

# Penerapan Random Forest dalam Klasifikasi Promosi Karyawan Berdasarkan Dataset HR Analytics Vidya

1<sup>st</sup> Dhi'fan Razaqa

Departemen Informatika,  
Fakultas Ilmu Komputer  
Universitas Pembangunan  
Nasional "Veteran" Jakarta  
Jakarta, Indonesia

[2210511091@mahasiswa.upnvj.ac.id](mailto:2210511091@mahasiswa.upnvj.ac.id)

2<sup>nd</sup> Fajar Ramadhan

Departemen Informatika,  
Fakultas Ilmu Komputer  
Universitas Pembangunan  
Nasional "Veteran" Jakarta  
Jakarta, Indonesia

[2210511097@mahasiswa.upnvj.ac.id](mailto:2210511097@mahasiswa.upnvj.ac.id)

3<sup>rd</sup> Muhammad Dwi Ananda

Departemen Informatika,  
Fakultas Ilmu Komputer  
Universitas Pembangunan  
Nasional "Veteran" Jakarta  
Jakarta, Indonesia

[2210511119@mahasiswa.upnvj.ac.id](mailto:2210511119@mahasiswa.upnvj.ac.id)

4<sup>th</sup> Karenina Nurmelita Malik

Departemen Informatika,  
Fakultas Ilmu Komputer  
Universitas Pembangunan  
Nasional "Veteran" Jakarta  
Jakarta, Indonesia

[2210511089@mahasiswa.upnvj.ac.id](mailto:2210511089@mahasiswa.upnvj.ac.id)

## I. PENDAHULUAN

Dalam era transformasi digital saat ini, persaingan bisnis semakin kompetitif sehingga perusahaan dituntut untuk mampu mengelola sumber daya manusianya secara efektif guna menjaga atau meningkatkan kualitas karyawannya. Salah satu strategi yang umum dilakukan adalah memberikan promosi sebagai bentuk peningkatan jenjang karier bagi karyawan untuk meningkatkan motivasi karyawan [1]. Namun, pada praktiknya, proses promosi sering menghadapi kendala, seperti dinilai terlalu subjektif, kurang transparansi, serta kriteria yang kurang jelas. Masalah ini menimbulkan dampak negatif terhadap motivasi karyawan, yang pada akhirnya menurunkan performa karyawan tersebut [2].

Maka dari itu, diperlukan pendekatan khusus untuk mengatasi masalah tersebut. Pendekatan berbasis data dapat menjadi solusi yang menjanjikan. Memanfaatkan data-data historis karyawan, perusahaan dapat melakukan *big data analytics* pada data tersebut untuk membangun sistem yang dapat mengidentifikasi apakah seorang karyawan berpotensi mendapatkan promosi secara lebih objektif [3]. Melalui pendekatan ini, perusahaan dapat menciptakan keputusan yang objektif, adil, dan transparan.

Teknologi yang dapat digunakan untuk mewujudkan hal ini adalah *machine learning*. *Machine learning* memungkinkan program untuk belajar mengenali pola dari data historis untuk memprediksi suatu keputusan [4]. Pada kasus identifikasi apakah seorang karyawan mendapatkan promosi atau tidak, *machine learning* digunakan untuk membangun model yang dapat melakukan klasifikasi guna memprediksi kemungkinan seorang karyawan mendapat promosi atau tidak berdasarkan fitur-fitur yang telah dipelajari dari data historis. Implementasi teknologi ini tidak hanya membuat proses promosi menjadi objektif, adil, dan transparan, tetapi juga membantu proses promosi menjadi lebih efisien.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi untuk melakukan klasifikasi promosi karyawan menggunakan *machine learning* dengan memanfaatkan *big data*. Pada penelitian ini, algoritma yang digunakan untuk melakukan klasifikasi adalah *Random Forest*.

## II. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan pendekatan kuantitatif berbasis data historis karyawan yang bersumber dari dataset "HR Analytics" yang tersedia di Kaggle. Tujuan utama dari penelitian ini adalah membangun model klasifikasi yang mampu memprediksi apakah seorang karyawan berpotensi mendapatkan promosi atau tidak yang diukur berdasarkan atribut-atribut yang terdapat dalam dataset.

### A. Akuisisi Data

Dataset yang digunakan dalam penelitian ini bersumber dari Kaggle dengan judul "HR Analytics" yang diambil dari Analytics Vidya. Dataset ini berisi berbagai atribut karyawan, seperti pendidikan, masa kerja, pengalaman sebelumnya, pelatihan, dan status promosi. Dataset ini mewakili skenario nyata dan mencerminkan tantangan dalam pengambilan keputusan sumber daya manusia berbasis data.

### B. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk mempersiapkan dataset agar dapat digunakan dalam proses pelatihan model. Proses ini melibatkan beberapa langkah seperti penanganan *missing values* atau data kosong untuk memastikan kelengkapan data, transformasi fitur kategorikal ke dalam bentuk numerik, dan normalisasi fitur numerik apabila diperlukan untuk menjaga konsistensi skala antar fitur [5].

