

Penerapan Algoritma C4.5 untuk Prediksi Risiko Turnover (Pergantian Karyawan) Menggunakan Data Kinerja dan Kepuasan Kerja

Ginta Khairunisa¹, Farah Fauziyyah Hasyim², Andi Aprilia Sidiuzdah³, Heri Purnomo⁴, Adia Nanda⁵, Annida Purnamawati⁶

¹²³⁴⁵⁶Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika (UBSI)

Jl. Kramat Raya No.98 Jakarta, Indonesia

Email: ¹19251332@bsi.ac.id, ²19251205@bsi.ac.id, ³19251601@bsi.ac.id, ⁴19251592@bsi.ac.id,

⁵19251543@bsi.ac.id, ⁶annida.npr@bsi.ac.id

Abstrak

Tingkat pergantian karyawan yang tinggi menyebabkan kerugian yang besar dari segi finansial dan operasional bagi perusahaan. Penelitian ini bertujuan untuk menciptakan model prediktif yang efisien dalam mengidentifikasi pegawai yang berisiko tinggi mengalami *Turnover* dengan memanfaatkan algoritma klasifikasi C4.5 yang berbasis pada pohon keputusan. Data yang digunakan berasal dari dataset publik *HR-Employee-Attrition* yang diambil dari platform Kaggle, yang berisi 1.470 entri serta 35 atribut, mencakup faktor demografis, kinerja, dan tingkat kepuasan. Proses pemodelan dilaksanakan dengan menggunakan perangkat lunak *RapidMiner*, yang mencakup langkah-langkah persiapan data dan validasi silang. Hasil dari pengujian model menunjukkan rata-rata akurasi sebesar 83,67%. Analisis kinerja kelas mengindikasikan adanya masalah ketidakseimbangan kelas; model memiliki *Recall* yang sangat baik untuk kelas "Tidak *Resign*" dengan persentase 95,78%. Namun sangat rendah untuk kelas "*Resign*" yaitu 20,68%. Model berbasis keputusan menunjukkan bahwa variabel umur merupakan atribut penyaring utama, diikuti oleh Pendapatan Bulanan dan Lama Sejak Promosi Terakhir. Kesimpulan dari studi ini adalah meskipun C4.5 memberikan akurasi yang tinggi secara keseluruhan, masih diperlukan penanganan atas ketidakseimbangan kelas untuk meningkatkan kemampuan dalam mengidentifikasi kejadian *Turnover* yang sebenarnya.

Kata kunci: C4.5, Prediksi Turnover, Penambangan Data, Pohon Keputusan, Manajemen SDM, RapidMiner

IMPLEMENTATION OF THE C4.5 ALGORITHM FOR TURNOVER RISK PREDICTION USING PERFORMANCE AND JOB SATISFACTION

Abstract

High employee turnover causes significant losses from financial and operational perspectives for companies. This study aims to create an efficient predictive model for identifying employees at high risk of turnover by utilizing the C4.5 classification algorithm based on the decision tree method. The data used comes from the public HR-Employee-Attrition dataset acquired from the Kaggle platform, containing 1470 entries and 35 attributes, covering demographic factors, performance, and job satisfaction levels. The modeling process was carried out using the RapidMiner software, including data preparation steps and cross-validation. The results of the model testing show an average accuracy of 83.67%. Class performance analysis indicates an issue of class imbalance; the model demonstrates excellent Recall for the "No Resign" class at 95.78%, but very low Recall for the "Resign" class at 20.68%. The decision-based model reveals that the Age variable is the primary distinguishing attribute, followed by Monthly Income and Years Since Last Promotion. The conclusion of this study is that although C4.5 provides high overall accuracy, handling the class imbalance is still necessary to enhance the capability of identifying actual turnover events.

Keywords: C4.5, Turnover Prediction, Data Mining, Decision Tree, HR Management, RapidMiner

1. PENDAHULUAN

Employee Turnover yang tinggi menimbulkan gangguan operasional dan kebocoran finansial kronis (Pratama et al., 2025). Setiyadi et al. menyatakan bahwa kehilangan karyawan merupakan masalah yang sangat penting dalam konteks

dinamika organisasi karena dampaknya yang signifikan terhadap produktivitas dan stabilitas tenaga kerja. Dampak kehilangan karyawan tidak hanya terbatas pada biaya penggantian, tetapi juga dapat berakibat pada kerusakan reputasi organisasi, mengganggu pencapaian tujuan, dan mengganggu adaptasi organisasi terhadap perubahan lingkungan,

yang dapat berakibat pada kehilangan kompetitivitas. Dalam era digital saat ini, hal tersebut dapat berakibat pada kehilangan pengetahuan dan keterampilan yang telah diakumulasikan, serta mengganggu proses operasional yang telah terintegrasi dengan teknologi (2024).

