https://jurnal.pustakagalerimandiri.co.id/index.php/pustakaai
DOI : https://doi.org/10.55382/jurnalpustakaai.v5i2.115

JURNAL PUSTAKA AI
JURNAL PUSAT AKSES KAJIAN TEKNOLOGI ARTIFICIAL INTELLIGENCE

Vol. 5 No. 2 (2025) 177 – 187 E ISSN: 2809-4069

Analisis Faktor yang Mempengaruhi Promosi Karyawan Menggunakan *Random Forest* pada *Dataset Employee Promotion*

Sapto Kurniawan[1], Agung Nugroho[2], Suherman[3]
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Jl. Inspeksi Kalimalang Tegal Danas Arah Deltamas, Cibatu, Cikarang, Kabupaten Bekasi
[1]sapto.kurniawan2000.sk@gmail.com [2]agung@pelitabangsa.ac.id [3]suherman@pelitabangsa.ac.id

Abstract

Subjectivity and lack of transparency in employee promotion processes remain significant challenges in human resource management. This study aims to develop a data-driven classification model for employee promotion using the Random Forest algorithm and the CRISP-DM methodology. The dataset employed is the publicly available Employee Promotion Dataset from Kaggle, which exhibits an imbalanced class distribution. To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied during data preprocessing. The model was evaluated by using classification metrics such as accuracy, precision, recall, and ROC AUC. The results show that the model achieved 93% accuracy, 85.13% precision, and 26.55% recall. Feature importance analysis revealed that training score, previous year's performance rating, length of service, KPI achievements, and awards are the dominant factors influencing promotion decisions. These findings are expected to support the development of a more objective and data-driven decision-making system for employee promotions.

Keywords: employee promotion, random forest, SMOTE, classification, HR analytics

Abstrak

Permasalahan subjektivitas dan kurangnya transparansi dalam proses promosi karyawan menjadi tantangan dalam manajemen sumber daya manusia. Penelitian ini bertujuan untuk membangun model klasifikasi promosi karyawan berbasis data menggunakan algoritma *Random Forest* dengan pendekatan CRISP-DM. Dataset yang digunakan merupakan *Employee Promotion Dataset* dari *Kaggle* yang memiliki distribusi kelas tidak seimbang. Untuk menangani hal tersebut, diterapkan teknik SMOTE *(Synthetic Minority Oversampling Technique)* pada data latih. Model dievaluasi menggunakan metrik klasifikasi seperti akurasi, precision, recall, dan ROC AUC. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi sebesar 93%, precision 85,13%, dan recall 26,55%. Analisis feature importance mengidentifikasi skor pelatihan, penilaian kinerja sebelumnya, lama bekerja, capaian KPI, dan penghargaan sebagai faktor dominan dalam keputusan promosi. Temuan ini diharapkan dapat mendukung sistem pengambilan keputusan promosi yang lebih objektif dan berbasis data.

Kata kunci: promosi karyawan, *random forest*, SMOTE, klasifikasi, *HR analytics*

1. Pendahuluan

Promosi karyawan merupakan salah satu proses strategis dalam manajemen sumber daya manusia (SDM) yang dapat berdampak langsung pada motivasi kerja dan keberlangsungan organisasi. Keputusan promosi yang adil dan tepat dapat meningkatkan loyalitas, sedangkan promosi yang subjektif cenderung memicu ketidakpuasan dan

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025

turnover. Dalam praktiknya, banyak perusahaan masih mengandalkan penilaian manual yang tidak sepenuhnya berbasis data, sehingga muncul kebutuhan untuk menerapkan pendekatan prediktif yang lebih objektif.

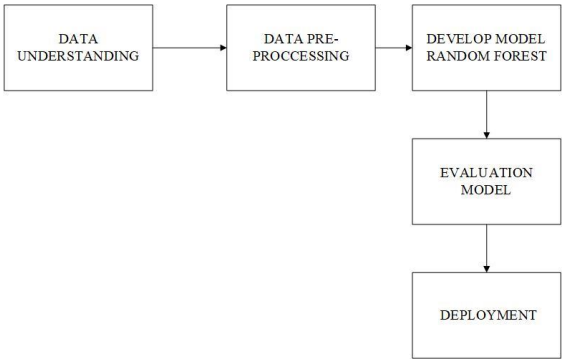
Seiring berkembangnya teknologi kecerdasan buatan, algoritma machine learning telah dimanfaatkan dalam pengambilan keputusan SDM, termasuk dalam klasifikasi promosi. Salah satu algoritma yang banyak digunakan adalah Random Forest, karena kemampuannya dalam menangani data kategorikal dan numerik secara bersamaan serta menghasilkan interpretasi melalui analisis feature importance. Studi oleh Sotarjua dan Santoso [1] menunjukkan bahwa Random Forest unggul dalam mengklasifikasi promosi pada data tidak seimbang, mengalahkan KNN dan Decision Tree. Penelitian lain juga membuktikan bahwa algoritma klasifikasi seperti C4.5 dapat diaplikasikan untuk prediksi promosi, namun memiliki keterbatasan dalam akurasi dan interpretabilitas [2]. Sebaliknya, penelitian oleh Miftahuljannah dkk. (2023) yang menggunakan algoritma regresi linear dalam memprediksi penjualan produk justru menunjukkan bahwa pemilihan atribut yang kurang relevan berdampak signifikan terhadap tingginya nilai kesalahan prediksi. Hal ini menegaskan bahwa pemilihan metode dan fitur yang tepat sangat penting untuk membangun model prediksi yang andal dalam pengambilan keputusan berbasis data [3].

