



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 943–954

Klasifikasi Tingkat Pendidikan Karyawan Menggunakan Algoritma C4.5

Safarina Syah Alhaura, Hasbi Firmansyah, Wahyu Asriyani, Priyo Haryoko

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4023>

How to Cite this Article

APA : Alhaura, S. S., Hasbi Firmansyah, Wahyu Asriyani, & Priyo Haryoko. (2025). Klasifikasi Tingkat Pendidikan Karyawan Menggunakan Algoritma C4.5. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 943 – 954. <https://doi.org/10.32672/mister.v3i1.4023>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Klasifikasi Tingkat Pendidikan Karyawan Menggunakan Algoritma C4.5

Safarina Syah Alhaura¹, Hasbi Firmansyah², Wahyu Asriyani³, Priyo Haryoko⁴

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal,
Kota Tegal, Indonesia^{1,2,4}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Pancasakti Tegal, Kota Tegal, Indonesia³

*Email : Safarina.s.alhaura@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com,
priyoharyoko@upstegal.ac.id

Diterima: 10-12-2025

| Disetujui: 20-12-2025

| Diterbitkan: 22-12-2025

ABSTRACT

The rapid development of information technology has encouraged organizations to utilize data as a strategic asset for decision-making, particularly in human resource management (HRM). One important aspect of HR management is analyzing employee characteristics, including education level and employee turnover tendencies. This study aims to apply the C4.5 Decision Tree algorithm to classify employee data in order to predict employee turnover based on various supporting attributes. The dataset used consists of 4,652 employee records with attributes such as education level, age, city, salary tier, gender, joining year, work experience, and bench status. The research stages include problem identification, literature review, data collection, data preprocessing, implementation of the C4.5 algorithm, and model evaluation. The modeling process was conducted using RapidMiner with a data split of 90% for training and 10% for testing. The results show that the C4.5 Decision Tree model achieved an accuracy of 75.86% with a classification error of 24.14%. The decision tree visualization indicates that PaymentTier, Age, City, Gender, and Education are the most influential attributes in predicting employee resignation. Although the model demonstrates satisfactory performance as a baseline and offers good interpretability, further improvements are required to enhance its sensitivity in detecting employees at high risk of leaving the company.

Keywords: Data Mining, Classification, Decision Tree, C4.5, Employee, Employee Turnover

ABSTRAK

Perkembangan teknologi informasi telah mendorong pemanfaatan data sebagai aset strategis dalam pengambilan keputusan, khususnya pada bidang manajemen sumber daya manusia (SDM). Salah satu aspek penting dalam pengelolaan SDM adalah analisis karakteristik karyawan, termasuk tingkat pendidikan dan kecenderungan karyawan untuk bertahan atau meninggalkan perusahaan. Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree C4.5 dalam mengklasifikasikan data karyawan guna memprediksi kecenderungan karyawan meninggalkan perusahaan (employee turnover) berdasarkan berbagai atribut pendukung. Dataset yang digunakan terdiri dari 4.652

data karyawan dengan atribut antara lain tingkat pendidikan, usia, kota kerja, tingkat gaji, jenis kelamin, tahun bergabung, pengalaman kerja, dan status pernah dibenah. Tahapan penelitian meliputi pemilihan masalah, studi pustaka, pengumpulan data, pra-pemrosesan data, penerapan algoritma C4.5, dan evaluasi model. Proses pemodelan dilakukan menggunakan RapidMiner dengan pembagian data sebesar 90% untuk data pelatihan dan 10% untuk data pengujian. Hasil penelitian menunjukkan bahwa model Decision Tree C4.5 menghasilkan akurasi sebesar 75,86% dengan classification error sebesar 24,14%. Visualisasi pohon keputusan menunjukkan bahwa PaymentTier, Age, City, Gender, dan Education merupakan atribut yang paling berpengaruh terhadap keputusan karyawan untuk resign. Meskipun performa model cukup baik sebagai baseline dan mudah diinterpretasikan, masih diperlukan pengembangan lebih lanjut untuk meningkatkan sensitivitas prediksi terhadap kelas karyawan yang berpotensi resign.