Oleh karena itu, keahlian organisasi dalam memikirkan serta mengurangi dampak buruk dari pergantian karyawan merupakan faktor penting untuk menjaga kemampuan bersaing serta kestabilan dalam jangka waktu yang panjang.

Dalam usaha menanggulangi masalah tersebut, bidang ilmu *Data Mining* memberikan penawaran solusi berupa prediksi yang berbasis pada data. *Data Mining* sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional basis data yang besar. Menurut Sulaiman Asyuti, Ahmad Aji Setyawan *Data Mining* merupakan suatu proses pengalihan data atau penyaringan data dengan memanfaatkan kumpulan data yang memiliki ukuran cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut (2022). Konsep dasar dari database adalah kumpulan dari catatan, atau potongan dari pengetahuan (Khosmi Pane et al., 2022). Salah satu peran utama basis data adalah efisiensi dalam penyimpanan dan pengelolaan data (Aulia et al., 2023).

Dengan melakukan pengkajian pada data terdahulu milik pegawai mulai dari data kependudukan, catatan upah, kinerja kerja, sampai pada tingkat kesenangan dalam bekerja pola-pola tersembunyi yang mengarah pada keputusan untuk resign dapat diketahui. Kajian ini memilih algoritma C4.5 karena kelebihanannya dalam menciptakan model berupa pohon keputusan (*decision tree*) yang sangat mudah untuk dipahami.

Algoritma C4.5 yaitu sebuah Algoritma yang digunakan untuk membangun *decision tree* pengambilan keputusan untuk memilih atribut akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada (Yusuf Maulana et al., 2022). Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki (Nasrullah, 2021). Cara untuk membuat model *Decision Tree* adalah dengan memecah data ke dalam kelompok yang lebih kecil berdasarkan atribut di dalam data (Ramadhon et al., 2024).

Kejelasan dari model ini menjadi hal yang penting, yang memungkinkan para pengelola SDM untuk memahami dengan pasti faktor-faktor mana saja yang paling berpengaruh dalam memicu terjadinya risiko *turnover*, dan tidak hanya sekedar menerima hasil perkiraan. Data set yang dipakai

dalam penelitian ini yaitu *HR-Employee-Attrition* yang diperoleh dari sebuah wadah bernama Kaggle. Pertimbangan menggunakan kumpulan data ini dilandasi dengan kelengkapan atribut yang ada di dalamnya, yang mencakup berbagai macam faktor penentu terjadinya *turnover* (seperti Usia, Penghasilan Per Bulan, Lama Waktu Promosi Terakhir). Meskipun terdapat klasifikasi lainnya seperti *Naïve Bayes* atau *K-Nearest Neighbor* (K-NN) yang telah dipakai dalam penelitian yang sejenis, muncul kebutuhan untuk menguji seberapa efektifkah C4.5 dalam data set yang rumit serta tidak seimbang (*imbalanced*) ini. Secara khusus, studi ini akan menggarisbawahi bagaimana kinerja model dalam mengenali kelompok minoritas (yaitu kasus *turnover* yang sebenarnya), yang menjadi hambatan utama dalam kegiatan prediksi di bidang HR.

2. METODE PENELITIAN

Pendekatan kuantitatif diimplementasikan dalam metodologi penelitian ini, yang memanfaatkan teknik *Data Mining* untuk mengembangkan serta memvalidasi model prediktif. Pendekatan kuantitatif sendiri adalah suatu konsep yang melibatkan diri pada perhitungan atau angka (Prayogi, 2022). *Data mining* adalah suatu proses penambangan informasi penting dari suatu data (Sudarsono et al., 2021). Langkah-langkah penelitian yang dilaksanakan secara terstruktur, sebagaimana diilustrasikan dalam tahapan berikut:

2.1. Tahapan Penelitian

Pada Gambar 1. ditunjukkan tahapan penelitian yang terdiri dari tujuh tahap utama, yaitu:



Gambar 1. Tahap Penelitian

Tahap awal dimulai dengan mengidentifikasi masalah, dan tujuannya adalah untuk mengembangkan pemahaman yang mendalam tentang risiko pergantian karyawan dan bagaimana teknologi *Data Mining* dapat digunakannya secara efisien dan tepat. Selanjutnya, tahap tinjauan pustaka, pada langkah ini berbagai kerangka teori yang diperoleh dari jurnal, buku, dan sumber lainnya dikaji, terutama yang berkaitan dengan Algoritma C4.5, pergantian karyawan, dan teknik klasifikasi. Klasifikasi adalah teknik *data mining* yang menetapkan kategori pada kumpulan data untuk membantu dalam memprediksi dan analisis yang lebih akurat (Aprilia et al., 2021).