Namun, sebagian besar studi terdahulu masih menghadapi tantangan pada ketidakseimbangan data (imbalanced dataset), di mana jumlah karyawan yang dipromosikan jauh lebih kecil dibandingkan yang tidak. Hal ini menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas. Beberapa penelitian belum secara eksplisit mengombinasikan Random Forest dengan teknik penyeimbangan data seperti SMOTE (Synthetic Minority Oversampling Technique) dalam konteks prediksi promosi [4]. Sementara itu, penelitian terbaru juga menunjukkan bahwa integrasi algoritma Random Forest dan teknik penyeimbangan seperti SMOTE mampu meningkatkan kinerja model klasifikasi secara signifikan dalam domain prediksi, termasuk untuk pengambilan keputusan bisnis berbasis data historis (Sangaji & Sutabri, 2025) [5]. Selain itu, algoritma K-Means Clustering telah berhasil digunakan untuk segmentasi pelanggan dan optimalisasi strategi penjualan (Sari dkk., 2025; Pangarso dkk., 2025) [6][7], yang menunjukkan tren pemanfaatan data-driven decision dalam berbagai konteks bisnis.Oleh karena itu, dibutuhkan studi lanjutan yang tidak hanya fokus pada akurasi model, tetapi juga pada identifikasi faktor-faktor penting yang memengaruhi promosi, menggunakan pendekatan yang lebih robust.

Penelitian ini bertujuan untuk membangun model klasifikasi promosi karyawan menggunakan algoritma Random Forest pada Employee Promotion Dataset. Model ini dievaluasi secara menyeluruh menggunakan metrik klasifikasi seperti akurasi, precision, recall, dan ROC AUC [5]. Selain itu, dilakukan analisis terhadap faktor-faktor utama yang memengaruhi promosi berdasarkan perhitungan feature importance. Hasil penelitian diharapkan dapat memberikan wawasan praktis kepada pihak manajemen SDM dalam menyusun kebijakan promosi yang lebih adil, transparan, dan berbasis data.

2. Metode Penelitian

Penelitian ini dilakukan berdasarkan pendekatan data mining dengan standar metodologi CRISP-DM (Cross Industry Standard Process for Data Mining), yang terdiri dari lima tahap utama: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment[6].

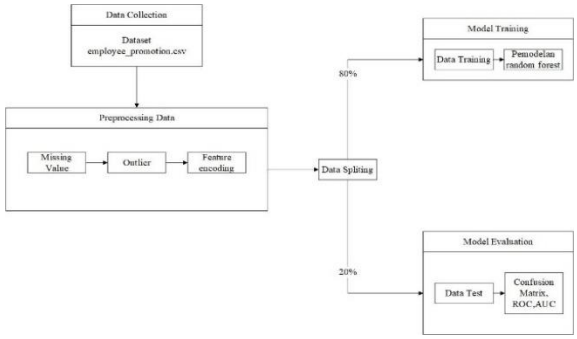


Gambar 1. Tahap Metodologi CRISP-DM

Secara umum, proses penelitian dimulai dari pemahaman tujuan bisnis yaitu bagaimana membangun model prediktif untuk promosi karyawan secara adil dan berbasis data. Setelah itu, dilakukan pemahaman dan eksplorasi terhadap dataset publik yang diperoleh dari Kaggle, yang mencakup data karyawan seperti usia, departemen, tingkat pendidikan, skor pelatihan, lama kerja, serta status promosi. Dataset ini bersifat tidak seimbang (imbalanced), di mana hanya sebagian kecil karyawan yang dipromosikan.

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025

Selanjutnya, dilakukan proses data preparation seperti pembersihan data, penanganan missing value, dan encoding fitur kategorikal. Untuk mengatasi ketidakseimbangan data, digunakan teknik SMOTE (Synthetic Minority Oversampling Technique) agar model dapat belajar lebih baik dalam mengenali kelas minoritas[8]. Tahapan ini dilanjutkan dengan pembangunan model menggunakan algoritma Random Forest yang dilatih pada data latih dan diuji pada data uji untuk melihat kinerjanya.
Evaluasi dilakukan menggunakan confusion matrix dan metrik seperti accuracy, precision, recall, dan F1-score, serta dilengkapi dengan analisis feature importance untuk mengidentifikasi variabel paling berpengaruh dalam proses promosi[9]. Hasil penelitian disusun dalam bentuk visualisasi dan kesimpulan yang dapat memberikan rekomendasi strategis kepada perusahaan dalam mengambil keputusan promosi karyawan secara lebih objektif dan data-driven.Lebih detailnya dapat dilihat pada gambar 2 Diagram Alir Penelitian dibawah ini.