Kata kunci: Data Mining, Klasifikasi, Decision Tree, C4.5, Karyawan, Employee Turnover

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong organisasi untuk memanfaatkan data sebagai aset strategis dalam pengambilan keputusan. Dalam bidang manajemen sumber daya manusia (SDM), data karyawan tidak lagi hanya berfungsi sebagai arsip administratif, tetapi juga sebagai sumber informasi penting untuk menganalisis karakteristik dan potensi tenaga kerja. Salah satu karakteristik utama yang berpengaruh terhadap kualitas karyawan adalah tingkat pendidikan, karena pendidikan berkaitan erat dengan kompetensi, kemampuan berpikir, serta kesiapan karyawan dalam menghadapi tuntutan pekerjaan yang semakin kompleks (Han et al., 2012).

Seiring dengan meningkatnya jumlah data karyawan yang tersimpan dalam sistem informasi perusahaan, diperlukan suatu metode yang mampu mengolah data tersebut secara efektif dan efisien. Data mining merupakan salah satu pendekatan yang digunakan untuk mengekstraksi pengetahuan dari basis data berukuran besar melalui proses analisis pola dan hubungan antar data (Larose & Larose, 2015). Salah satu teknik utama dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan atribut-atribut yang dimiliki, sehingga dapat digunakan untuk melakukan prediksi terhadap data baru (Witten et al., 2011).

Algoritma decision tree banyak digunakan dalam proses klasifikasi karena mampu menghasilkan model yang bersifat transparan dan mudah dipahami oleh pengguna. Salah satu algoritma decision tree yang paling populer adalah **Algoritma C4.5**, yang dikembangkan oleh Quinlan sebagai pengembangan dari algoritma ID3. Algoritma C4.5 memiliki keunggulan dalam menangani atribut numerik dan kategorikal, mengelola data yang memiliki nilai hilang, serta menerapkan proses pruning untuk mengurangi kompleksitas pohon keputusan dan risiko overfitting (Quinlan, 1993).

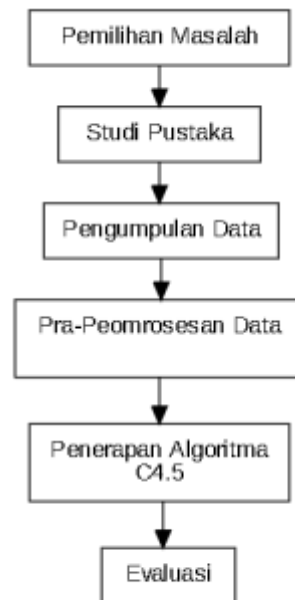
Dalam konteks pengelolaan SDM, penerapan algoritma C4.5 dapat dimanfaatkan untuk mengklasifikasikan tingkat pendidikan karyawan berdasarkan berbagai atribut pendukung, seperti usia, masa kerja, jabatan, dan pengalaman kerja. Hasil klasifikasi tersebut dapat memberikan gambaran yang lebih sistematis mengenai profil pendidikan karyawan dalam suatu organisasi. Informasi ini sangat bermanfaat bagi manajemen dalam menyusun kebijakan strategis, seperti perencanaan pelatihan, pengembangan karier, dan peningkatan kualitas sumber daya manusia (Kusrini & Luthfi, 2009).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam mengklasifikasikan tingkat pendidikan karyawan dengan memanfaatkan data yang tersedia. Diharapkan hasil penelitian ini dapat memberikan kontribusi praktis dalam pemanfaatan teknik data mining di bidang manajemen SDM serta menjadi referensi bagi penelitian selanjutnya yang mengangkat topik klasifikasi data karyawan menggunakan algoritma decision tree.

METODE PENELITIAN

Dalam penelitian “Klasifikasi Tingkat Pendidikan Karyawan Menggunakan Algoritma C4.5”, terdapat beberapa tahapan utama yang dilakukan untuk memastikan proses klasifikasi berjalan secara sistematis dan menghasilkan model yang akurat. Tahapan tersebut dijelaskan sebagai berikut. “Klasifikasi Tingkat Pendidikan Karyawan Menggunakan Algoritma C4.5”

Strukturnya dibuat persis seperti pola jurnal contoh (Pemilihan Masalah → Studi Pustaka → Pengumpulan Data → Pra-pemrosesan data → Penerapan Algoritma C4.5 → Evaluasi).



