riwayat pelatihan, pengalaman kerja, hasil evaluasi atau tes sebelumnya, serta informasi lain yang berhubungan dengan kinerja anggota satuan pengamanan.
- Sifat dan Sumber Data**
Sumber data yang digunakan adalah data internal PT. Republic Technology Nusantara yang bersifat privat, meliputi data personal, profil, dan catatan kinerja anggota satuan pengamanan.

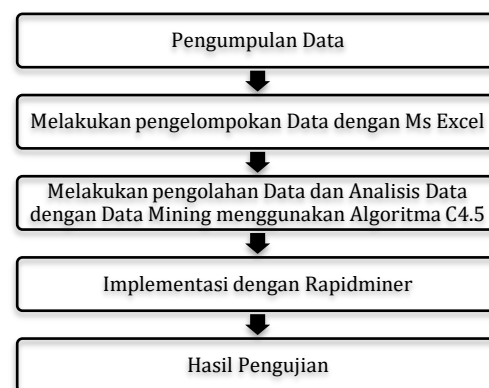
Penelitian ini dilakukan dengan mematuhi aturan dan regulasi yang berlaku untuk menjaga kerahasiaan dan privasi data. Peneliti juga telah mendapatkan izin dari PT. Republic Technology Nusantara untuk mengakses dan menganalisis data tersebut.

3.1. Teknik Pengumpulan Data

Pada tahap pengumpulan data, dilakukan observasi sebagai metode utama, yaitu proses mengumpulkan data dengan cara mengamati langsung lokasi penelitian dan mencatat secara sistematis berbagai aspek yang relevan dengan permasalahan yang akan diteliti. [14] Observasi ini dirancang untuk mendukung tujuan penelitian, di mana peneliti secara terencana dan terstruktur mencatat data serta fakta terkait evaluasi kinerja anggota satuan pengamanan di PT. Republic Technology Nusantara.

3.2. Tahapan Penelitian

Pada Gambar 1 berikut ini ditampilkan tahapan-tahapan penelitian yang menggambarkan alur sistematis dari proses yang dilakukan:



Gambar 1. Tahapan Penelitian [15]

Berikut ini adalah penjelasan mengenai tahapan-tahapan penelitian:

- Pengumpulan Data**
Langkah awal dalam penerapan metodologi adalah mengumpulkan data yang relevan dari PT. Republic Technology Nusantara. Proses ini mencakup akses ke sistem atau basis data perusahaan yang menyimpan informasi terkait anggota satuan pengamanan, seperti kinerja, absensi, dan atribut lain yang berhubungan dengan evaluasi mereka.
- Pra-Pemrosesan**
Setelah data terkumpul, tahap berikutnya adalah prapemrosesan data. Proses ini mencakup:
 - Pembersihan Data:** Menghapus data yang tidak valid, duplikat, atau tidak relevan.
 - Pemilihan Atribut:** Menentukan atribut yang paling relevan atau memiliki pengaruh besar terhadap klasifikasi anggota satuan pengamanan.
- Penanganan Nilai Hilang atau Outlier**
Memperbaiki atau menghapus nilai yang hilang, serta menangani data yang berada di luar batas wajar.

- d. Pemilihan Fitur
Menggunakan teknik seperti analisis korelasi, analisis informasi, atau pemilihan berbasis model untuk memilih atribut yang paling relevan dan informatif dalam proses klasifikasi.
- e. Penerapan Algoritma C4.5
Setelah tahap prapemrosesan, algoritma C4.5 diterapkan pada data yang telah disiapkan.

