

Penerapan Algoritma *Decision Tree* Untuk Klasifikasi Departemen Karyawan Berdasarkan Data *HR Analytics*

Implementation of the Decision Tree Algorithm for Employee Department Classification Using HR Analytics Data

Muhammad Raka Muhyidin¹⁾, Rofid Safwanda¹⁾, Mohammad Difa Ibnu Rizky¹⁾, Rafli Gilang Ramadhan¹⁾

¹⁾Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Cikarang

Diajukan 12 Januari 2026 / Disetujui 30 Maret 2026

Abstrak

Perkembangan teknologi informasi mendorong organisasi untuk memanfaatkan data sebagai dasar pengambilan keputusan, termasuk dalam pengelolaan sumber daya manusia. Salah satu permasalahan yang sering dihadapi adalah pengelompokan atau klasifikasi departemen karyawan secara akurat berdasarkan karakteristik data yang dimiliki. Penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* dalam melakukan klasifikasi departemen karyawan menggunakan data *HR Analytics* serta mengevaluasi kinerja model yang dihasilkan. Data yang digunakan terdiri dari beberapa atribut, antara lain usia, tingkat pendidikan, jarak tempat tinggal ke kantor, kepuasan kerja, pendapatan bulanan, keterlibatan kerja, dan pengalaman kerja. Proses pengolahan data dilakukan menggunakan perangkat lunak RapidMiner dengan metode validasi *Split Validation*, yaitu 70% data sebagai data latih dan 30% sebagai data uji. Hasil pengujian menunjukkan bahwa algoritma *Decision Tree* mampu menghasilkan tingkat akurasi sebesar 97,05%. Meskipun demikian, analisis lebih lanjut menggunakan *Confusion Matrix* menunjukkan adanya perbedaan performa antar kelas, di mana kelas *Human Resources* memiliki nilai *recall* yang lebih rendah dibandingkan kelas *Research & Development* dan *Sales* akibat ketidakseimbangan data. Oleh karena itu, penelitian ini menegaskan bahwa evaluasi model klasifikasi tidak cukup hanya berdasarkan nilai akurasi, tetapi juga perlu mempertimbangkan metrik *recall* per kelas agar sistem klasifikasi yang dihasilkan lebih adil dan andal.

Kata Kunci: *Data Mining*, *Decision Tree*, Klasifikasi, *HR Analytics*, RapidMiner

Abstract

The development of information technology has encouraged organizations to utilize data as a basis for decision making, including in human resource management. One of the common problems faced is the accurate classification of employee departments based on available data characteristics. This study aims to implement the *Decision Tree* algorithm to classify employee departments using *HR Analytics* data and to evaluate the performance of the resulting model. The dataset consists of several attributes such as age, education level, distance from home to office, job satisfaction, monthly income, job involvement, and work experience. Data processing was conducted using RapidMiner software with the *Split Validation* method, where 70% of the data was used as training data and 30% as testing data. The experimental results show that the *Decision Tree* algorithm achieved an accuracy of 97.05%. However, further analysis using a *Confusion Matrix* indicates performance differences among classes, where the *Human Resources* class has a lower recall value compared to the *Research & Development* and *Sales* classes due to data imbalance. Therefore, this study emphasizes that classification model evaluation should not rely solely on accuracy but also consider class-based recall to ensure a fair and reliable classification system.

Keywords: *Data Mining*, *Decision Tree*, Classification, *HR Analytics*, RapidMiner

*Korespondensi Penulis:
E-mail: dipaibnu25@gmail.com

Pendahuluan

Dalam era transformasi digital dan Revolusi Industri 4.0, data menjadi aset strategis bagi organisasi dalam mendukung pengambilan keputusan. Salah satu bidang yang menghasilkan data dalam jumlah besar adalah manajemen sumber daya manusia (Rizqiyah et al., 2021). Data *HR Analytics* mencakup berbagai informasi mengenai karakteristik dan kinerja karyawan yang dapat dimanfaatkan untuk meningkatkan efektivitas pengelolaan organisasi. Namun, tanpa teknik pengolahan data yang tepat, data tersebut tidak dapat memberikan informasi yang bernilai (Priliasari, 2023; Sayaputra, 2025).