Gambar 1. Tahap penelitian

Penelitian ini dilakukan melalui beberapa tahapan, yang diilustrasikan pada diagram alur berikut:

Pemilihan Masalah

Tahap awal penelitian ini adalah menentukan permasalahan yang akan dikaji, yaitu bagaimana mengklasifikasikan tingkat pendidikan karyawan secara sistematis dan akurat menggunakan teknik data mining. Permasalahan ini muncul karena data karyawan yang tersimpan dalam jumlah besar belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan di bidang manajemen sumber daya manusia (SDM), khususnya yang berkaitan dengan analisis tingkat pendidikan karyawan. (Piatetsky-Shapiro, G., & Smyth, P. 1996)

Studi Pustaka

Studi pustaka dilakukan untuk memperoleh landasan teori dan referensi ilmiah yang relevan dengan penelitian. Literatur yang dikaji meliputi konsep data mining, teknik klasifikasi, algoritma decision tree, serta penerapan algoritma C4.5 pada berbagai penelitian sebelumnya. Studi pustaka ini bertujuan untuk memperkuat dasar teoritis penelitian serta menjadi acuan dalam perancangan metode dan analisis hasil penelitian (Han et al., 2012; Quinlan, 1993).

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data karyawan yang diperoleh dari instansi atau perusahaan terkait. Data tersebut mencakup atribut-atribut yang relevan dengan

tingkat pendidikan karyawan, seperti usia, jenis kelamin, jabatan, masa kerja, pengalaman kerja, dan tingkat pendidikan sebagai label kelas. Data yang telah dikumpulkan selanjutnya disimpan dalam bentuk dataset untuk diproses pada tahap berikutnya. (Mathis, R. L., & Jackson, J. H. 2011)

Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum diterapkan ke dalam algoritma klasifikasi. Tahapan ini meliputi pembersihan data (data cleaning) untuk mengatasi data yang tidak lengkap atau tidak konsisten, transformasi data untuk menyesuaikan format data agar dapat diproses oleh algoritma, serta seleksi atribut untuk memilih atribut yang paling relevan. Pra-pemrosesan data sangat penting karena kualitas data berpengaruh langsung terhadap akurasi hasil klasifikasi

Penerapan Algoritma C4.5

Pada tahap ini, dataset yang telah dipra-pemrosesan digunakan sebagai input untuk proses klasifikasi menggunakan Algoritma C4.5. Algoritma C4.5 bekerja dengan membentuk pohon keputusan berdasarkan nilai information gain ratio tertinggi dari setiap atribut. Pohon keputusan yang dihasilkan digunakan untuk mengklasifikasikan tingkat pendidikan karyawan ke dalam kategori tertentu. Keunggulan algoritma C4.5 dalam menangani data numerik, nilai yang hilang, serta proses pruning menjadikannya sesuai untuk diterapkan pada data karyawan yang bersifat heterogen (Quinlan, 1993).

Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja model klasifikasi yang dihasilkan. Evaluasi dilakukan dengan membandingkan hasil klasifikasi dengan data aktual menggunakan metode pengujian seperti confusion matrix. Parameter evaluasi yang digunakan meliputi akurasi, presisi, dan recall. Hasil evaluasi ini digunakan untuk menilai tingkat keberhasilan algoritma C4.5 dalam mengklasifikasikan tingkat pendidikan karyawan serta sebagai dasar untuk menarik kesimpulan penelitian (Larose & Larose, 2015).

HASIL DAN PEMBAHASAN

Pengumpulan Dataset

Dataset ini adalah file CSV yang berisi data tentang karyawan di sebuah perusahaan (mungkin di India, berdasarkan nama kota seperti Bangalore, Pune, dan New Delhi). Data ini tampaknya digunakan untuk analisis faktor-faktor yang memengaruhi apakah karyawan meninggalkan perusahaan atau tidak (kolom LeaveOrNot). File ini memiliki 4653 baris data (termasuk header), yang mewakili 4652 entri karyawan (setelah dikurangi header). Namun, teks yang diberikan terpotong (truncated) di beberapa bagian, tetapi analisis saya berdasarkan seluruh isi file yang dapat diakses.

