

# Klasifikasi Kinerja Pegawai Menggunakan Algoritma Random Forest

Jelly Helia Putri <sup>1\*</sup>, Sunarsan Sitohang <sup>2</sup>

<sup>1</sup> Universitas Putra Batam; pd210210100@upbatam.ac.id

<sup>2</sup> Universitas Putra Batam; sunarsan@puterabatam.ac.id

**Abstrak:** Evaluasi kinerja pegawai merupakan aspek penting dalam manajemen sumber daya manusia, namun penerapannya masih sering dilakukan secara subjektif sehingga berpotensi menimbulkan bias dan ketidakkonsistenan. Penelitian-penelitian sebelumnya umumnya berfokus pada penerapan algoritma klasifikasi tanpa menekankan kesenjangan penelitian, kontribusi ilmiah, serta analisis faktor dominan yang memengaruhi kinerja pegawai. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi kinerja pegawai yang objektif serta mengidentifikasi faktor-faktor utama penentu kinerja pegawai menggunakan algoritma Random Forest. Penelitian ini menggunakan dataset *INX Future Inc. Employee Performance* yang diperoleh dari Kaggle dengan tiga kategori kinerja, yaitu Low, Medium, dan High. Proses penelitian dilakukan melalui tahapan data mining yang meliputi seleksi data, prapemrosesan, transformasi data, pemodelan, dan evaluasi. Performa model dievaluasi menggunakan metrik akurasi, presisi, recall, dan f1-score, serta dibandingkan berdasarkan beberapa skenario pembagian data latih dan data uji. Hasil penelitian menunjukkan bahwa pembagian data latih dan data uji sebesar 80%:20% menghasilkan performa terbaik dengan tingkat akurasi sebesar 94,58%, yang lebih unggul dibandingkan skenario lainnya. Analisis *feature importance* menunjukkan bahwa kepuasan lingkungan kerja, sistem reward, pengembangan karier, dan pengalaman kerja merupakan faktor yang paling berpengaruh terhadap kinerja pegawai. Simpulan penelitian ini menegaskan bahwa algoritma Random Forest efektif dalam klasifikasi kinerja pegawai sekaligus memberikan kontribusi ilmiah dalam mengidentifikasi faktor penentu kinerja, serta implikasi praktis sebagai sistem pendukung keputusan evaluasi kinerja pegawai yang lebih objektif dan berbasis data.

**Keywords:** Kinerja Pegawai; Random Forest; Machine Learning; Data Mining; Klasifikasi

DOI: <https://doi.org/10.47134/jacis.v6i1.170>

\*Correspondensi: Jelly Helia Putri

Email: [pb210210100@upbatam.ac.id](mailto:pb210210100@upbatam.ac.id)

Receive: 12 Januari 2026

Accepted: 06 Maret 2026

Published: 11 Maret 2026



**Copyright:** © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstrak:** Employee performance evaluation plays a crucial role in human resource management; however, it is often conducted subjectively, leading to potential bias and inconsistency. Previous studies have mainly focused on the application of classification algorithms without clearly addressing research gaps, scientific contributions, or the analysis of dominant factors influencing employee performance. Therefore, this study aims to develop an objective employee performance classification model and identify key factors affecting employee performance using the Random Forest algorithm. This study utilizes the *INX Future Inc. Employee Performance* dataset obtained from Kaggle, consisting of three performance categories: Low, Medium, and High. The research process follows data mining stages, including data selection, preprocessing, transformation, modeling, and evaluation. Model performance is assessed using accuracy, precision, recall, and f1-score metrics and compared across several training and testing data split scenarios. The results indicate that the

80% training and 20% testing data split achieves the best performance with an accuracy of 94.58%, outperforming other scenarios. Furthermore, feature importance analysis reveals that work environment satisfaction, reward systems, career development, and work experience are the most influential factors affecting employee performance. This study concludes that the Random Forest algorithm is effective for employee performance classification and provides scientific contributions in identifying key performance determinants, as well as practical implications as a data-driven decision support system for objective employee performance evaluation.