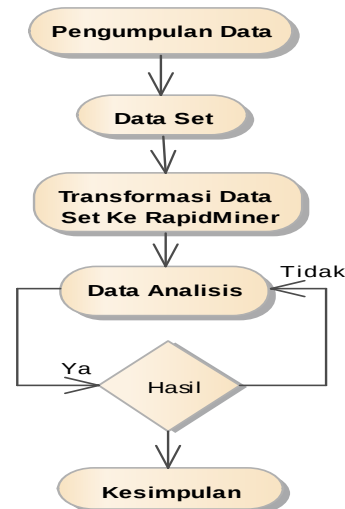
Langkah-langkah dalam penerapan algoritma ini adalah sebagai berikut:

- a. Pengklasifikasian: Melakukan proses klasifikasi untuk mengevaluasi kinerja anggota satuan pengaman di PT. Republic Technology Nusantara. Data yang dikumpulkan digunakan untuk memprediksi hasil penilaian kinerja anggota pengamanan, berdasarkan kinerja, perilaku, dan absensi yang tercatat selama tahun pertama evaluasi.
- b. Pengelompokan Data: Data dikelompokkan menggunakan Microsoft Excel untuk mempermudah analisis, dengan pengelompokan berdasarkan bulan dan atribut yang relevan.
- c. Pengolahan dan Analisis Data: Setelah pengelompokan, data diproses dan dianalisis menggunakan algoritma C4.5. Data karyawan anggota satuan pengaman kemudian diolah dan diklasifikasikan untuk menghasilkan pohon keputusan (decision tree).
- d. Implementasi dengan RapidMiner: Data yang telah diolah diuji menggunakan aplikasi RapidMiner untuk menentukan hasil akhir dari penelitian.
- e. Hasil Penelitian: Penetapan masa evaluasi kinerja karyawan satuan pengaman selama satu tahun, yang dihasilkan dari proses klasifikasi data sebelumnya menggunakan aplikasi RapidMiner dalam bentuk pohon keputusan dan aturan (rules).

3.3. Diagram Aktifitas Penelitian

Penelitian ini mengikuti alur kerja yang dijelaskan dalam diagram aktivitas pada Gambar 2,

yang menggambarkan proses identifikasi masalah dan tujuan penelitian melalui penggunaan aplikasi RapidMiner. [16]



Gambar 2. Diagram Aktifitas Penelitian

Data yang digunakan telah dikelompokkan ke dalam kelas-kelas tertentu. Selanjutnya, proses perhitungan akar pohon dilakukan dengan memilih atribut berdasarkan nilai gain tertinggi. Untuk mendapatkan nilai gain, langkah pertama adalah menghitung nilai entropy dari atribut yang tersedia. Tahapan algoritma Decision Tree C4.5 meliputi:

- a. Menyiapkan data training.
- b. Menentukan akar dari pohon.
- c. Menghitung nilai gain untuk setiap atribut.

4. HASIL DAN PEMBAHASAN

4.1. Data Mentah

Tabel 1 menyajikan informasi mengenai data karyawan yang mencakup atribut kehadiran, kinerja, preferensi gaji, dan potensi turnover (perpindahan kerja). Data ini diperoleh dari hasil pengumpulan secara langsung dan telah disusun dengan rapi dalam bentuk tabel. Berikut tabelnya:

Tabel 1. Data Turnover Karyawan

No	Nama Karyawan	Kehadiran	Kinerja	Preferensi Gaji	Turnover
1	Farah Citra	Kurang	Kurang	Mau Dibayar UMR	Ya
2	Farah Joko	Cukup	Rajin	Mau Dibayar UMR	Tidak
3	Joko Hana	Cukup	Cukup	Tidak Mau Dibayar UMR	Ya
4	Eko Gilang	Rajin	Rajin	Tidak Mau Dibayar UMR	Tidak
5	Farah Indra	Rajin	Kurang	Mau Dibayar UMR	Tidak
6	Hana Hana	Kurang	Cukup	Tidak Mau Dibayar UMR	Ya
7	Citra Budi	Kurang	Kurang	Mau Dibayar UMR	Tidak
8	Farah Eko	Cukup	Cukup	Mau Dibayar UMR	Tidak
9	Indra Dewi	Cukup	Rajin	Mau Dibayar UMR	Ya
10	Eko Achmad	Kurang	Kurang	Tidak Mau Dibayar UMR	Ya
11	Budi Farah	Kurang	Cukup	Mau Dibayar UMR	Ya
12	Gilang Indra	Rajin	Rajin	Tidak Mau Dibayar UMR	Ya
13	Hana Gilang	Rajin	Cukup	Mau Dibayar UMR	Tidak
14	Hana Joko	Kurang	Kurang	Mau Dibayar UMR	Ya