2.2. Pengumpulan Data

Tahap selanjutnya melibatkan pengumpulan data. Menurut Maesaroh et al. Teknik pengumpulan data merupakan komponen fundamental dalam proses penelitian yang memungkinkan peneliti untuk memperoleh informasi yang diperlukan guna menjawab pertanyaan penelitian dan mencapai

tujuan penelitian yang telah ditetapkan (2025). Dalam penelitian ini, sumber data berasal dari kumpulan data *HR-Employee-Attrition* yang dapat diakses secara terbuka, yang bersumber dari platform Kaggle. Alih-alih mengandalkan kuesioner, penelitian ini memanfaatkan data sekunder yang ada,

- a. Jumlah Data: Dataset ini terdiri dari 1470 *record* karyawan.
- b. Atribut: Mencakup 35 atribut, termasuk faktor demografi, kinerja, kompensasi, dan kepuasan kerja.
- c. Variabel Target: Variabel target adalah *Attrition* (Yes/No).

2.3. Pra-pemrosesan Data

Tahap selanjutnya melibatkan pra-pemrosesan data, di mana informasi yang terkumpul menjalani prosedur penataan dan penyempurnaan. Menurut Haq et al. Tujuan utama pra-pemrosesan adalah untuk membersihkan dan mempersiapkan data sehingga lebih mudah dianalisis dan menghasilkan hasil yang lebih akurat (2024). Data dimuat ke *RapidMiner* menggunakan operator *Read CSV* dengan *Import Configuration Wizard* untuk memastikan data dibaca dengan separator yang tepat.

- a. Penyesuaian Tipe Data: Kolom *Attrition*, *Gender*, dan *OverTime* diubah menjadi tipe *Binominal*, dan kolom kategori lainnya (*Business Travel*, *Job Role*, dll.) diubah menjadi *Polynomial*.
- b. Pemilihan Atribut dan Penetapan Peran: Atribut non-prediktif seperti *Employee Count* dan *Standard Hours* dieliminasi, dan atribut *Attrition* ditetapkan sebagai label (variabel target).

2.4. Pengujian Algoritma dan Validasi

- a. Algoritma C4.5
Penelitian ini menggunakan algoritma *Decision Tree* C4.5. Algoritma ini dipilih karena kemampuannya menghasilkan pohon keputusan yang mudah diinterpretasikan dengan memanfaatkan metrik *Information Gain Ratio* untuk menentukan atribut mana yang paling efektif untuk membagi data pada setiap simpul (*node*).
- b. Proses Pemodelan di *RapidMiner*

Model dibangun menggunakan alur kerja sebagai berikut:

Representasi alur RapidMiner: Read CSV → Set Role → Select Attributes → Cross Validation (di dalamnya: Decision Tree → Apply Model → Performance)

c. Validasi Model

Metode *Cross-Validation* digunakan untuk mengevaluasi kinerja model secara objektif dan memitigasi risiko *overfitting*. Metode ini membagi dataset menjadi k bagian (umumnya 10-fold); model dilatih pada k-1 bagian data dan diuji pada 1 bagian yang tersisa, proses ini diulang hingga semua bagian digunakan sebagai data uji.

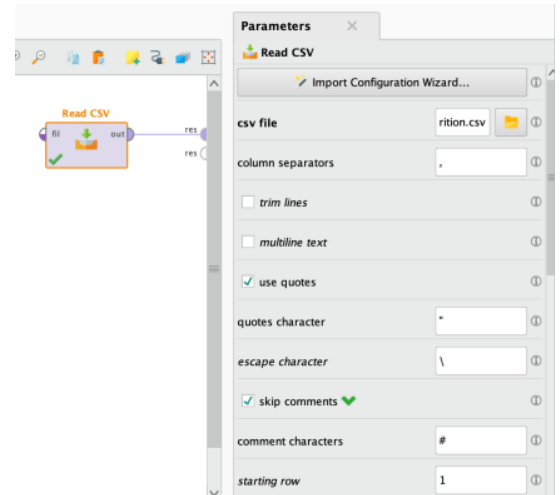
2.5. Evaluasi Kinerja

Kinerja model diukur menggunakan operator *Performance* di *RapidMiner*, dengan fokus pada metrik berikut: Akurasi yaitu Persentase prediksi yang benar secara keseluruhan. *Precision* dan *Recall* yaitu Metrik spesifik kelas yang sangat penting untuk mengevaluasi kemampuan model dalam memprediksi kelas minoritas (*Attrition 'Yes'*), terutama *Recall Yes* untuk mengukur kemampuan model mendeteksi kasus *turnover* yang sebenarnya.