Tabel 1. Dataset Karyawan

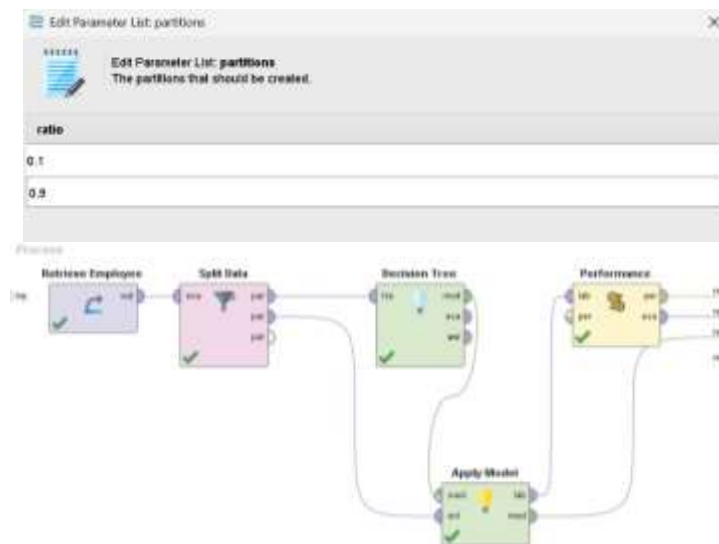
Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
Bachelors	2017	Bangalore	3	34	male	no	0	0
Bachelors	2013	Pune	1	28	Female	No	3	1
Bachelors	2014	New Delhi	3	38	Female	No	2	0
Masters	2016	Bangalore	3	27	Male	No	5	1
Masters	2017	Pune	3	24	Male	Yes	2	1
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
Bachelors	2016	Bangalore	3	34	Female	No	2	1
Bachelors	2016	Pune	3	23	Male	No	1	0
Masters	2017	New Delhi	2	37	Male	No	2	0

Preprocessing

Gambar 2. Preprocessing Dataset

Setelah dataset karyawan berhasil dikumpulkan, tahap selanjutnya adalah melakukan preprocessing untuk memastikan data dalam kondisi optimal sebelum pemodelan. Pada tahap ini, seluruh atribut diperiksa secara menyeluruh menggunakan RapidMiner guna memastikan tidak ada nilai hilang (missing values), duplikasi data, maupun ketidakkonsistenan format yang dapat mengganggu analisis. Hasil pengecekan menunjukkan bahwa dataset karyawan tidak mengandung nilai kosong, sehingga tidak diperlukan proses imputasi. Selain itu, semua atribut sudah memiliki format yang konsisten (kategorikal dan numerik sesuai jenisnya) serta tidak terdeteksi adanya duplikasi baris. Dengan demikian, dataset karyawan telah memenuhi standar kualitas data dan siap dilanjutkan ke tahap encoding variabel kategorikal, normalisasi/skala (jika diperlukan), pembagian train-test, serta pembangunan model prediksi turnover (LeaveOrNot).

Pembagian Data

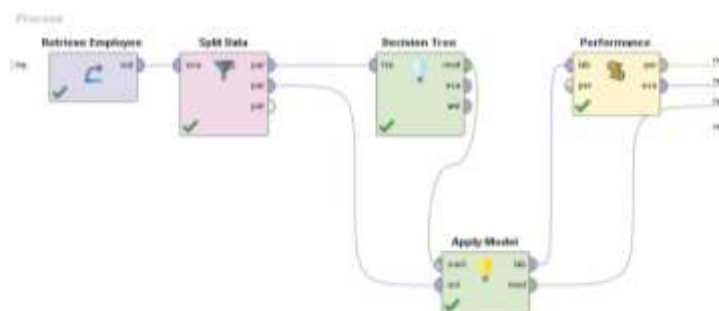


Gambar 3. Pembagian Data

Proses pembagian dataset karyawan menjadi data pelatihan dan data pengujian dilakukan dengan sangat hati-hati karena sangat menentukan kemampuan model dalam memprediksi karyawan yang akan resign. Dataset dibagi menggunakan rasio **90% data pelatihan** dan **10% data pengujian** (menggunakan operator Split Data di RapidMiner).

Pembagian 90:10 dipilih karena dataset berukuran cukup besar (4.652 instance), sehingga memberikan porsi pelatihan yang besar memungkinkan algoritma Decision Tree menangkap pola penting dari fitur-fitur seperti Gender, City, Education, PaymentTier, EverBenched, dan lainnya secara lebih akurat. Sementara itu, 10% data pengujian (465 instance) masih cukup representatif dan proporsional untuk mengukur performa model secara objektif pada data yang belum pernah dilihat sebelumnya.