4.2. Perhitungan Gani dan Entropy

Tabel 2 Hasil Perhitungan Gain Dan Entropy Pada Node 1

Ket	Jumlah K	Turnover		Entropy	Gain
		Ya	Tidak		
Total	14	8	6	0,95228	
Kehadiran					1,224155134
Kurang	6	5	1	0,65002	
Cukup	4	2	2	1	
Rajin	4	1	3	0,81128	
Kinerja					2,427376486
Gaji				0,99108	
Mau	9	4	5		
Tidak Mau	5	4	1	0,72193	0,721928095

Tabel 2 tersebut menganalisis pengaruh atribut Gaji terhadap turnover (perpindahan kerja) berdasarkan nilai entropy dan gain. Berikut adalah penjelasan rinci:

a. Kolom-Kolom

- Ket (Keterangan): Menunjukkan kategori atau atribut yang dianalisis.
- Jumlah K: Jumlah total data dalam kategori tersebut.
- Turnover Ya: Jumlah karyawan yang berpotensi melakukan turnover.
- Turnover Tidak: Jumlah karyawan yang tidak berpotensi melakukan turnover.
- Entropy: Ukuran ketidakpastian atau keragaman data dalam kategori tersebut.
- Gain: Informasi yang diperoleh dari atribut, yang menunjukkan seberapa besar atribut tersebut memengaruhi keputusan.

b. Penjelasan Per Baris

Total: Dari 5 data total, 2 orang berpotensi turnover, dan 3 tidak. Nilai entropy keseluruhan adalah 0,970950594, mencerminkan tingkat ketidakpastian yang cukup tinggi dalam data.

c. Gaji

- Mau: Dari 4 karyawan yang mau dibayar sesuai UMR, 1 orang berpotensi turnover, dan 3 tidak, dengan entropy 0,811278124.
- Tidak Mau: Dari 1 karyawan yang tidak mau dibayar UMR, seluruhnya (1 orang) berpotensi turnover, sehingga entropy bernilai 0 (tidak ada ketidakpastian).

d. Gain

Gain untuk atribut Gaji bernilai negatif (-0,04315). Nilai negatif ini mengindikasikan bahwa atribut Gaji tidak memberikan kontribusi positif terhadap pengurangan ketidakpastian dalam keputusan turnover pada node ini.

e. Kesimpulan

Atribut Gaji pada data ini memiliki pengaruh yang rendah terhadap turnover, sebagaimana ditunjukkan oleh nilai gain negatif. Meskipun entropy lebih rendah pada kategori Tidak Mau, data ini tidak cukup signifikan untuk menyimpulkan bahwa atribut Gaji menjadi faktor utama yang memengaruhi turnover.

Tabel 3 menyajikan hasil perhitungan gain dan entropy pada node 2.

Tabel 3. Hasil Perhitungan Gain Dan Entropy Pada Node 3

Node 2	Keterangan	Jumlah Kasus (s)	Turnover		Entropy	Gain
			(ya)	(tidak)		
Total		5	2	3	0,97095059	
Gaji						-0,0432
	Mau	4	1	3	0,81127812	
	Tidak	1	1	0	0	

Tabel 3 merupakan tabel yang memuat informasi analisis atribut Gaji terhadap turnover pada node 2. Berikut adalah penjelasan rinci berdasarkan isi tabel:

a. Struktur Tabel

- Node: Menunjukkan level atau posisi analisis (dalam hal ini, node 2).
- Keterangan: Atribut atau kategori yang dianalisis (misalnya, Total dan Gaji).
- Jumlah Kasus (s): Total jumlah kasus data pada kategori tersebut.
- Turnover (Ya): Jumlah karyawan yang berpotensi mengalami turnover.