3. HASIL DAN PEMBAHASAN

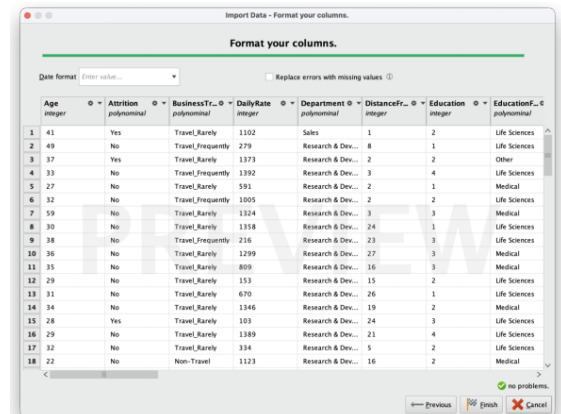
3.1. Pengumpulan Data

Tahap pengumpulan data merupakan langkah penting yang menentukan kualitas dan relevansi model prediktif yang akan dikembangkan. Tujuan Pengumpulan Data ini adalah untuk mendapatkan kumpulan data historis yang lengkap, yang mencakup beragam elemen yang telah divalidasi secara empiris untuk mempengaruhi keputusan karyawan terkait pengunduran diri atau retensi. Retensi karyawan adalah strategi organisasi untuk mempertahankan karyawan yang potensial dimiliki oleh organisasi untuk tetap loyal kepada organisasi (Harvida, 2020). Data ini bertindak sebagai dasar pengetahuan yang akan digunakan Algoritma C4.5 untuk mengidentifikasi pola tersembunyi yang memicu pergantian karyawan, sehingga memungkinkan pembentukan model prediktif yang akurat.



Gambar 2. Operator *Read CSV* Saat Memuat File *Attrition*

Pada Gambar 2 adalah membuka file *Attrition.csv* menggunakan *Import Configuration Wizard*. Wizard membantu *RapidMiner* mengenali struktur file secara otomatis mulai dari separator koma, penggunaan tanda kutip, karakter *escape*, hingga baris awal data. Tujuannya agar file dapat dibaca dengan format yang tepat tanpa harus mengatur semuanya secara manual.

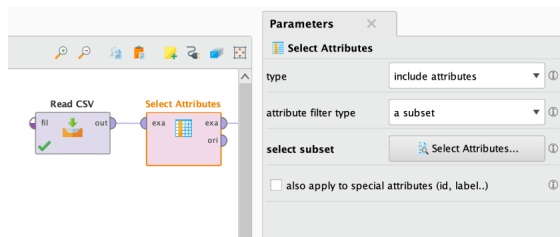


Gambar 3. Mengatur dan mengubah tipe data

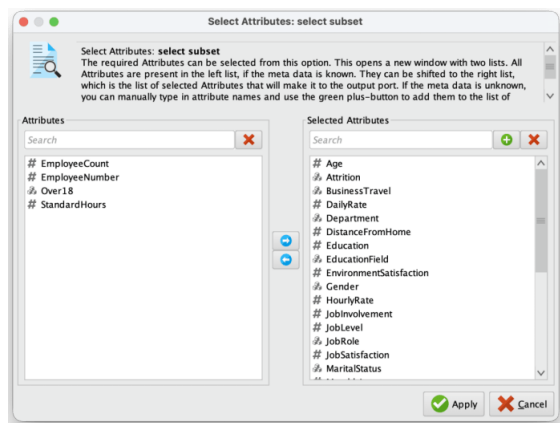
Sumber Data yang digunakan adalah data sekunder, yaitu dataset publik *HR-Employee-Attrition* yang diperoleh dari platform *Kaggle*. Spesifikasi Data: Jumlah *Entri Dataset* terdiri dari 1470 *record* (baris), yang merepresentasikan data dari 1470 karyawan. Jumlah Atribut Meliputi 35 atribut yang kaya akan informasi mencakup dimensi penting seperti Faktor Demografi (*Age, Gender, Marital Status*), Faktor Organisasi (*Job Role, Job Level, Total Working Years, Years Since Last Promotion*), dan Faktor Kepuasan (*Environment Satisfaction, Job Satisfaction*). Variabel Target yang menjadi fokus klasifikasi adalah *Attrition*, dengan dua kelas: "Yes" (*turnover* atau *resign*) dan "No" (karyawan bertahan).