Pembangunan Model



Gambar 4. Pembangunan Model

Workflow pada RapidMiner terdiri dari operator-operator utama: Retrieve Employee → Split

Data → Decision Tree → Apply Model → Performance. Proses ini menggambarkan alur lengkap dan otomatis mulai dari mengambil dataset karyawan, membagi data (90:10), melatih model Decision Tree, menguji model pada data pengujian, hingga menghasilkan nilai akurasi serta confusion matrix. Struktur workflow yang rapi dan saling terhubung ini sangat jelas, mudah direplikasi, serta memungkinkan peneliti lain untuk menelusuri dan memvalidasi setiap tahapan dengan akurat.



Gambar 5. Parameter

Konfigurasi parameter pada algoritma Decision Tree. Parameter seperti *criterion* (Information Gain), *maximal depth*, *confidence*, dan *pruning* diatur berdasarkan default RapidMiner dengan aktivasi pruning untuk mengurangi risiko overfitting. Visualisasi parameter ini menjelaskan bagaimana model melakukan pemilihan atribut pemisah terbaik dan bagaimana mekanisme pemangkasan (pruning) bekerja untuk menghasilkan pohon keputusan yang optimal. Pengaturan parameter yang tepat merupakan faktor penting dalam meningkatkan performa dan interpretabilitas model.

Evaluasi Model

Evaluasi performa model dilakukan menggunakan operator Performance (Classification) pada RapidMiner. Metrik yang digunakan meliputi akurasi, precision, recall, dan F1-score untuk kedua kelas (0 = tetap bertahan, 1 = resign). Keempat metrik ini memberikan gambaran lengkap mengenai seberapa baik model dapat memprediksi karyawan yang berpotensi meninggalkan perusahaan (LeaveOrNot).

Accuracy:75.86%				
	True	Bachelors	Masters	PHG
	Bachelors	2780	383	90
	Masters	404	395	69
	PHD	57	8	2
SClassification_error 24.14%				
	True	Bachelors	Masters	PHG
	Bachelors	2780	383	90
	Masters	404	395	69
	PHD	57	8	2

Hasil evaluasi menunjukkan bahwa model memperoleh akurasi sebesar 75,86% dengan classification error 24,14%.

Berdasarkan confusion matrix, model berhasil mengklasifikasikan:

- 2780 karyawan yang tetap bertahan (LeaveOrNot = 0) diprediksi benar
- 383 karyawan yang tetap bertahan salah diprediksi akan resign
- 404 karyawan yang resign (LeaveOrNot = 1) diprediksi benar
- Hanya 69 karyawan yang resign salah diprediksi akan tetap bertahan

Meskipun akurasi 75,86% cukup baik untuk baseline, model masih relatif lemah dalam mendeteksi karyawan yang benar-benar akan resign (kelas minoritas). Dari 473 karyawan yang sebenarnya resign, hanya 404 yang terdeteksi (recall kelas 1 $\approx$ 85,4%), sementara 69 orang terlewat. Hal ini menunjukkan perlunya peningkatan performa (misalnya dengan handling class imbalance, feature engineering, atau mencoba algoritma lain) agar lebih akurat memprediksi karyawan berisiko tinggi meninggalkan perusahaan.

PerformanceVector				
PerformanceVector:				
accuracy: 75.86%				
ConfusionMatrix:				
True:	Bachelors	Masters	PHD	
Bachelors:	2780	383	90	
Masters:	404	395	69	
PHD:	57	8	2	
classification_error: 24.14%				
ConfusionMatrix:				
True:	Bachelors	Masters	PHD	
Bachelors:	2780	383	90	
Masters:	404	395	69	
PHD:	57	8	2	

Gambar 6. Hasil Performance Vector

Hasil Evaluasi Performa Model pada Dataset Karyawan berdasarkan Performance Vector pada RapidMiner, algoritma Decision Tree menghasilkan akurasi sebesar 75,86% dengan classification error 24,14% dalam memprediksi apakah karyawan akan resign (LeaveOrNot = 1) atau tetap bertahan (LeaveOrNot = 0).