Gambar 2. Diagram Alir Penelitian

Dataset yang digunakan adalah Employee Promotion Dataset yang tersedia secara publik melalui platform Kaggle. Dataset ini berisi 54.808 data karyawan dengan berbagai atribut seperti umur, jenis kelamin, departemen, pendidikan, masa kerja, hasil pelatihan, penilaian kinerja, penghargaan, dan status promosi. Atribut target adalah is_promoted, dengan distribusi kelas tidak seimbang (hanya ±8,5% karyawan yang dipromosikan).
Adapun deskripsi dari masing-masing atribut dalam dataset ini dapat dilihat pada tabel berikut:

Tabel 1. Deskripsi Atribut pada Employee Promotion Dataset

No.	Nama Atribut	Tipe Data	Deskripsi
1	employee_id	Integer	ID unik setiap karyawan
2	department	Kategorikal	Departemen tempat karyawan bekerja (contoh: Sales & Marketing, Finance)
3	region	Kategorikal	Wilayah kerja karyawan (34 kategori, contoh: region_1, region_2, dst.)
4	education	Kategorikal	Tingkat pendidikan terakhir (Bachelor's, Master's, Below Secondary)
5	gender	Kategorikal	Jenis kelamin (m = laki-laki, f = perempuan)
6	recruitment_channel	Kategorikal	Jalur rekrutmen (other, sourcing, referred)
7	no_of_trainings	Numerik	Jumlah pelatihan yang diikuti (rata-rata: 1, maksimum: 10)
8	age	Numerik	Usia karyawan (antara 20–60 tahun, rata-rata 34,8 tahun)
9	previous_year_rating	Numerik	Nilai kinerja tahun sebelumnya (1–5, sebagian ada yang kosong)
10	length_of_service	Numerik	Lama kerja di perusahaan (1–37 tahun, rata-rata 5,8 tahun)
11	awards_won	Boolean	Penghargaan: 1 = ya, 0 = tidak (hanya 2,3% karyawan yang mendapat)
12	avg_training_score	Numerik	Skor pelatihan rata-rata (dari 39–99, rata-rata 63,7)
13	is_promoted	Label	Target prediksi: 1 = dipromosikan, 0 = tidak (hanya sekitar 8,5% yang ya)

2.1. Alur Penelitian Berdasarkan CRISP-DM
Tahapan penelitian dilakukan secara berurutan sebagai berikut:
a. Business Understanding
Tahap ini bertujuan memahami permasalahan bisnis, yaitu bagaimana proses promosi karyawan yang sering kali tidak objektif dapat dibantu melalui analisis data. Tujuan utama adalah membangun model prediksi promosi dan memahami fitur-fitur yang paling berpengaruh dalam keputusan tersebut[10].
b. Data Understanding
Peneliti melakukan eksplorasi awal terhadap struktur dataset. Ditemukan adanya data kosong pada kolom education dan previous_year_rating, serta ketidakseimbangan distribusi kelas target. Analisis statistik dan visualisasi dilakukan untuk memahami karakteristik data[11].
c. Data Preparation
Langkah-langkah prapemrosesan data meliputi:
Imputasi nilai kosong dengan modus (untuk kategori) dan median (untuk numerik).

One-hot encoding untuk fitur kategorikal seperti department, region, dan education.
Normalisasi tidak dilakukan karena Random Forest tidak sensitif terhadap skala data[12].
Penanganan data tidak seimbang menggunakan teknik SMOTE (Synthetic Minority Oversampling Technique) untuk memperbanyak sampel kelas minoritas.
d. Modeling
Model dibangun menggunakan algoritma Random Forest Classifier dari library Scikit-Learn. Dataset dibagi menjadi data latih dan uji dengan rasio 80:20. Parameter awal menggunakan pengaturan default, tanpa hyperparameter tuning[13].
e. Evaluation
Evaluasi model dilakukan dengan metrik klasifikasi meliputi:

1) Accuracy untuk mengukur keseluruhan prediksi benar,
2) Precision untuk melihat ketepatan dalam memprediksi kelas promosi,
3) Recall untuk mengukur sensitivitas terhadap kelas minoritas,
4) F1-score sebagai harmonisasi antara precision dan recall,
5) ROC AUC untuk mengukur kualitas pemisahan antara dua kelas.
6) Selain itu, digunakan confusion matrix dan feature importance untuk interpretasi performa dan kontribusi masing-masing fitur[14].

f. Deployment
Tahap ini mencakup penyusunan hasil dalam bentuk visualisasi, analisis fitur, dan identifikasi calon karyawan potensial. Model tidak diimplementasikan dalam bentuk aplikasi, tetapi hasil penelitian ini dapat menjadi dasar untuk sistem pengambilan keputusan berbasis data di bidang HR.

2.2. Alat dan Bahasa Pemrograman
Penelitian ini dilakukan sepenuhnya menggunakan platform Google Colaboratory (Colab) sebagai lingkungan eksperimen berbasis cloud yang mendukung kolaborasi dan skalabilitas[15]. Adapun alat dan pustaka utama yang digunakan mencakup:

1) Python 3.9 sebagai bahasa pemrograman utama,
2) Scikit-learn untuk proses modeling dan evaluasi,
3) Pandas & NumPy untuk manipulasi dan analisis data,
4) Matplotlib & Seaborn untuk visualisasi eksploratif dan evaluatif,
5) Imbalanced-learn (imblearn) untuk penerapan teknik SMOTE[16].