3.2. Pembersihan Data (*Data Cleaning*)

Pembersihan data adalah tahapan yang melibatkan identifikasi, perbaikan, serta penghapusan kesalahan dan ketidakakuratan dalam data guna meningkatkan kualitas dan keandalan data (Rizki et al., 2024). Tujuan pembersihan data adalah untuk menyempurnakan dan menyiapkan dataset yang belum disempurnakan, sehingga memungkinkan pemanfaatannya secara efisien oleh algoritma klasifikasi. Prosedur ini melibatkan penentuan dan penghapusan atribut yang tidak diperlukan atau atribut yang kurang memiliki variasi nilai, karena atribut tersebut tidak membantu proses pembelajaran. Untuk meningkatkan kualitas data yang dimasukkan dan memusatkan proses pembelajaran model pada atribut prediktif yang benar-benar signifikan, tujuannya adalah untuk memfasilitasi pembuatan model yang beroperasi lebih efektif.



Gambar 4. Konfigurasi Operator *Select Attributes* untuk Pembersihan Data



Gambar 5. Pemilihan *Attributes*

Dalam dataset *HR-Employee Attrition*, terdapat atribut-atribut tertentu yang dianggap konstan, yang menunjukkan bahwa setiap 1470 entri data memiliki nilai yang sama untuk atribut-atribut ini. Atribut-atribut yang tidak berubah ini tidak mampu membedakan antara karyawan yang meninggalkan perusahaan (“*Yes*”) dan mereka yang bertahan (“*No*”). Atribut-atribut spesifik yang dikecualikan dari proses analisis menggunakan operator *Select Attributes* di RapidMiner meliputi (*EmployeeCount*, *EmployeeNumber*, *StandardHours*, dan *Over18*). Untuk menangani data yang hilang telah diverifikasi bahwa kumpulan data ini sepenuhnya bebas dari

nilai yang hilang, artinya tidak diperlukan prosedur imputasi data khusus. Proses penghilangan atribut yang tidak relevan ini dapat dilihat pada konfigurasi operator *Select Attributes* di *RapidMiner* pada Gambar 4 dan Gambar 5.

3.3. Transformasi Data (*Data Transformation*)

Tahap transformasi data dilakukan untuk memastikan bahwa setiap atribut dalam dataset memiliki format tipe data yang tepat dan memenuhi persyaratan Algoritma C4.5. Langkah ini sangat penting mengingat C4.5 mencapai kinerja puncak dengan data kategorikal (*binomial/polinomial*) dan numerik. Tujuan Transformasi: Mengubah format data mentah menjadi format yang dapat diproses oleh algoritma klasifikasi.

3.3.1. Konversi Tipe Data

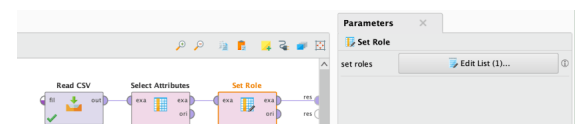
Transformasi tipe data ditangani secara bertahap dalam *RapidMiner*, yang mencakup Panduan Konfigurasi Impor serta operator lain untuk penyesuaian yang tepat.

a. *Binomial*: Kolom yang hanya memiliki dua nilai kategori, yaitu *Attrition* (target: *Yes/No*), *Gender*, dan *OverTime*, diubah menjadi tipe data *Binomial*.

b. *Polynomial*: Kolom kategori yang memiliki lebih dari dua nilai unik (*multinomial*), seperti *Business Travel*, *Department*, *EducationField*, *JobRole*, dan *Marital Status*, diubah menjadi tipe data *Polynomial*.

c. *Integer/Real*: Kolom yang mengandung nilai kuantitatif (numerik) seperti *Age*, *MonthlyIncome*, dan *DailyRate* dipertahankan atau dikonversi menjadi tipe data *Integer* atau *Real*.

3.3.2. Penetapan Peran (*Set Role*)



Gambar 6. Konfigurasi Tipe Data dan Penetapan Peran (*Set Role*)



Gambar 7. Atribut *Attrition* dipilih dan diberi peran label

Set Role yang digunakan untuk mengatur peran atribut. Atribut *Attrition* dipilih dan diberi peran label, menjadi target yang akan diprediksi oleh model.

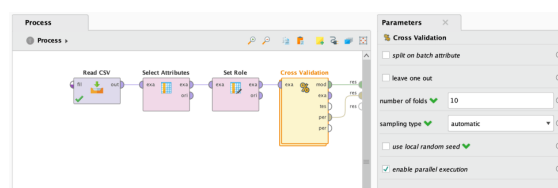
a. Atribut *Attrition* ditetapkan sebagai Label (variabel dependen/target) karena merupakan variabel yang akan diprediksi oleh model.

b. Atribut *Employee Number* ditetapkan sebagai ID dan diabaikan selama proses pelatihan model, karena fungsinya hanya sebagai pengenalan unik dan tidak berkontribusi pada pola klasifikasi.

Proses penetapan tipe data dan peran ini adalah langkah wajib sebelum data masuk ke modul *Cross Validation*.