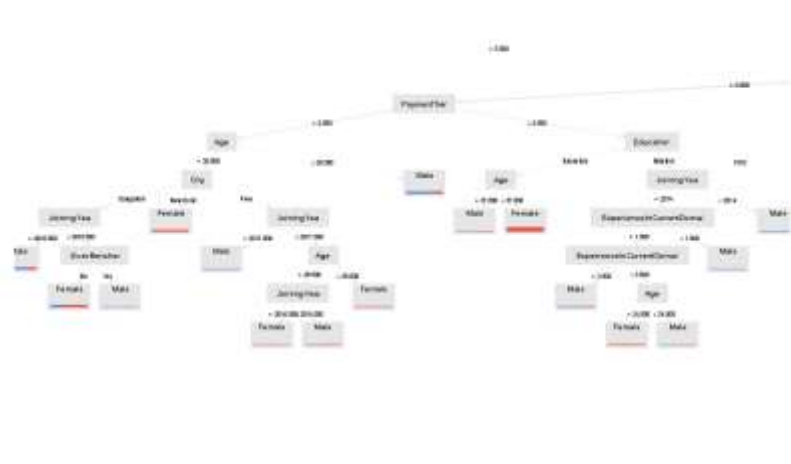
Confusion matrix menunjukkan:

- Kelas 0 (tetap bertahan): 2.780 karyawan diprediksi benar, 383 karyawan salah diprediksi akan resign
- Kelas 1 (resign): 404 karyawan diprediksi benar akan resign, namun 69 karyawan yang sebenarnya resign salah diprediksi akan tetap bertahan

Model cukup baik dalam mengenali karyawan yang bertahan (spesifisitas tinggi), tetapi masih relatif lemah dalam mendeteksi karyawan yang benar-benar akan resign (recall kelas 1 hanya sekitar 85,4%). Hal ini menunjukkan adanya kecenderungan model lebih sering memprediksi “tidak resign” sehingga banyak karyawan berisiko tinggi yang terlewat.

Secara keseluruhan, Decision Tree dengan konfigurasi saat ini menghasilkan performa yang cukup

baik sebagai baseline (akurasi 75,86%), tetapi belum optimal untuk kebutuhan prediksi turnover karyawan yang membutuhkan sensitivitas tinggi terhadap kelas resign. Performa dapat ditingkatkan lebih lanjut melalui penanganan ketidakseimbangan kelas (class imbalance), feature engineering, encoding yang lebih baik, atau mencoba algoritma lain seperti Random Forest, Gradient Boosted Trees, atau XGBoost.



Gambar 7. Hasil Decision Tree

Hasil visualisasi struktur pohon keputusan (Decision Tree) pada dataset karyawan Pohon keputusan menunjukkan dengan sangat jelas bagaimana model memprediksi apakah karyawan akan resign (LeaveOrNot = 1) atau tetap bertahan (LeaveOrNot = 0). Berikut urutan atribut paling penting yang dipilih secara bertingkat:

Root node (paling atas & paling kuat): PaymentTier

→ Karyawan dengan gaji tertinggi (Tier 1) hampir semuanya bertahan.

Cabang utama berikutnya:

- Age
- City
- Gender
- Education
- JoiningYear
- EverBenched
- ExperienceInCurrentDomain

Aturan-aturan utama yang terbaca dari pohon ini (dari kiri ke kanan):

- Karyawan dengan PaymentTier tinggi (Tier 1 & sebagian Tier 2) → sangat kecil kemungkinan resign.
- Jika Age ≤ 25 tahun → risiko resign lebih tinggi.
- Jika bekerja di Pune → risiko resign jauh lebih besar dibandingkan Bangalore atau New Delhi.
- Wanita (Female) secara konsisten berada di cabang dengan probabilitas resign lebih tinggi untuk resign.
- Karyawan bergelar Masters atau PHD cenderung resign lebih sering dibandingkan Bachelors.
- Karyawan yang bergabung pada tahun 2017–2018 memiliki kecenderungan resign lebih tinggi.

- Karyawan yang pernah EverBenched = Yes juga lebih sering keluar.
- Pengalaman di domain saat ini (ExperienceInCurrentDomain) menjadi penentu tambahan pada beberapa cabang.

KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi menggunakan algoritma Decision Tree (C4.5) untuk memprediksi kecenderungan karyawan meninggalkan perusahaan (employee turnover) berdasarkan dataset karyawan sebanyak 4.652 records. Setelah melalui tahapan preprocessing yang menunjukkan data sudah bersih tanpa missing value maupun duplikasi, dataset dibagi dengan rasio 90% training dan 10% testing. Model yang dihasilkan mencapai akurasi 75,86% dengan classification error 24,14%.