**Keywords:** Employee Performance; Random Forest; Machine Learning; Data Mining; Classification

---

## PENDAHULUAN

Kinerja pegawai merupakan indikator utama dalam menilai efektivitas pengelolaan sumber daya manusia dan keberhasilan organisasi. Berbagai penelitian terdahulu menunjukkan bahwa sistem penilaian kinerja yang masih dilakukan secara manual dan berbasis penilaian subjektif berpotensi menghasilkan evaluasi yang tidak konsisten serta kurang akurat[1], sehingga berdampak pada kualitas pengambilan keputusan manajerial, seperti penentuan promosi, pemberian insentif, dan pengembangan karier pegawai[2]. Kondisi ini menjadi permasalahan serius terutama pada organisasi dengan jumlah pegawai yang besar dan karakteristik data kinerja yang kompleks.

Oleh karena itu, diperlukan pendekatan penilaian kinerja yang objektif dan berbasis data untuk meminimalkan bias subjektivitas serta meningkatkan keandalan hasil evaluasi. Pemanfaatan metode *machine learning* dalam evaluasi kinerja pegawai menawarkan solusi ilmiah yang mampu mengolah berbagai atribut kinerja secara simultan dan menghasilkan klasifikasi kinerja yang lebih konsisten sebagai dasar pengambilan keputusan manajemen[3].

Perkembangan teknologi informasi mendorong pemanfaatan *machine learning* dalam evaluasi kinerja pegawai, khususnya untuk mendukung keputusan manajemen sumber daya manusia yang bersifat krusial, seperti penentuan promosi, pemberian insentif, perencanaan pengembangan karier, dan identifikasi pegawai yang memerlukan pembinaan kinerja. Pendekatan berbasis *machine learning* memungkinkan pengolahan atribut kinerja yang kompleks dan multidimensi secara lebih sistematis, sehingga dapat mengurangi subjektivitas penilaian dan meningkatkan konsistensi keputusan manajerial[4]. Penelitian terdahulu telah menerapkan berbagai algoritma klasifikasi, seperti *Decision Tree*[5], *Naive Bayes*[6], *Support Vector Machine*[7], dan *K-Nearest Neighbor*[8], untuk mengelompokkan kinerja pegawai berdasarkan indikator tertentu. Namun, sebagian besar studi tersebut masih menghadapi keterbatasan, antara lain performa yang kurang stabil pada data dengan distribusi kelas tidak seimbang[8], sensitivitas terhadap *noise* data[7], serta keterbatasan dalam memberikan interpretasi faktor dominan yang memengaruhi hasil klasifikasi. Keterbatasan ini menyebabkan hasil evaluasi belum sepenuhnya optimal sebagai dasar pengambilan keputusan strategis SDM, sehingga diperlukan pendekatan klasifikasi yang

lebih robust sekaligus interpretatif untuk mendukung evaluasi kinerja pegawai secara objektif dan komprehensif[9],[10].

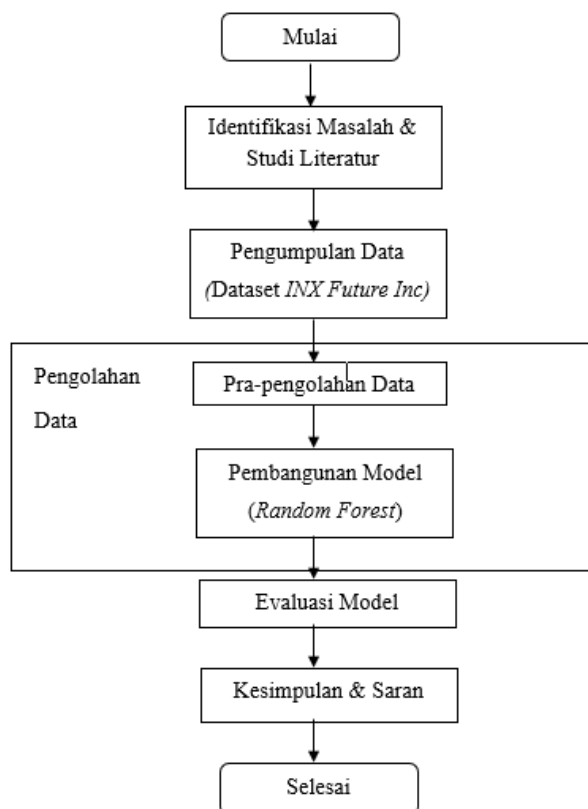
Meskipun berbagai pendekatan klasifikasi telah diterapkan dalam evaluasi kinerja pegawai, sejumlah penelitian terdahulu masih menunjukkan keterbatasan yang spesifik. Beberapa studi hanya mengelompokkan kinerja pegawai ke dalam dua kategori, seperti *high* dan *low*[11][12], atau tiga kategori tanpa mempertimbangkan distribusi kelas yang tidak seimbang[8][12]. Selain itu, evaluasi performa model pada penelitian-penelitian tersebut umumnya masih terbatas pada penggunaan metrik akurasi, tanpa disertai metrik lain seperti precision, recall, f1-score, atau confusion matrix yang lebih representatif dalam menilai kinerja model klasifikasi. Di sisi lain, pemanfaatan algoritma *ensemble* pada beberapa penelitian sebelumnya masih belum dioptimalkan untuk menangani kompleksitas data dan ketidakseimbangan kelas secara efektif. Temuan-temuan tersebut, sebagaimana dilaporkan dalam berbagai studi sebelumnya, menunjukkan adanya celah penelitian yang membuka peluang pengembangan model klasifikasi kinerja pegawai yang lebih robust, komprehensif, dan evaluatif[13].