3.4. Proses Data Mining

Tahap ini merupakan inti dari penelitian, di mana algoritma klasifikasi diimplementasikan pada data yang telah ditransformasi. Perangkat lunak *RapidMiner* digunakan untuk melakukan pemodelan dan pengujian. Tujuannya adalah untuk membuat model yang meramalkan pergantian karyawan dengan memanfaatkan Algoritma C4.5, dan untuk menilai seberapa baik kinerja model melalui penerapan metode validasi standar.



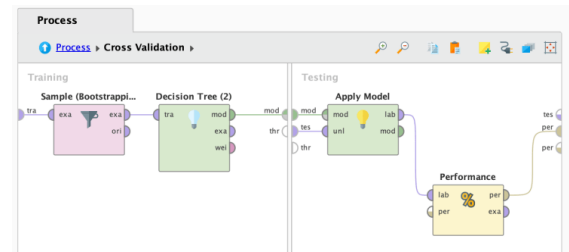
Gambar 8. Alur Proses Data Mining Menggunakan *RapidMiner*

3.4.1. Konversi Tipe Data

Model klasifikasi dibangun menggunakan operator Pohon Keputusan, yang mewujudkan

representasi Algoritma C4.5. Algoritma ini dipilih karena kemampuannya menghasilkan struktur pohon keputusan yang mudah dipahami, sehingga terbukti bermanfaat untuk menganalisis faktor-faktor yang berkontribusi terhadap pergantian karyawan.

3.4.2. Metode Validasi



Gambar 9. Training dan Testing pada Operator *Cross Validation*

Cross validation atau estimasi rotasi merupakan suatu teknik validasi model untuk menilai dan mengetahui berapa hasil statistik analisis yang akan menggeneralisasi himpunan data independen (Tuntun et al., 2022). *Cross Validation* K-Fold digunakan untuk memastikan penilaian yang tidak memihak terhadap kinerja model dan mengurangi potensi *overfitting*. Metode ini melibatkan pembagian dataset menjadi k segmen (K Fold). Studi ini menggunakan *Cross Validation* 10-Fold, yang membagi data menjadi total 10 subset yang berbeda. Dengan menggunakan sembilan subset data, model dilatih, kemudian diuji menggunakan subset yang tersisa. Untuk memastikan setiap subset digunakan sebagai data uji setidaknya sekali, prosedur ini dilakukan 10 kali.

3.4.3. Alur Proses *RapidMiner*

Seluruh tahapan, mulai dari pemrosesan hingga validasi, dikemas dalam alur proses tunggal di *RapidMiner* (lihat Gambar 8). Alur kerja utama terdiri dari:

a. *Input Data*: Data yang telah ditransformasi dimasukkan ke operator *Cross Validation*.

b. *Training Subprocess* (Pelatihan): Di dalam CV, operator *Decision Tree* digunakan untuk melatih model.

c. *Testing Subprocess* (Pengujian): Operator *Apply Model* digunakan untuk menerapkan model yang telah dilatih ke data pengujian, dan operator *Performance* digunakan untuk menghitung metrik kinerja.

Alur lengkap proses *Data Mining* ini memastikan bahwa model diuji secara

menyeluruh sebelum hasil akhirnya diinterpretasikan.

3.5. Evaluasi Pola

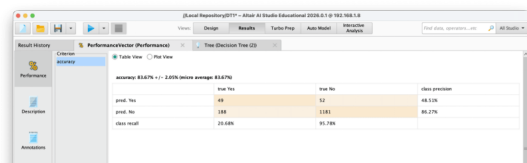
Tahap akhir penelitian ini adalah evaluasi pola, yang melibatkan penilaian efektivitas model melalui metrik kuantitatif dan menguraikan pola keputusan yang dihasilkan oleh algoritma C4.5.

3.5.1. Analisis Kinerja Model dan Keterbatasan Deteksi

Operator Kinerja digunakan untuk mengukur efektivitas model, menghasilkan *Performance Vector* dan *Confusion Matrix* yang berasal dari proses *Cross Validation*. Hasil ini menunjukkan daya prediktif model, yang mencakup presisi umum (akurasi) dan kemampuan spesifik kelasnya (*recall*).

Tabel 1. Ringkasan Matriks Kinerja Model C4.5

Metrik	Nilai (%)	Interpretasi Kinerja
Akurasi (<i>Overall</i>)	83,67% ($\pm 2,05\%$)	Akurasi rata-rata yang menunjukkan kemampuan klasifikasi umum model.
<i>Recall</i> (<i>Resign/Yes</i>)	20,68%	Kemampuan model mengidentifikasi kasus <i>turnover</i> yang sebenarnya (kelas minoritas).
<i>Recall</i> (<i>Tidak Resign/No</i>)	95,78%	Kemampuan model mengidentifikasi karyawan yang bertahan (kelas mayoritas).