Visualisasi pohon keputusan mengungkap bahwa faktor paling dominan yang memengaruhi keputusan karyawan untuk resign secara berurutan adalah PaymentTier (tingkat gaji), diikuti Age, City, Gender, Education, JoiningYear, EverBenched, dan ExperienceInCurrentDomain. Karyawan dengan gaji terendah (Tier 3), berjenis kelamin perempuan, bekerja di Pune, bergelar Masters atau PHD, serta usia muda memiliki risiko resign paling tinggi, sedangkan karyawan dengan gaji tertinggi (Tier 1) hampir selalu bertahan.

Meskipun akurasi yang diperoleh tergolong cukup baik sebagai model baseline dan aturan yang dihasilkan sangat mudah diinterpretasikan oleh tim HR, performa terhadap kelas minoritas (karyawan yang resign) masih perlu ditingkatkan karena recall kelas 1 hanya sekitar 85,4%. Hal ini menunjukkan bahwa masih terdapat karyawan berisiko tinggi yang tidak terdeteksi. Oleh karena itu, model Decision Tree ini sudah dapat digunakan sebagai alat bantu pengambilan keputusan awal dalam manajemen retensi karyawan, namun disarankan untuk dilanjutkan dengan teknik penanganan ketidakseimbangan kelas dan percobaan algoritma ensemble guna meningkatkan sensitivitas prediksi terhadap karyawan yang benar-benar akan meninggalkan perusahaan

DAFTAR PUSTAKA

- Aliyah, S., Desi, E., Nasution, F. P., & Tahel, F. (2025). *Analisis Algoritma C4 . 5 dalam Mengukur Tingkat Kepuasan Staf terhadap Kinerja Teknisi Komputer Abstrak*. 6(1), 334–345.
- Azuaje, F. (2006). *Review of "Data Mining : Practical Machine Learning Tools and Techniques " by Witten and Frank*. 2, 1–2. <https://doi.org/10.1186/1475-925X-5-51>
- C, M. A. (2025). *Classification of New Employee Selection Using the C4 . 5 Algorithm Klasifikasi Seleksi Penerimaan Karyawan Baru*. 5(January), 26–34.
- Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hera, A., Rian, A. Al, Faruque, O., Mohtasam, M., Sizan, H., Khan, N. A., Rahaman, A., & Ali, M. J. (2024). *Leveraging Information Systems for Strategic Management : Enhancing Decision-Making and Organizational Performance*. 1045–1061. <https://doi.org/10.4236/ajibm.2024.148054>
- Informasi, S., Bina, U., & Informatika, S. (2019). *Prediksi Promosi Jabatan Karyawan Dengan Algoritma C4 . 5 (Studi Kasus : Apartemen Senayan Jakarta)*. 18(4), 288–298.
- Kasus, S., Mercu, U., & Yogyakarta, B. (2021). *Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4 . 5*. 5, 117–127.
- Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics (Wiley Series on Methods*

- and Applications in Data Mining): 9781118116197: Computer Science Books @ Amazon.com. Wiley Series, 794. <https://doc.lagout.org/Others/Data Mining/Data Mining and Predictive Analytics %5BLarose %26 Larose 2015-03-16%5D.pdf>
- Outsourcing, P., Sahabat, P. T., Pemuda, D. U. A., Palmerah, J., No, B., Rw, R. T., Utara, G., Lama, K. K., Selatan, J., Khusus, D., & Jakarta, I. (2024). *Klasifikasi Proses Menentukan Kelayakan Karyawan Baru dengan Metode*. 4.
- Ross, J., Morgan, Q., & Publishers, K. (1994). *Book Review : C4 . 5 : Programs for Machine Learning*. 240, 235–240.
- Siburian, R. S., Rozalia, O., Alpianita, P., & Dermawan, A. A. (2024). *Klasifikasi Ketidakhadiran Karyawan Menggunakan Metode Algoritma Decision Tree C4 . 5*. 9(01), 61–73.
- Sivi, N. A., Hartono, R., & Hanafi, P. (2023). *Penerapan Algoritma C4 . 5 untuk Prediksi Kelulusan Mahasiswa berdasarkan Data Akademik*. 1(5).
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.