Data mining merupakan salah satu pendekatan yang dapat digunakan untuk mengekstraksi pola dan pengetahuan dari kumpulan data dalam jumlah besar (Andry et al., 2023). Salah satu fungsi utama dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas-kelas tertentu berdasarkan karakteristik yang dimiliki (Ahmad et al., 2021; Munoz-Ausecha et al., 2023). Dalam konteks organisasi, klasifikasi departemen karyawan seperti *Research & Development*, *Sales*, dan *Human Resources* sangat penting untuk mendukung proses administrasi, analisis kinerja, serta pengambilan keputusan manajerial (Henglie et al., 2024).

Algoritma *Decision Tree* merupakan salah satu metode klasifikasi yang banyak digunakan karena mudah dipahami, bersifat interpretatif, dan mampu menangani data kategorikal maupun numerik (Xie, 2024). Selain itu, *Decision Tree* dapat menghasilkan aturan keputusan yang jelas sehingga memudahkan pihak manajemen dalam memahami hasil klasifikasi (Kamila et al., 2025). Meskipun demikian, model klasifikasi sering kali hanya dievaluasi berdasarkan nilai akurasi secara keseluruhan,

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan algoritma *Decision Tree* untuk klasifikasi departemen karyawan menggunakan data *HR Analytics* serta melakukan evaluasi kinerja model secara lebih mendalam menggunakan *Confusion Matrix*. Evaluasi ini bertujuan untuk mengetahui performa model pada setiap kelas dan mengidentifikasi potensi kelemahan, khususnya pada kelas dengan jumlah data yang lebih sedikit.

Metode Penelitian

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian terapan (*applied research*). Pendekatan ini dipilih karena penelitian berfokus pada penerapan metode data mining untuk menyelesaikan permasalahan klasifikasi departemen karyawan secara nyata (Irwanto et al., 2024).

Sumber dan Jenis Data

Data yang digunakan merupakan data *HR Analytics* yang berisi atribut-atribut karyawan, antara lain usia, tingkat pendidikan, jarak tempat tinggal ke kantor, kepuasan kerja, pendapatan bulanan, keterlibatan kerja, serta pengalaman kerja (Irwanto et al., 2024). Data tersebut memiliki tiga kelas target, yaitu *Research & Development*, *Sales*, dan *Human Resources*.

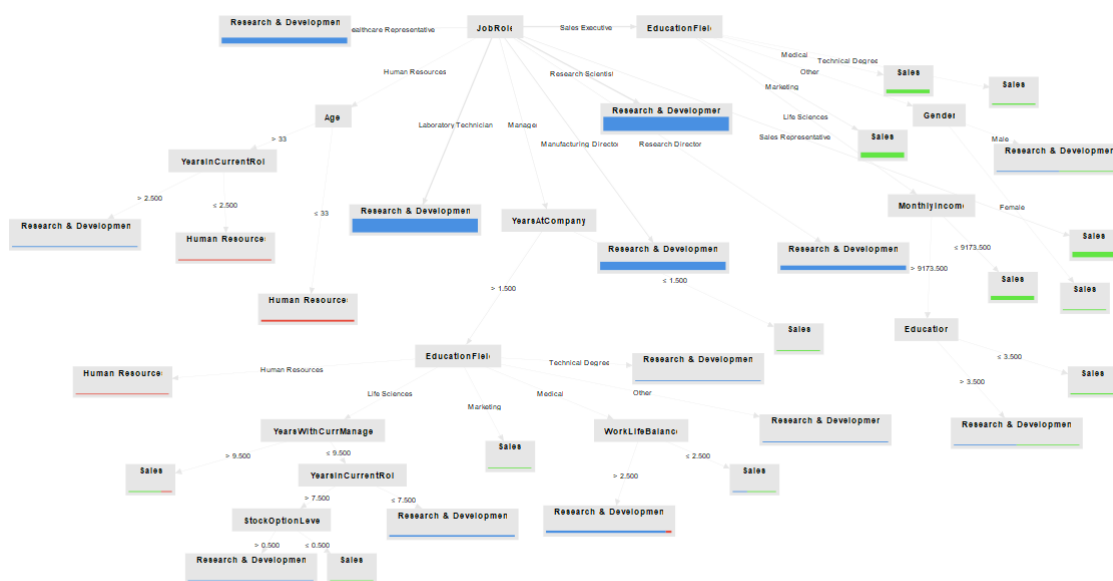
Hasil Dan Pembahasan

Hasil Pengujian Model Klasifikasi Menggunakan RapidMiner

Pengujian model klasifikasi departemen karyawan dilakukan menggunakan perangkat lunak RapidMiner dengan algoritma *Decision Tree* sebagai model utama. Proses pengujian diawali dengan

pembagian dataset menggunakan metode *Split Validation*, di mana 70% data digunakan sebagai data latih (*training set*) dan 30% sebagai data uji (*testing set*). Pendekatan ini bertujuan untuk memastikan bahwa model yang dibangun mampu melakukan generalisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya (Putri & Situmeang, 2024).