- Turnover (Tidak): Jumlah karyawan yang tidak berpotensi mengalami turnover.
- Entropy: Ukuran tingkat ketidakpastian atau keragaman data.
- Gain: Nilai yang menunjukkan seberapa signifikan atribut tersebut memengaruhi keputusan turnover.

b. Penjelasan Data

Total: Jumlah kasus total pada node 2 adalah 5 orang. Dari jumlah tersebut, 2 orang berpotensi turnover, dan 3 orang tidak. Entropy total adalah 0,970950594, menunjukkan tingkat ketidakpastian data yang cukup tinggi.

c. Gaji

- Mau: Ada 4 kasus di mana karyawan ingin dibayar sesuai UMR. Dari jumlah tersebut, 1 orang berpotensi turnover, dan 3 orang tidak. Nilai entropy pada kategori ini adalah 0,811278124, yang berarti ketidakpastian cukup berkurang dibandingkan total.
- Tidak: Ada 1 kasus di mana karyawan tidak mau dibayar sesuai UMR. Dari jumlah tersebut, 1 orang berpotensi turnover, dan tidak ada yang tidak berpotensi. Entropy pada kategori ini adalah 0, karena tidak ada ketidakpastian dalam data.

d. Gain

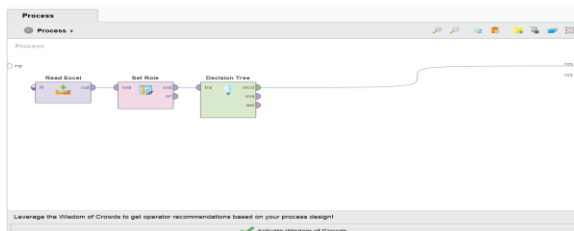
Nilai gain untuk atribut Gaji adalah -0,04315. Nilai negatif ini menunjukkan bahwa atribut Gaji tidak memberikan kontribusi signifikan dalam mengurangi ketidakpastian pada node 2. Dengan kata lain, atribut Gaji tidak efektif dalam membantu prediksi turnover pada data ini.

e. Kesimpulan

Atribut Gaji pada node 2 tidak memberikan pengaruh yang signifikan terhadap keputusan turnover, sebagaimana terlihat dari nilai gain yang negatif. Entropy kategori Tidak menunjukkan tidak ada keragaman, tetapi jumlah kasusnya terlalu kecil (hanya 1 orang) untuk menjadi dasar analisis yang kuat.

4.3. Penerapan Rapid Miner

Pada RapidMiner, proses pengujian dimulai dengan mengimpor dataset menggunakan modul Read Excel. Kemudian, modul Set Role digunakan sebagai target prediksi untuk atribut "label". Model Decision Tree dibangun dengan data pelatihan yang telah diatur pada tahap sebelumnya. Modul Read Excel mengimpor dataset pengujian, dan modul Set Role mengubah atributnya. Modul Performance menilai hasil prediksi untuk menentukan kebenaran model, sedangkan modul Apply Model menerapkan model yang dikembangkan ke data uji. Model Decision Tree yang dapat digunakan untuk memprediksi label dataset dihasilkan dengan menghubungkan semua modul untuk membuat alur pemrosesan data yang terstruktur. Berikut Gambar 2. Penerapan dengan Rapid Miner dari Data set.



Gambar 2. gambar disgin pohon Keputusan



Gambar 3. Hasil pohon Keputusan

Berdasarkan gambar 3:

a. Struktur Pohon Keputusan

Diagram di atas menunjukkan proses klasifikasi Turnover (Ya/Tidak) berdasarkan dua variabel utama, yaitu:

- Kinerja: Terdiri dari tiga kategori - Rajin, Cukup, dan Kurang.
- Preferensi Gaji: Apakah individu bersedia dibayar UMR atau tidak.

b. Alur Keputusan

- Node Utama: Variabel Kinerja dipilih sebagai pemisah pertama karena memiliki nilai *gain* tertinggi (dari analisis sebelumnya).