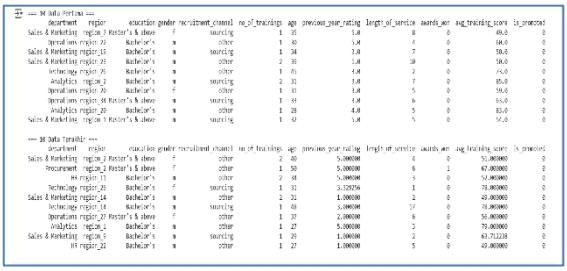
Penggunaan ekosistem teknologi tersebut mendukung replikasi penelitian, serta memastikan bahwa pendekatan yang digunakan bersifat modular, transparan, dan sesuai standar penelitian berbasis data terkini. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoritis, tetapi juga memiliki nilai aplikatif yang tinggi untuk mendukung kebijakan berbasis data dalam pengelolaan sumber daya manusia.

3. Hasil dan Pembahasan

Penelitian ini bertujuan untuk membangun model prediktif yang dapat mengidentifikasi karyawan berpotensi promosi berdasarkan data historis. Dalam upaya ini, dilakukan serangkaian proses mulai dari eksplorasi data, pra-pemrosesan, pelatihan model, hingga evaluasi performa menggunakan beberapa metrik.

3.1 Deskripsi Dataset

Dataset yang digunakan berjudul employee_promotion.csv yang diperoleh dari platform Kaggle, terdiri dari 54.808 data karyawan. Dataset mencakup 13 atribut yang menggambarkan informasi personal, performa kerja, dan status promosi. Kolom target dalam dataset adalah is_promoted, yang bernilai 1 untuk karyawan yang dipromosikan dan 0 jika tidak. Distribusi data menunjukkan adanya ketimpangan kelas, dengan hanya sekitar 8,52% karyawan yang mendapatkan promosi. Dari dataset ini dapat ditampilkan 10 data awal dan 10 data akhir sebagai berikut:

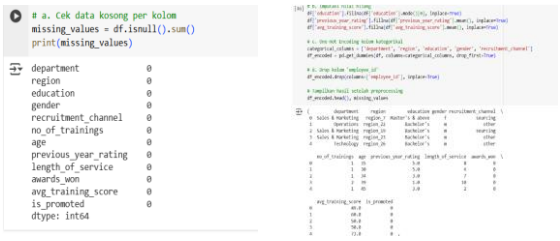


Gambar 3. Tampilan 10 data awal dan 10 data akhir

Oleh karena itu, dibutuhkan teknik penyeimbangan data agar hasil prediksi tidak bias terhadap kelas mayoritas.

3.2 Pra-pemrosesan Data

Tahapan pra-pemrosesan meliputi penanganan missing value menggunakan metode imputasi modus dan median[17]. Transformasi fitur kategorikal dilakukan menggunakan one-hot encoding.



Gambar 4. Proses menghilangkan nilai kosong (missing value)

Teknik SMOTE (Synthetic Minority Over-sampling Technique) diterapkan untuk mengatasi ketidakseimbangan kelas dalam data latih, yang menghasilkan distribusi kelas yang lebih proporsional. Selain itu, analisis terhadap outlier dilakukan, khususnya pada fitur length_of_service, dan diputuskan untuk tetap mempertahankan data tersebut karena masih relevan secara logis.

3.3 Pelatihan dan Evaluasi Model

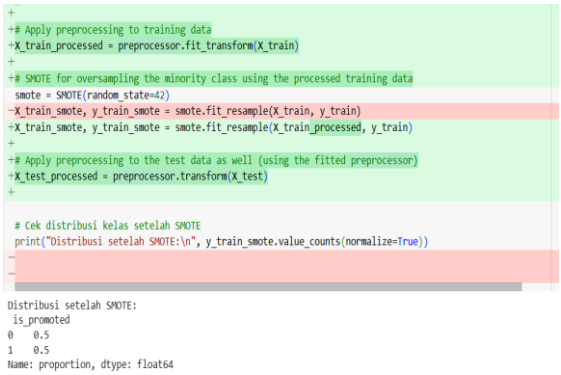
Model dilatih menggunakan algoritma Random Forest Classifier dengan data latih yang telah diolah. Evaluasi dilakukan menggunakan data uji yang tidak dimodifikasi.

Gambar 5. Proses pemisahan fitur is_promoted

Dataset dibagi menjadi dua bagian, yaitu 80% untuk data latih (training data) dan 20% untuk data uji (testing data)**.** Pembagian ini dilakukan secara acak tetapi dengan mempertahankan proporsi distribusi kelas menggunakan teknik *stratified split* agar representasi antara data promosi dan tidak promosi tetap seimbang di kedua subset. Model kemudian dilatih pada data latih dan dilakukan pengujian menggunakan data uji untuk mengukur kinerjanya secara obyektif[18].

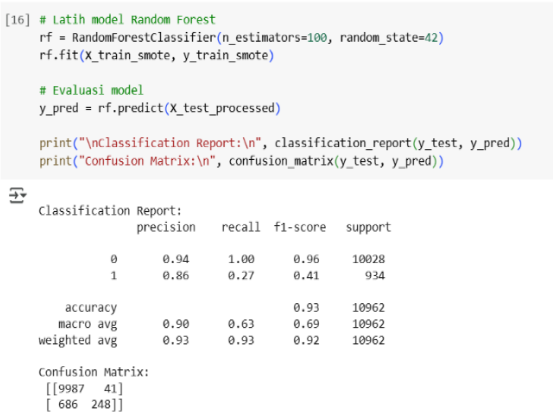
Untuk mengatasi ketidakseimbangan kelas dalam dataset pelatihan, digunakan metode SMOTE (Synthetic Minority Over-sampling Technique). SMOTE bekerja dengan cara menghasilkan data sintetis untuk kelas minoritas, yaitu karyawan yang dipromosikan (is_promoted = 1), sehingga distribusi antara dua kelas menjadi seimbang. Setelah dilakukan SMOTE, distribusi kelas menjadi 50% untuk masing-masing kelas, memungkinkan model belajar secara seimbang dari kedua jenis label.