### C. Eksplorasi dan Data Analisis (EDA)

Exploratory Data Analysis dilakukan untuk memahami lebih dalam karakteristik dan pola dari dataset yang digunakan. Analisis ini mencakup visualisasi pada tiap fitur terhadap target, analisis korelasi antar fitur numerik, identifikasi nilai-nilai outlier, serta eksplorasi awal terhadap fitur-fitur yang berpotensi memiliki pengaruh signifikan terhadap hasil prediksi. Proses EDA dalam data HR juga penting dilakukan untuk memahami karakteristik data dan mengidentifikasi fitur-fitur utama yang mempengaruhi hasil prediksi, termasuk dalam konteks promosi karyawan {cite}.

### D. Pemodelan

Pada penelitian ini digunakan pemodelan dengan teknik algoritma Random Forest Classifier, yaitu sebuah metode

*ensemble* yang menggabungkan prediksi dari sejumlah algoritma Decision Tree untuk menghasilkan klasifikasi akhir yang lebih akurat dan stabil. Hal ini juga diperkuat oleh penelitian yang menunjukkan bahwa Random Forest memiliki akurasi yang lebih tinggi dibandingkan Decision Tree dalam berbagai kasus klasifikasi, salah satunya pada penelitian yang dilakukan oleh [6] dalam topik data kesehatan. Klasifikasi menggunakan Random Forest mencapai akurasi hingga 99%, mengungguli Decision Tree yang hanya mencapai sekitar 75%.

Algoritma ini dipilih dengan alasan kemampuannya dalam menangani fitur numerik dan kategorikal secara bersamaan, ketahanan terhadap overfitting, dan kapabilitasnya dalam memberikan interpretasi melalui *feature importance*. Hal ini juga didukung oleh penelitian pada [7] yang menunjukkan bahwa Random Forest cocok untuk analisis prediktif dalam data HR, termasuk prediksi promosi, karena dapat menangani berbagai jenis fitur dan memberikan hasil klasifikasi yang stabil.

Data dibagi menggunakan metode train-test split dengan rasio 80:20, di mana sejumlah 80% data digunakan untuk pelatihan dan 20% data digunakan untuk pengujian. Selain itu, dilakukan tuning terhadap beberapa hyperparameter, kedalaman maksimum pohon, dan kriteria pemisahan untuk memperoleh performa model yang optimal.

#### E. Evaluasi Model

Evaluasi model dilakukan dengan menggunakan beberapa metrik atau pengukuran umum diterapkan dalam klasifikasi, yaitu akurasi, presisi, recall, dan F1-score [8]. Pengukuran-pengukuran ini memberikan gambaran mengenai seberapa baik model dalam mengklasifikasikan data secara tepat dan seimbang. Selain itu, Random Forest juga menyediakan informasi mengenai setiap fitur yang dapat dimanfaatkan untuk mengidentifikasi atribut yang paling berpengaruh dalam menentukan keputusan promosi karyawan.

#### DAFTAR PUSTAKA

- [1] F. Syarief, K. P. Utomo, M. A. Winardi, and A. Kurniawan, "PENGARUH PROMOSI JABATAN DAN KUALITAS TERHADAP MOTIVASI PEMERINTAH KOTA BEKASI," *JURISMA J. Ris. Bisnis Manaj.*, vol. 11, no. 2, pp. 207–224, Oct. 2021, doi: 10.34010/jurisma.v11i2.2872.
- [2] G. Tachie-Donkor, T. G. Mensah, and E. A. Nsowah, "Staff Promotion and Employee Commitment to Work: A Survey in a Ghanaian University Library," *J. Advocacy Res. Educ.*, vol. 11, no. 2, Aug. 2024, doi: 10.13187/jare.2024.2.222.
- [3] R. M. Sari, "PEMANFAATAN ARTIFICIAL INTELLEGENCE (AI) PADA PROSES PENGAMBILAN KEPUTUSAN MANAJEMEN: MENGAJI TREN, PELUANG DAN TANTANGAN," *J. Innov. Res. Knowl.*, vol. 4, no. 5, Art. no. 5, Oct. 2024.
- [4] A. Gevindo and B. Hendrik, "PENERAPAN MACHINE LEARNING UNTUK MENDETEKSI SERANGAN ANOMALI DALAM JARINGAN KOMPUTER : SYSTEMATIC LITERATURE REVIEW," vol. 9, no. 3, 2025.
- [5] S. F. Sari and K. M. Lhaksana, "Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification," *J. Comput. Syst. Inform. JoSYC*, vol. 3, no. 4, Art. no. 4, Sep. 2022, doi: 10.47065/josyc.v3i4.2099.
- [6] H. Oktavianto, H. W. Sulisty, G. Wijaya, D. Irawan, and G. Abdurrahman, "Analisis Perbandingan Decision Tree dan Random Forest Pada Klasifikasi Teks Data Kesehatan," *BINA INSANI ICT J.*, vol. 11, no. 1, pp. 56–65, Jun. 2024, doi: 10.51211/biict.v11i1.2928.

- [7] S. L. V. Papineni, M. R. A. S. Yarlagadda, S. Yarlagadda, and H. Akkineni, "An Extensive Analytical Approach on Human Resources using Random Forest Algorithm," *Int. J. Eng. Trends Technol.*, vol. 69, no. 5, pp. 119–127, May 2021, doi: 10.14445/22315381/IJETT-V69I5P217.
- [8] S. K. and S. Sidharth, "HR Analytics: Employee Attrition Analysis using Random Forest," *Int. J. Perform. Eng.*, vol. 18, no. 4, p. 275, Apr. 2022, doi: 10.23940/ijpe.22.04.p5.275281.