• Cabang 1: Kinerja – Rajin

Jika kinerja Rajin, maka hasilnya adalah Turnover = Tidak. Grafik menunjukkan distribusi data yang lebih banyak pada kategori "Tidak".

c. Cabang 2: Kinerja – Cukup

Jika kinerja Cukup, maka variabel Preferensi Gaji digunakan sebagai pemisah berikutnya:

- Mau dibayar UMR → Turnover = Tidak.
- Tidak mau dibayar UMR → Turnover = Ya.

d. Cabang 3: Kinerja – Kurang

Jika kinerja Kurang, maka hasilnya adalah Turnover = Ya. Grafik memperlihatkan distribusi dominan pada "Ya".

e. Interpretasi Visual Grafik

Grafik batang di setiap node menggambarkan distribusi proporsi data:

- Hijau: Turnover = Tidak.
- Merah: Turnover = Ya.
- Biru: Distribusi tambahan yang menunjukkan kontribusi data di setiap cabang.

f. Kesimpulan

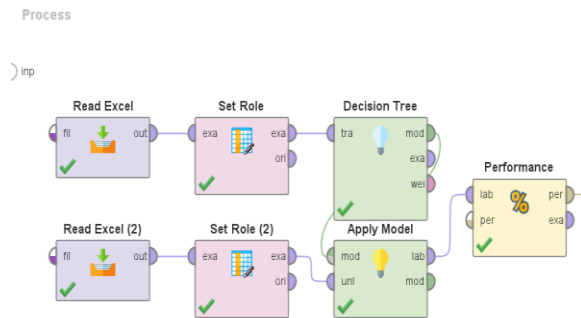
- Kinerja adalah faktor pertama yang memengaruhi keputusan Turnover.
 - Rajin cenderung tidak melakukan Turnover.
 - Kurang memiliki kecenderungan kuat untuk melakukan Turnover.

- Preferensi Gaji menjadi faktor tambahan bagi karyawan dengan kinerja Cukup.

Mereka yang tidak mau dibayar UMR lebih cenderung melakukan Turnover.

Dengan demikian, pohon keputusan ini membantu memetakan faktor-faktor utama yang memengaruhi keputusan Turnover karyawan berdasarkan Kinerja dan Preferensi Gaji.

4.4. Desain Performance



Gambar 4. Desain Performance

Gambar 4 metupakan alur proses untuk membangun dan mengevaluasi model pohon keputusan menggunakan data dari file Excel. Proses dimulai dengan modul Read Excel, yang membaca dua dataset: satu untuk pelatihan model (training data) dan satu untuk pengujian model (test data). Selanjutnya, modul Set Role digunakan untuk menentukan peran masing-masing kolom dalam dataset, seperti menetapkan kolom target (label) yang akan diprediksi dan fitur-fitur yang digunakan dalam proses pembelajaran. Setelah itu, modul Decision Tree membangun model pohon keputusan berdasarkan training data, mempelajari pola dari data untuk memprediksi nilai target. Model yang telah dilatih kemudian diterapkan pada test data melalui modul Apply Model untuk menghasilkan prediksi. Akhirnya, modul Performance digunakan untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi terhadap nilai sebenarnya dalam test data, menggunakan metrik evaluasi seperti akurasi, presisi, atau recall. Proses ini menghasilkan output yang menunjukkan hasil analisis model. Diagram ini menggambarkan alur kerja analisis data yang biasa dilakukan pada platform seperti RapidMiner.

4.5. Tampilan hasil performance

Table View Plot View

accuracy: 50.00%

	true Ya	true Tidak	true tidak	true ya	class precision
pred. Ya	3	2	0	1	50.00%
pred. Tidak	0	2	0	1	66.67%
pred. tidak	0	0	0	0	0.00%
pred. ya	1	1	1	2	40.00%
class recall	75.00%	40.00%	0.00%	50.00%	

Gambar 5. Hasil Performance

a. Accuracy

- Accuracy menunjukkan seberapa sering model membuat prediksi yang benar.
- Nilai akurasi adalah 50.00%, artinya dari semua data, 50% prediksi model sesuai dengan label sebenarnya.