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025



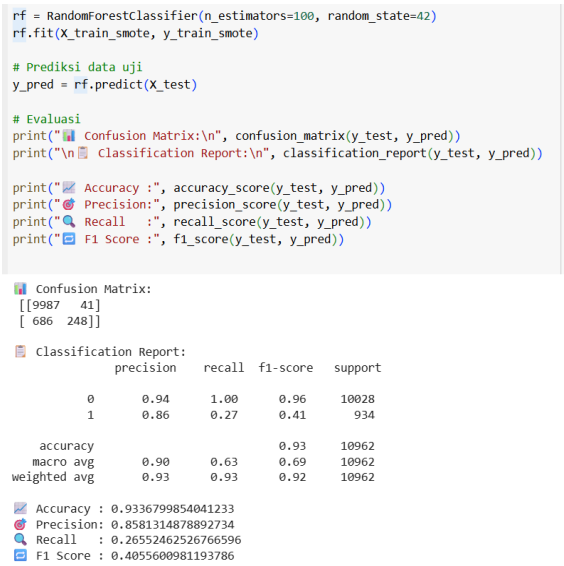
Gambar 6. Proses SMOTE pada training data

Setelah data seimbang, dilakukan pelatihan model menggunakan Random Forest Classifier. Model dilatih menggunakan data hasil SMOTE dan diuji terhadap data uji yang tidak diubah (tetap mencerminkan distribusi asli). Evaluasi performa dilakukan dengan metrik classification_report, confusion_matrix, dan ROC AUC.



Gambar 7. Proses Latih Model Random Forest

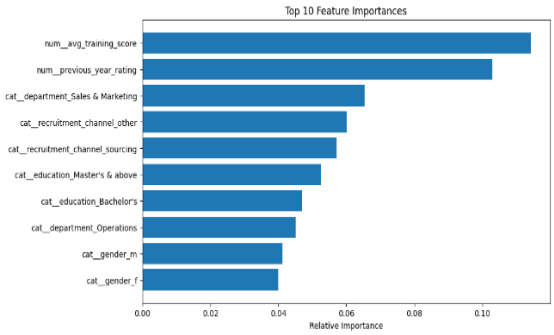
Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu akurasi**,** precision**,** recall**,** F1-score, dan ROC AUC**.** Hasil evaluasi model pada data uji ditampilkan sebagai berikut:



Gambar 8. Proses Evaluasi dataset

Hasilnya menunjukkan akurasi sebesar 93%, precision 85,1%, recall 26,5%, F1-score 40,5%, dan AUC 0,78. Meskipun precision tergolong tinggi, nilai recall masih rendah akibat ketidakseimbangan kelas. Visualisasi feature

importance menunjukkan bahwa fitur previous_year_rating, avg_training_score, dan length_of_service merupakan kontributor utama dalam keputusan promosi.



Gambar 9. Grafik 10 fitur dengan pengaruh terbesar prediksi promosi

Berdasarkan hasil analisis fitur menggunakan algoritma Random Forest, diperoleh 10 fitur dengan pengaruh terbesar dalam proses prediksi promosi karyawan. Fitur avg_training_score dan previous_year_rating menempati posisi teratas, yang menunjukkan bahwa performa dan keberhasilan dalam pelatihan sangat menentukan kemungkinan promosi. Selanjutnya, variabel kategorikal seperti department_Sales & Marketing dan education_Master's & above juga memiliki kontribusi yang signifikan, menunjukkan bahwa departemen tempat bekerja serta jenjang pendidikan berpengaruh terhadap keputusan promosi. Fitur terkait saluran rekrutmen (recruitment_channel) juga turut mempengaruhi, yang dapat mengindikasikan perbedaan performa antar jalur perekrutan. Analisis ini memperkuat pemahaman bahwa promosi karyawan tidak hanya dipengaruhi oleh satu aspek saja, melainkan kombinasi dari performa kerja, kompetensi, pengalaman, dan latar belakang personal serta profesional.

3.4 Identifikasi Karyawan Potensial Dipromosikan

Model memprediksi sebanyak 289 karyawan dalam data uji berpotensi dipromosikan. Dari analisis terhadap karakteristik karyawan tersebut, ditemukan bahwa usia rata-rata mereka adalah 34,37 tahun, dengan masa kerja sekitar 5,72 tahun, skor pelatihan 71,53, dan rating performa tahun sebelumnya mencapai 3,94. Persentase penerima penghargaan dan jumlah pelatihan lebih dari satu juga lebih tinggi di antara karyawan yang dipromosikan. Hal ini mengindikasikan bahwa promosi tidak bersifat acak, melainkan terkait erat dengan performa historis dan keterlibatan karyawan dalam program pengembangan.