Model *Decision Tree* yang dihasilkan membentuk struktur pohon keputusan yang terdiri dari node akar, node cabang, dan node daun. Setiap node merepresentasikan atribut yang paling berpengaruh dalam menentukan kelas departemen karyawan, sedangkan node daun menunjukkan hasil klasifikasi akhir berupa departemen *Research & Development*, *Sales*, atau *Human Resources*. Pemilihan atribut pada setiap node dilakukan berdasarkan nilai *information gain* tertinggi, sehingga atribut yang berada pada tingkat atas pohon memiliki kontribusi paling besar dalam proses klasifikasi.



Gambar 1. Hasil *Decision Tree*

Hasil pengujian model menunjukkan bahwa *Decision Tree* mampu mempelajari pola hubungan antara atribut karyawan dengan departemen tempat mereka bekerja ditampilkan pada gambar 1. Hal ini mengindikasikan bahwa data HR Analytics yang digunakan memiliki keterkaitan yang cukup kuat dengan kelas target, sehingga layak digunakan sebagai dasar pengambilan keputusan berbasis data.

Evaluasi kinerja model secara keseluruhan

Evaluasi kinerja model dilakukan menggunakan metrik evaluasi yang dihasilkan oleh operator *Performance (Classification)* pada RapidMiner. Salah satu metrik utama yang dihasilkan adalah nilai akurasi keseluruhan (*overall accuracy*). Berdasarkan hasil pengujian, model *Decision Tree* memperoleh tingkat akurasi sebesar **97,05%**, yang berarti sebagian besar data uji berhasil diklasifikasikan dengan benar ke dalam departemen yang sesuai.

Nilai akurasi yang tinggi ini secara umum menunjukkan bahwa model memiliki performa yang sangat baik. Dalam konteks penelitian data mining, akurasi di atas 90% sering dianggap sebagai indikator keberhasilan model klasifikasi. Namun demikian, akurasi yang tinggi tidak selalu

mencerminkan performa model yang adil dan merata, terutama pada kasus klasifikasi multikelas dengan distribusi data yang tidak seimbang.

Oleh karena itu, evaluasi tidak hanya berhenti pada nilai akurasi, tetapi dilanjutkan dengan analisis *Confusion Matrix* untuk memperoleh gambaran performa model secara lebih rinci pada setiap kelas departemen.

Analisis *Confusion Matrix*

Confusion Matrix digunakan untuk membandingkan hasil prediksi model dengan data aktual pada data uji. Matriks ini memberikan informasi mengenai jumlah prediksi yang benar (*True Positive*) dan prediksi yang salah (*False Positive* dan *False Negative*) untuk setiap kelas.

Berdasarkan *Confusion Matrix* yang dihasilkan, diketahui bahwa jumlah data uji terdiri dari tiga kelas, yaitu *Research & Development*, *Sales*, dan *Human Resources*, dengan distribusi data yang tidak seimbang. Kelas *Research & Development* memiliki jumlah data paling besar, diikuti oleh *Sales*, sedangkan *Human Resources* memiliki jumlah data paling sedikit.

accuracy: 97.05% +/- 2.47% (micro average: 97.04%)

	true Research & Develo...	true Sales	true Human Resources
pred. Research & Devel...	623	10	5
pred. Sales	5	243	1
pred. Human Resources	5	1	20
class recall	98.42%	95.67%	76.92%

Gambar 2. Hasil *Confusion Matrix*

Hasil *Confusion Matrix* menunjukkan bahwa sebagian besar data *Research & Development* dan *Sales* berhasil diklasifikasikan dengan benar oleh model ditampilkan pada tabel 2. Sebaliknya, pada kelas *Human Resources* masih terdapat sejumlah data yang salah diklasifikasikan ke dalam kelas lain, terutama ke kelas *Research & Development* dan *Sales*. Temuan ini menunjukkan adanya kecenderungan model untuk lebih dominan memprediksi kelas mayoritas.

Simpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma *Decision Tree* mampu mengklasifikasikan departemen karyawan menggunakan data *HR Analytics* dengan tingkat akurasi yang tinggi. Meskipun demikian, evaluasi menggunakan *Confusion Matrix* menunjukkan bahwa performa model belum merata pada semua kelas, khususnya pada kelas *Human Resources* yang memiliki jumlah data lebih sedikit. Oleh karena itu, evaluasi kinerja model klasifikasi tidak cukup hanya berdasarkan akurasi, tetapi juga perlu mempertimbangkan metrik recall per kelas agar sistem klasifikasi yang dihasilkan lebih adil dan efektif.

Daftar Pustaka

Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1), 1–29.

<https://doi.org/10.1002/ett.4150>

- Andry, J. F., Hartono, H., Honni, Deny, & Jo, J. (2023). Analysis and Prediction Supermarket Sales with Data Mining Using Rapidminer. *AIP Conference Proceedings*, 2693(1). <https://doi.org/10.1063/5.0118725>
- Henglie, R., Purnomo, Y., & Ginting, J. A. (2024). Memprediksi Peningkatan H-Indeks untuk Jurnal Penelitian dengan Menggunakan Algoritma Cost-Sensitive Selective Naive Bayes Classifiers. *Jurnal Algoritma, Logika Dan Komputasi*, 7(1), 643–651. <https://doi.org/10.30813/j-alu.v7i1.6028>
- Irwanto, Saputra, R. W., Ayuningtyas, & Priono, S. (2024). Pengaruh Disiplin Pegawai dan Sarana Prasarana terhadap Kinerja Pegawai Melalui Penerapan Absensi Online di Kantor Kecamatan Banjarbaru Selatan. *Jurnal Riset Inspirasi Manajemen Dan Kewirausahaan*, 8(1), 12–22. <https://doi.org/10.35130/jrimk.v8i1.509>
- Kamila, A. R., Lee, F. S., & Andry, J. F. (2025). Peningkatan Performa Model Machine Learning Untuk Deteksi Dini Polycystic Ovary Syndrome Melalui Kombinasi Metode Preprocessing. *Infotech: Jurnal of Techology Information*, 11(2), 267–274. <https://doi.org/10.37365/jti.v11i2.448>
- Munoz-Ausecha, C., Gómez, J. E. G., Ruiz-Rosero, J., & Ramirez-Gonzalez, G. (2023). Asset Ownership Transfer and Inventory Using RFID UHF TAGS and Ethereum Blockchain NFTs. *Electronics (Switzerland)*, 12(6), 1–20. <https://doi.org/10.3390/electronics12061497>
- Prihasari, E. (2023). Perlindungan Data Pribadi Konsumen dalam Transaksi E-Commerce. *Jurnal Rechts Vinding: Media Pembinaan Hukum Nasional*, 12(2). <https://doi.org/10.33331/rechtsvinding.v12i2.1285>
- Putri, N., & Situmeang, S. E. (2024). Analisis Perbandingan Metode Decision Tree, Random Forest, dan Support Vector Machine (SVM) dalam Memprediksi Kesehatan Janin. *Departemen Statistika Institut Teknologi Sepuluh Nopember*, 1, 1–8. <https://doi.org/10.12962/J27213862.VXIX.XXXX>
- Rizqiyah, R. N., Kriyantono, R., & Sujoko, A. (2021). Manajemen Media Relations Dalam Meningkatkan Citra Positif Dan Reputasi Perusahaan Di Era Artificial Intelligence. *Metacommunication: Journal of Communication Studies*, 6(1), 53–66. <https://doi.org/10.20527/mc.v6i1.10135>
- Sayaputra, A. (2025). Implementation of Machine Learning Algorithms for Predicting Student Final GPA Using Multiclass Classification Models. *Journal of Artificial Intelligence and Software Engineering*, 5(2), 660–675. <https://doi.org/10.30811/jaise.v5i2.7015>
- Xie, Y. (2024). Efficiency of Hybrid Decision Tree Algorithms in Evaluating the Academic Performance of Students. *The Science and Information Organization*, 15(2), 480–491. <https://thesai.org/Publications/ViewPaper?Volume=15&Issue=2&Code=IJACSA&SerialNo=51>