b. Baris dan Kolom

- Baris: Mewakili prediksi model (pred. Ya, pred. Tidak, pred. tidak, pred. ya).
- Kolom: Mewakili label sebenarnya (true Ya, true Tidak, true tidak, true ya).

c. Interpretasi Data per Baris

- Baris pred. Ya:
 - Model memprediksi "Ya".
 - 3 prediksi benar (true Ya).
 - 2 prediksi salah (true Tidak).
 - 1 prediksi salah (true ya).
 - Class Precision: 50.00%, artinya setengah dari prediksi "Ya" benar.
- Baris pred. Tidak:
 - Model memprediksi "Tidak".
 - 2 prediksi benar (true Tidak).
 - 1 prediksi salah (true ya).
 - Class Precision: 66.67%, artinya sebagian besar prediksi "Tidak" benar.
- Baris pred. tidak:
 - Model memprediksi "tidak".
 - Semua prediksi salah (0 nilai di semua kolom).
 - Class Precision: 0.00%, artinya tidak ada prediksi "tidak" yang benar.
- Baris pred. ya:
 - Model memprediksi "ya".
 - 1 prediksi benar (true ya).
 - 1 salah prediksi untuk true Ya dan true Tidak.
 - 2 prediksi benar (true ya).
 - Class Precision: 40.00%, menunjukkan hanya sebagian kecil prediksi "ya" yang benar.

d. Recall per Kelas

Recall mengukur seberapa baik model mengidentifikasi semua contoh positif sebenarnya untuk setiap kelas:

- true Ya: 75.00% (model berhasil menangkap 75% data yang benar-benar "Ya").
- true Tidak: 40.00%.
- true tidak: 0.00% (tidak ada prediksi yang benar).
- true ya: 50.00%.

5. KESIMPULAN DAN SARAN

Kesimpulan dari analisis pohon keputusan menunjukkan bahwa faktor utama yang memengaruhi keputusan Turnover karyawan adalah variabel Kinerja, diikuti oleh Preferensi Gaji sebagai faktor tambahan. Kinerja karyawan menjadi pemisah utama dalam struktur pohon keputusan, dengan hasil yang menunjukkan bahwa karyawan dengan kinerja Rajin cenderung tidak melakukan Turnover, sementara mereka yang memiliki kinerja Kurang sangat berpotensi untuk melakukan Turnover. Untuk karyawan dengan kinerja Cukup, Preferensi Gaji menjadi variabel penentu berikutnya. Karyawan yang bersedia menerima gaji sesuai UMR cenderung tidak melakukan Turnover, sedangkan mereka yang tidak mau dibayar UMR lebih berisiko untuk meninggalkan perusahaan. Distribusi data pada grafik batang di setiap node memberikan visualisasi yang jelas tentang proporsi keputusan Turnover, dengan warna hijau mewakili "Tidak", merah untuk "Ya", dan biru untuk

distribusi tambahan. Secara keseluruhan, analisis ini memberikan wawasan yang mendalam tentang bagaimana Kinerja dan Preferensi Gaji memengaruhi keputusan Turnover, sehingga dapat digunakan untuk merancang strategi retensi yang lebih efektif berdasarkan data tersebut. Akurasi rendah (50%) menunjukkan bahwa model masih belum optimal dan memiliki banyak prediksi yang salah. Kelas dengan class recall rendah (seperti "true tidak") memerlukan perhatian khusus dalam perbaikan model.