Tabel 2 menyajikan ringkasan statistik dari karyawan yang dipromosikan berdasarkan nilai rata-rata dan persentase:

Tabel 2. Ringkasan Statistik Karyawan yang Dipromosikan

Fitur	Nilai Rata-rata / Persentase
Usia Rata-rata	34,37 tahun
Rata-rata Lama Bekerja (tahun)	5,72 tahun
Rata-rata Skor Pelatihan	71,53
Rata-rata Rating Kinerja Tahun Sebelumnya	3,94
Persentase Penerima Penghargaan	11,98%
Persentase dengan >1 Pelatihan	16,24%

Berdasarkan data tersebut, karyawan yang dipromosikan umumnya memiliki usia sekitar 34 tahun, dengan lama masa kerja hampir 6 tahun. Mereka juga cenderung memiliki skor pelatihan yang tinggi, yaitu sekitar 71,53, serta rating kinerja tahun sebelumnya yang mendekati nilai maksimal, yaitu 3,94 dari skala 5. Selain itu, sekitar 12% dari karyawan yang dipromosikan pernah menerima penghargaan, dan 16% mengikuti lebih dari satu pelatihan selama periode yang diamati.

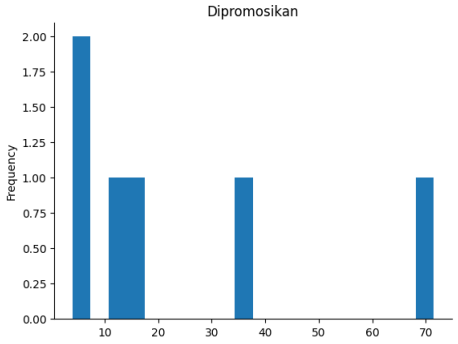
Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025

Selain ringkasan agregat tersebut, di bawah ini ditampilkan 3 contoh karyawan yang berhasil dipromosikan berdasarkan data aktual, untuk memberikan ilustrasi nyata:

Tabel 3. Contoh 3 karyawan yang dipromosikan

wan	Departemen	Pendidikan	Usia	Lama Bekerja (thn)	Rating Tahun Lalu	Skor Pelatihan	Penghargaan	Jumlah Pelati
402	Sales & Marketing	Master's & above	32	5	5	85	Ya	2
776	Operations	Bachelor's	36	7	4	78	Tidak	1
30	Technology	Master's & above	33	6	5	89	Ya	3

Ketiga karyawan di atas memiliki nilai-nilai yang konsisten dengan faktor-faktor penting yang telah diidentifikasi dalam analisis feature importance, seperti skor pelatihan yang tinggi, rating kinerja tahun sebelumnya di atas rata-rata, serta lama masa kerja yang cukup signifikan. Mereka juga berasal dari departemen strategis seperti Sales, Operations, dan Technology, serta menunjukkan partisipasi aktif dalam program pelatihan internal.



Gambar 10. Grafik Batang karyawan yang dipromosikan

Temuan ini menunjukkan bahwa promosi karyawan cenderung diberikan kepada mereka yang menunjukkan konsistensi dalam performa, partisipasi aktif dalam pelatihan, serta pengalaman kerja yang cukup panjang. Meskipun penghargaan dan jumlah pelatihan lebih dari satu tidak dialami oleh mayoritas, indikator tersebut tetap menjadi pembeda yang cukup signifikan antara karyawan yang dipromosikan dan yang tidak.

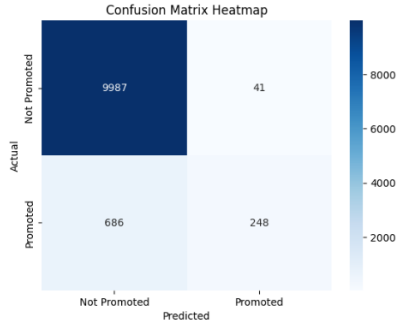
Namun demikian, meskipun model Random Forest memiliki precision yang tinggi, nilai recall yang dihasilkan masih tergolong rendah, yaitu 26,5%. Hal ini menunjukkan bahwa model masih belum sepenuhnya sensitif dalam mendeteksi seluruh karyawan yang seharusnya dipromosikan. Rendahnya recall ini kemungkinan besar disebabkan oleh ketidakseimbangan kelas pada dataset, di mana hanya sekitar 8,5% data berlabel is_promoted = 1. Meskipun telah dilakukan penanganan ketidakseimbangan data dengan metode SMOTE, metode ini hanya berdampak pada data pelatihan dan tidak selalu menjamin perbaikan kinerja pada data uji. Selain itu, kemungkinan adanya tumpang tindih karakteristik antara kelas minoritas dan mayoritas turut menyulitkan model dalam memisahkan kedua kelompok secara akurat.

Selain aspek teknis model, hasil interpretasi fitur penting juga dapat digunakan secara praktis oleh manajemen sumber daya manusia (HR). Fitur seperti avg_training_score, previous_year_rating, dan length_of_service terbukti memiliki korelasi tinggi terhadap peluang promosi, sehingga dapat dijadikan dasar dalam merancang kebijakan SDM. Beberapa implikasi praktis yang bisa dilakukan antara lain: menyusun jalur karier berbasis performa nyata, mendorong karyawan untuk aktif dalam pelatihan internal, menetapkan sistem penghargaan berbasis KPI yang terukur, serta mengidentifikasi talenta potensial lebih dini untuk difasilitasi dalam program pengembangan karier. Dengan demikian, penerapan machine learning dalam konteks ini tidak hanya bersifat prediktif, tetapi juga dapat mendorong praktik manajemen SDM yang lebih objektif, adil, dan berbasis data.