Berdasarkan celah penelitian yang ada, penelitian ini berfokus pada klasifikasi kinerja pegawai ke dalam tiga kategori, yaitu *Low*, *Medium*, dan *High*, menggunakan algoritma Random Forest. Kebaruan penelitian terletak pada pengembangan kerangka evaluasi kinerja pegawai berbasis data yang mengombinasikan analisis performa model melalui berbagai skenario pembagian data latih dan data uji serta interpretasi faktor penentu kinerja melalui analisis *feature importance*. Pendekatan ini tidak hanya menghasilkan klasifikasi yang akurat, tetapi juga memberikan pemahaman terhadap kontribusi setiap atribut kinerja. Penelitian ini bertujuan mengembangkan model klasifikasi kinerja pegawai yang *robust* dan interpretatif sebagai kontribusi akademik dalam pengembangan metode evaluasi kinerja pegawai berbasis machine learning, serta sebagai sistem pendukung keputusan yang objektif dalam praktik manajemen sumber daya manusia.

## METODE

### Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk klasifikasi kinerja pegawai, karena metode ini mampu mengolah data kinerja yang bersifat multidimensi, kompleks, dan berjumlah besar[14], serta lebih adaptif dibandingkan metode statistik konvensional dalam menangani hubungan nonlinier dan ketidakseimbangan kelas[15]. Alur penelitian dilakukan secara sistematis, diawali dengan identifikasi masalah dan studi literatur untuk merumuskan fokus penelitian, dilanjutkan dengan pengumpulan data menggunakan dataset *INX Future Inc. Employee Performance*. Tahap pra-pengolahan data mencakup pembersihan, penanganan data tidak lengkap, dan transformasi data agar siap digunakan dalam pemodelan. Selanjutnya, model klasifikasi dibangun menggunakan algoritma Random Forest untuk mengelompokkan kinerja pegawai ke dalam kategori *Low*, *Medium*, dan *High*. Tahap akhir penelitian meliputi evaluasi performa model menggunakan metrik yang relevan serta penarikan kesimpulan dan penyusunan saran berdasarkan hasil evaluasi yang diperoleh.



Gambar 1 Diagram Alur Penelitian

### Konsep Kinerja Pegawai

Kinerja pegawai diposisikan sebagai variabel dependen yang merepresentasikan tingkat pencapaian pegawai dalam melaksanakan tugas dan tanggung jawabnya berdasarkan data kinerja yang tersedia pada dataset. Kinerja pegawai dioperasionalkan sebagai variabel klasifikasi dengan tiga kategori, yaitu *Low*, *Medium*, dan *High*, yang ditentukan berdasarkan atribut-atribut kinerja pegawai yang telah melalui proses pra-pengolahan data. Kategorisasi ini digunakan sebagai dasar dalam pembangunan dan evaluasi model klasifikasi untuk menganalisis pola kinerja pegawai secara objektif dan berbasis data[16][17].