DAFTAR PUSTAKA

- [1] D. Pitoyo and D. S. Rohman, "Analisis Faktor-Faktor Yang Mempengaruhi Turnover Karyawan Pada RM. Cikutra Raya," *Rekayasa Ind. dan Mesin*, vol. 1, no. 2, p. 116, 2020, doi: 10.32897/retims.2020.1.2.324.
- [2] I. Holy, H. Haedar, and S. R. Dewi, "Pengaruh Pelatihan Dan Pengembangan Karyawan Terhadap Produktivitas Kerja Karyawan," *Jesya*, vol. 6, no. 2, pp. 1761–1771, 2023, doi: 10.36778/jesya.v6i2.1134.
- [3] E. Devia, "Penerapan Decision Tree Dengan Algoritma C4.5 Untuk Menentukan Rekomendasi Kenaikan Jabatan Karyawan," *J. Inf. Syst.*, vol. 3, no. 1, pp. 28–37, 2023, doi: 10.61488/jis.v3i1.257.
- [4] H. S. Kasman, "Perbandingan Efektifitas Algoritma K-Means Clustering -Topsis dan K-Medoids Clustering -Topsis dalam Menentukan Karyawan IT dengan Kinerja Terbaik A. Peneliti Terdahulu K-Means merupakan metode clustering non-hierarkis yang berusaha membagi data ke dalam," vol. 4, no. 2, 2024.
- [5] A. M. Husein and M. Brutu, "Prediksi Penerimaan Calon Karyawan Dengan Menggunakan Algoritma C4.5 Pada Biro Kesejahteraan Rakyat Provinsi Sumatera Utara," *Digit. Transform. Technol.*, vol. 2, no. 1, pp. 16–20, 2022, doi: 10.47709/digitech.v2i1.1769.
- [6] Sulaeman, Nana Suarna, Abdul Ajiz, Agus Bahtiar, and Fathurrohman, "Perbandingan Kinerja Algoritma Naïve Bayes Dan C.45 Dalam Klasifikasi Spam Email," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 6, no. 1, pp. 8–14, 2022, doi: 10.32485/kopertip.v6i1.130.
- [7] N. Salsabila and B. F. Tumanggor, "Pengaruh Kepuasan Kerja dan Stres Kerja Terhadap Tingkat Turnover Intention Pegawai Di PT PKSS Cabang Jakarta 1," *Jurna Sumber Daya Apar.*, vol. 5, no. 2, pp. 1–19, 2023.
- [8] D. A. Harvida and C. Wijaya, "Faktor Yang Mempengaruhi Turnover Karyawan Dan Strategi Retensi Sebagai Pencegahan Turnover Karyawan : Sebuah Tinjauan Literatur," *J. Ilmu Adm. Negara*, vol. 16, no. 1, pp. 13–23, 2020.
- [9] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [10] C. I. Ifaru *et al.*, "Penentuan Sparepart Genset Paling Sering Digunakan Pada Operator Indosat Ooredoo Hutchison Dengan Algoritma C4.5," *J. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 300–309, 2024, doi: 10.35957/jtsi.v5i1.7481.
- [11] D. M. Musa *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 168–182, 2024, doi: 10.37012/jtik.v10i1.1985.
- [12] Annisa, "Algoritma C4.5 Dalam Data Mining: Pengertian, Langkah dan Keunggulannya," 28 september, 2023. <https://fikti.umsu.ac.id/algoritma-c4-5-dalam-data-mining-pengertian-langkah-dan-keunggulannya/>.
- [13] I. Iddrus and D. W. Sari, "Penerapan Data Mining Menggunakan Algoritma Decision Tree C4.5 Untuk Memprediksi Mahasiswa Drop Out Di Universitas Wiraraja," *J. Adv. Res. Inform.*, vol. 1, no. 02, pp. 1–7, 2023, doi: 10.24929/jars.v1i02.2684.
- [14] S. Dwiasnati, W. Gunawan, R. R. Oprasto, B. O. Lubis, and B. Santoso, *Algoritma dan Pemrograman Implementasi Menggunakan Python*. Bandar Lampung: CV. Keranjang Teknologi Media, 2023.
- [15] I. Carolina, B. Lubis, A. Supriyatna, and R. Komarudin, "Application K-Nearest Neighbor Method with Particle Swarm Optimization for Classification of Heart Disease," in *ICAISD 2023*, 2024, no. Icaisd 2023, pp. 181–186, doi: 10.5220/0012446200003848.
- [16] P. P. Haryoto, H. Okprana, and I. S. Saragih, "Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," *TIN Terap. Inform. Nusantara*, vol. 2, no. 5, pp. 358–364, 2021.