4.5 Visualisasi dan Interpretasi Confusion Matrix

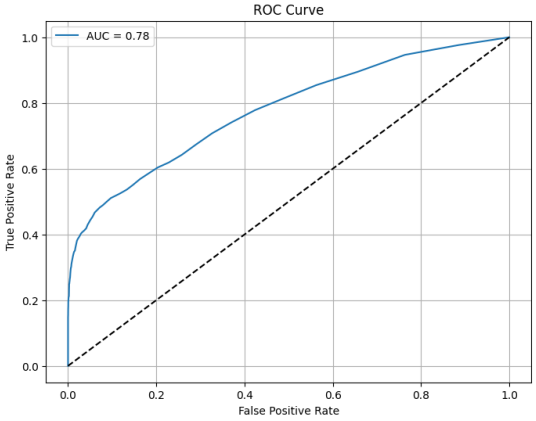
Confusion matrix divisualisasikan dalam bentuk heatmap. Hasil menunjukkan bahwa model mampu mengidentifikasi 9987 karyawan yang tidak dipromosikan dan 248 yang dipromosikan secara tepat. Namun,

terdapat 686 kasus false negative dan 41 false positive. Ini menandakan bahwa meskipun model cukup akurat dalam memprediksi kelas mayoritas, sensitivitas terhadap kelas minoritas masih perlu ditingkatkan.



Gambar 11. Grafik Confusion matric Heatmap evaluasi model

Kondisi ini menunjukkan bahwa model memiliki kemampuan cukup baik dalam mengenali karyawan yang tidak dipromosikan, namun masih perlu ditingkatkan dalam mengenali karyawan yang seharusnya dipromosikan. Hal ini wajar mengingat dataset memiliki karakteristik imbalanced, yaitu distribusi data yang tidak seimbang antara kelas promosi dan tidak promosi. Oleh karena itu, meskipun nilai akurasi tergolong tinggi, perhatian khusus perlu diberikan terhadap metrik lain seperti recall untuk kelas minoritas agar sistem tidak bias terhadap kelas mayoritas.



Gambar 12. Grafik Kurva ROC Model Random Forest

Nilai ROC AUC sebesar 0,78 memperkuat bukti bahwa model memiliki kemampuan yang baik dalam membedakan antara kelas promosi dan tidak promosi.Dengan performa seperti ini, model yang dibangun cukup dapat diandalkan sebagai sistem pendukung keputusan bagi manajemen sumber daya manusia (HR), terutama dalam menyaring kandidat yang layak dipromosikan berdasarkan data historis dan kinerja aktual.

4.6 Perbandingan Kinerja Model

Untuk memperkuat temuan, dilakukan perbandingan model Random Forest dengan dua algoritma lain, yaitu Logistic Regression dan Decision Tree. Hasil evaluasi disajikan pada Tabel berikut:

Tabel 4. 1 Perbandingan Kinerja Model

Model	**Akurasi**	**Precision**	**Recall**	**F1-Score**	**AUC**
Logistic Regression	91,2%	74,8%	15,4%	25,6%	0,71
Decision Tree	92,5%	79,2%	19,6%	31,1%	0,74
Random Forest	**93,0%**	**85,1%**	**26,5%**	**40,5%**	**0,78**

Dari hasil tersebut, model Random Forest unggul pada hampir semua metrik, terutama precision dan AUC. Hal ini menunjukkan bahwa Random Forest lebih mampu menangkap kompleksitas data dan menghasilkan prediksi yang lebih andal. Sementara Logistic Regression memiliki kinerja terendah dalam hal recall dan F1-score, menjadikannya kurang ideal untuk kasus dengan data tidak seimbang.

4.7 Pembahasan

Nilai recall yang masih rendah meskipun telah diterapkan SMOTE mengindikasikan bahwa diperlukan pendekatan tambahan seperti hybrid sampling atau penalized learning. Selain itu, proses tuning hyperparameter seperti n_estimators, max_depth, dan class_weight perlu dieksplorasi lebih lanjut. Secara praktis, model ini memberikan wawasan berharga bagi pengambil kebijakan HR. Faktor-faktor seperti skor pelatihan dan rating performa sebelumnya dapat dijadikan indikator objektif dalam proses promosi. Dengan demikian, integrasi model ini dalam sistem manajemen SDM dapat meningkatkan transparansi dan keadilan dalam promosi karyawan.