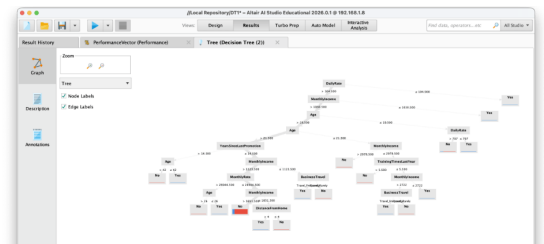
Gambar 10. *Performance Vector*

Model C4.5 menunjukkan tingkat akurasi umum yang baik, mencapai angka 83,67%. Walaupun begitu, angka ini harus diartikan sama dengan nilai *Recall* 'Yes' yang ternyata hanya sebesar 20,68%. Kinerja deteksi yang kurang memuaskan pada kelas

'Yes' (yaitu kasus keluarnya karyawan) menandakan adanya bias yang signifikan ke arah kelas mayoritas, yaitu 'No'. Jika Merujuk pada *Confusion Matrix*, dari seluruh kejadian *turnover* yang ada, sebanyak 188 kejadian tidak diketahui dengan benar sebagai potensi risiko yang tinggi. Kelemahan dalam mendeteksi ini menandakan bahwa model perlu ditingkatkan kapasitasnya dalam mengenali kasus-kasus yang jarang terjadi (minoritas) agar dapat bekerja secara optimal sebagai sistem deteksi dini.

3.5.2. Interpretasi Pola Keputusan (*Decision Tree*)

Pola keputusan yang dibuat melalui C4.5 dinilai untuk menentukan atribut dan aturan yang paling berdampak.



Gambar 11. Struktur Pohon Keputusan Model C4.5

Pola Penentu Atribut Utama: Pola pada pohon dimulai dengan atribut yang memiliki *Information Gain Ratio* terbesar. Hasil analisis pola mengungkapkannya hierarki atribut sebagai berikut:

- a. *Age* (Usia): Terletak sebagai simpul akar (*root node*). Ini memperkuat gagasan bahwa faktor usia adalah atribut tunggal yang paling mendasar dan dominan dalam memicu keputusan *turnover*.
- b. *Monthly Income* (Pendapatan Bulanan): Berfungsi sebagai pemisah utama setelah *Age*. Pola *turnover* berisiko tinggi sering terbentuk ketika karyawan muda dikombinasikan dengan tingkat pendapatan bulanan yang berada di bawah ambang batas tertentu.
- c. *Years Since Last Promotion*: Atribut ini menjadi faktor penting dalam percabangan lebih lanjut, yang menunjukkan bahwa lamanya waktu tanpa pengakuan atau kemajuan karier (promosi) merupakan pola yang kuat dalam keputusan pengunduran diri.

Derivasi Pola Keputusan: Pola-pola yang ditemukan oleh model dapat diterjemahkan menjadi aturan IF-THEN yang memberikan panduan

strategis bagi Manajemen SDM. Contoh pola risiko tinggi yang teridentifikasi adalah:

Pola Resiko Tinggi: Jika umur ≤ 31 tahun dan pendapatan bulanan dibawah ambang batas yang ditentukan model, maka karyawan diprediksi memiliki Risiko *Turn Over* Tinggi (*Yes*).

Analisis pola ini memberikan wawasan taktis untuk intervensi retensi yang ditargetkan pada segmen karyawan yang berada dalam kombinasi atribut risiko tinggi tersebut.

4. KESIMPULAN

Penelitian ini secara efektif memanfaatkan algoritma C4.5, yang beroperasi pada pohon keputusan, untuk membuat model prediksi risiko *turnover* karyawan, dengan memanfaatkan dataset *HR-Employee-Attrition*. Hasil pengujian dan evaluasi model menyoroiti tiga aspek utama.

Kinerja Model Secara Umum: Model C4.5 mencapai akurasi keseluruhan rata-rata sebesar 83,67 persen. Tingkat akurasi ini diperkuat oleh kemampuannya yang luar biasa dalam memprediksi kelas dominan, yaitu karyawan yang bertahan, dibuktikan dengan nilai *Recall "No"* sebesar 95,78 persen.

Keterbatasan Deteksi: Model menghadapi keterbatasan kemampuan deteksi yang signifikan pada kelas minoritas (*turnover*), dengan nilai *Recall 'Yes'* hanya 20,68%. Keterbatasan ini disebabkan oleh adanya ketidakseimbangan kelas (*class imbalance*) pada data, yang membuat model bias dan tidak optimal dalam mengidentifikasi kasus *turnover* yang sebenarnya.