4. Kesimpulan

Penelitian ini berhasil membangun model prediksi promosi karyawan menggunakan algoritma Random Forest dengan pendekatan CRISP-DM, yang mencakup tahapan pemahaman bisnis, eksplorasi data, pemodelan, hingga evaluasi. Untuk mengatasi masalah ketidakseimbangan kelas dalam data, diterapkan teknik SMOTE (Synthetic Minority Oversampling Technique) yang terbukti efektif dalam meningkatkan sensitivitas model terhadap kelas minoritas.Hasil evaluasi model menunjukkan performa prediktif yang cukup baik, dengan nilai akurasi sebesar 93%, precision 85,13%, recall 26,55%, F1-score 40,55%, dan ROC AUC 0,78. Meskipun nilai recall masih tergolong rendah, model mampu mengidentifikasi kandidat promosi dengan tingkat presisi tinggi, menjadikannya sebagai alat bantu pengambilan keputusan yang andal, terutama ketika mempertimbangkan risiko promosi yang salah sasaran.Berdasarkan analisis feature importance, faktor-faktor seperti previous_year_rating, avg_training_score, length_of_service, KPIs_met >80%, dan awards_won merupakan variabel yang paling berpengaruh dalam penentuan promosi. Hal ini menegaskan pentingnya integrasi data kinerja historis dalam proses pengambilan keputusan SDM yang lebih objektif.Sebagai pengembangan lebih lanjut, disarankan untuk melakukan tuning hyperparameter, mengeksplorasi algoritma alternatif seperti XGBoost atau ensemble models, serta memperluas cakupan data dengan menambahkan fitur kontekstual dan variabel eksternal. Implementasi sistem dalam bentuk dashboard interaktif juga direkomendasikan agar hasil model dapat digunakan secara praktis oleh manajemen dalam mendukung kebijakan promosi yang adil, transparan, dan berbasis data

Daftar Rujukan

[1] F. Elfaladonna, V. E. Putri, and D. Apryanty, “Analisis Algoritma Naïve Bayes Untuk Pendukung Keputusan Promosi Jabatan (PT. Xyz),” 2024.

[2] L. Madaerdo Sotarjua, D. Budhi Santoso, U. H. Singaperbangsa Karawang Jl Ronggo Waluyo, K. Telukjambe Timur, K. Karawang, and J. Barat, “PERBANDINGAN ALGORITMA KNN, DECISION TREE,*DAN RANDOM*FOREST PADA DATA IMBALANCED CLASS UNTUK KLASIFIKASI PROMOSI KARYAWAN,” vol. 7, no. 2, p. 2022.

[3] Miftahuljannah, M., Sunge, A. S., & Zy, A. T. (2023). Analisis prediksi penjualan dengan metode regresi linear di PT. Eagle Industry Indonesia. JINTEKS (Jurnal Informatika Teknologi dan Sains), 5(3), 398–403

[4]. P. C. Promosi Jabatan Karyawan Dengan Algoritma, S. Kasus, and A. Senayan Jakarta, “Prediction of Employee Position Promotion Using C4.5 Algorithm (Case Study: Senayan Apartment Jakarta),” 2019.

[5] Sangaji, D., & Sutabri, T. (2025). *Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla*. Jurnal Pustaka AI, 5(1), 13–18.

[6] Sari, D. P., Buana, W., & Sari, M. F. M. (2025). *Implementasi Data Mining pada Penjualan Barang dengan Teknik K-Means*. Jurnal Pustaka AI, 5(1), 106–112.

[7] Pangarso, A. W., Kurniati, N., & Editya, A. S. (2025). *Analisis Kesetiaan Pelanggan pada Penjualan Filter Rokok di CV. Karunia Abadi dengan Metode K-Means*. Jurnal Pustaka AI, 5(1), 58–63.

[8] M. N. Fahmi, “Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning),” Sains Data Jurnal Studi Matematika dan Teknologi, vol. 1, no. 2, pp. 87–96, Dec. 2023, doi: 10.52620/sainsdata.v1i2.31.

[9] M. N. Tentua and S. Sumarmi, “Prediksi Promosi Pegawai Menggunakan Metode Extremely Randomized Trees,” Jurnal Dinamika Informatika, vol. 12, no. 2, 2023, [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html

[10] A. Maehendrayuga, A. Setyanto, and Kusnawi, “Analisa Prediksi Turnover Karyawan menggunakan Machine Learning,” bit-Tech, vol. 7, no. 2, pp. 648–659, Dec. 2024, doi: 10.32877/bt.v7i2.1999.

[11] D. Fernando and R. G. Guntara, “Model Klasifikasi Penyebab Turnover Karyawan Menggunakan Kerangka Kerja CRISP-DM”.

[12] I. Putu, R. Paramaditya, and C. Pramartha, “Implementasi Algoritma Random Forest Dalam Menentukan Kualitas Susu Sapi,” 2022. [Online]. Available: https://www.kaggle.com/datasets/yrohit199/milk-quality

[13] S. A1 and Meiriyama, “Penerapan Algoritma Random Forest untuk Klasifikasi Jenis Daun Herbal,” 2023.

[14] D. N. Handayani and S. Qutub, “Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan,” RIGGS: Journal of Artificial Intelligence and Digital Business, vol. 4, no. 2, pp. 405–412, May 2025, doi: 10.31004/riggs.v4i2.512.

[15] F. T. Kurniati and D. Pramana, “Identifikasi Objek Menggunakan Random Forest dan Multi-Fitur,” 2023.

[16]N. Hidayah, "8 | Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan … Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," Jurnal Akademik Pendidikan Matematika, vol. 10, no. 1, 2024, doi: 10.55340/japm.v10i1.1491.

[17] S. Clara dkk., Implementasi Seleksi Fitur Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Penghasilan Pada Adult Income Dataset. 2021.

[18] R. Gelar Guntara, "Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab," Jurnal Ilmiah Multidisiplin, vol. 2, no. 6, 2023.

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025