Atribut Penentu Pola: Analisis Pohon Keputusan C4.5 secara transparan mengidentifikasi hirarki atribut yang paling berpengaruh. *Age* (Umur) ditetapkan sebagai atribut paling dominan (simpul akar), diikuti oleh *Monthly Income* (Pendapatan Bulanan) dan *Years Since Last Promotion*. Pola risiko *turnover* tertinggi teridentifikasi pada kombinasi karyawan muda dengan pendapatan di bawah ambang batas yang ditentukan model.

5. DAFTAR PUSTAKA

- Pratama, A., Mujiyono, S., & Sanjaya, U. P. (2025). Analisis Klasifikasi Resign Karyawan dengan Random Forest. *JURNAL UNITEK*, 18(1), 156–166. <https://doi.org/10.52072/unitek.v18i1.1415>
- Setiyadi, P., Prayogi, M. N., & Solichin, A. (2024). OPTIMALISASI PREDIKSI KEHILANGAN KARYAWAN MENGGUNAKAN TEKNIK RFE, SMOTE, DAN ADABOOST. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(4), 2131–2145. <https://doi.org/10.29100/jupi.v9i4.5642>
- Sulaiman Asyuti, Ahmad Aji Setyawan. (2022). DATA MINING DALAM PENGGUNAAN PRESENSI KARYAWAN DENGGA CLUSTER MEANS. *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, 1(1), 1–10. <https://doi.org/10.59024/jiti.v1i1.6>
- Khomsni Pane, S. Y., Ramadhan, N. G., & Adhinata, F. D. (2022). Perancangan basis data menggunakan normalisasi tabel pada perusahaan dagang barokah abadi. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, 2(2), 90–96. <https://doi.org/10.20895/dinda.v2i2.563>
- Aulia, B. W., Rizki, M., Prindiyana, P., & Surgana, S. (2023). Peran Krusial Jaringan Komputer dan Basis Data dalam Era Digital. *JUSTINFO | Jurnal Sistem Informasi Dan Teknologi Informasi*, 1(1), 9–20. <https://doi.org/10.33197/justinfo.vol1.iss1.2023.1253>
- Yusuf Maulana, Riki Winanjaya, & Fitri Rizki. (2022). Penerapan Data Mining dengan Algoritma C.45 Dalam Memprediksi Penjualan Tempe. *Bulletin of Computer Science Research*, 2(2), 53–58. <https://doi.org/10.47065/bulletincsr.v2i2.163>
- Nasrullah, A. H. (2021). IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS. *JURNAL ILMIAH ILMU KOMPUTER*, 7(2), 45–51. <https://doi.org/10.35329/jiik.v7i2.203>
- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, 3(2), 1860–1874. <https://doi.org/10.30997/karimahtauhid.v3i2.11952>
- Prayogi, A. (2022). TELAAH KONSEPTUAL PENDEKATAN KUANTITATIF DALAM SEJARAH. *Kalpataru: Jurnal Sejarah Dan Pembelajaran Sejarah*, 8(1). <https://doi.org/10.31851/kalpataru.v8i2.8970>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER. *JBASE - Journal of Business and*

- Audit Information Systems, 4(1).
<https://doi.org/10.30813/jbase.v4i1.2729>
- Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *SISTEMASI*, 10(1), 163.
<https://doi.org/10.32520/stmsi.v10i1.1129>
- Maesaroh, I., Miladia, U. A., Fithriyan, M., & Nulhakim, L. (2025). TEKNIK PENGUMPULAN DATA DALAM PENELITIAN. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 10(02), 315–325.
<https://doi.org/10.23969/jp.v10i02.27110>
- Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, 13(1), 1079–1084.
<https://doi.org/10.33395/jmp.v13i1.13937>
- Harvida, C. W. D. A. (2020). FAKTOR YANG MEMPENGARUHI TURNOVER KARYAWAN DAN STRATEGI RETENSI SEBAGAI PENCEGAHAN TURNOVER KARYAWAN: SEBUAH TINJAUAN LITERATUR. *JIANA (Jurnal Ilmu Administrasi Negara)*, 18(2), 13–23.
<https://doi.org/10.46730/jiana.v18i2.7926>
- Rizki, I. N., Prayoga, D., Puspita, M. L., & Huda, M. Q. (2024). IMPLEMENTASI EXPLORATORY DATA ANALYSIS UNTUK ANALISIS DAN VISUALISASI DATA PENDERITA STROKE KALIMANTAN SELATAN MENGGUNAKAN PLATFORM TABLEAU. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1).
<https://doi.org/10.23960/jitet.v12i1.3856>
- Tuntun, R., Kusrini, K., & Kusnawi, K. (2022). Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2111.
<https://doi.org/10.30865/mib.v6i4.4681>