### Faktor dan Indikator Kinerja Pegawai

Dalam penelitian ini, faktor-faktor yang memengaruhi kinerja pegawai dioperasionalkan melalui atribut-atribut yang tersedia pada dataset *INX Future Inc. Employee Performance*. Faktor motivasi dan kepuasan kerja direpresentasikan oleh atribut kepuasan lingkungan kerja dan sistem reward, sedangkan kemampuan dan kompetensi pegawai diwakili oleh atribut tingkat pendidikan, pengalaman kerja, serta pengembangan karier. Aspek disiplin dan kehadiran dicerminkan melalui atribut ketepatan waktu dan kehadiran, sementara beban kerja dan dukungan organisasi direpresentasikan oleh atribut tekanan kerja dan dukungan manajemen. Seluruh atribut tersebut digunakan sebagai variabel input dalam pembangunan model klasifikasi, dengan kinerja pegawai sebagai variabel target yang dikategorikan ke dalam kelas *Low*, *Medium*, dan *High*.

### Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset *INX Future Inc. Employee Performance* yang tersedia di platform Kaggle. Dataset ini

dipilih karena memiliki data kinerja pegawai yang terstruktur, jumlah observasi yang memadai (1.200 data), serta atribut yang merepresentasikan karakteristik individu, pekerjaan, dan organisasi sehingga relevan untuk klasifikasi kinerja pegawai berbasis machine learning. Kinerja pegawai digunakan sebagai variabel dependen, sedangkan atribut-atribut terkait digunakan sebagai variabel independen. Meskipun demikian, penggunaan dataset publik memiliki keterbatasan, seperti keterbatasan konteks organisasi yang spesifik serta ketergantungan pada kualitas dan kelengkapan data yang tersedia, sehingga hasil penelitian diinterpretasikan sesuai dengan konteks dataset yang digunakan[18].

### **Pra-pengolahan Data**

Tahap pra-pengolahan data dilakukan untuk memastikan kualitas data sebelum pemodelan. Proses ini meliputi pembersihan data dengan menghapus data duplikat dan inkonsistensi, serta penanganan nilai yang hilang menggunakan metode imputasi berbasis nilai statistik, yaitu pengisian nilai rata-rata untuk atribut numerik dan nilai modus untuk atribut kategorikal. Atribut kategorikal kemudian dikonversi ke dalam format numerik menggunakan teknik *label encoding* agar dapat diproses oleh algoritma klasifikasi. Tahapan pra-pengolahan ini bertujuan untuk meningkatkan konsistensi dan kualitas data sehingga model klasifikasi dapat dibangun secara optimal dan dapat direproduksi[19].

### **Pembangunan Model Random Forest**

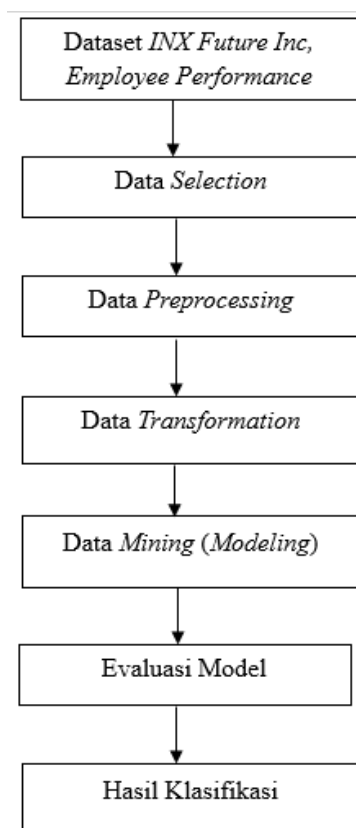
Model klasifikasi kinerja pegawai dikembangkan menggunakan algoritma Random Forest. Algoritma ini dipilih karena mampu mengatasi data dengan jumlah atribut yang banyak dan hubungan nonlinier[20], serta cenderung lebih stabil dibandingkan algoritma klasifikasi tunggal seperti Decision Tree yang rentan terhadap overfitting. Selain itu, dibandingkan metode lain seperti Naive Bayes dan K-Nearest Neighbor, Random Forest lebih adaptif terhadap data yang kompleks dan tidak seimbang[21], serta mampu memberikan informasi tambahan melalui analisis *feature importance*. Sejumlah penelitian empiris sebelumnya juga menunjukkan bahwa Random Forest menghasilkan performa yang lebih konsisten dalam klasifikasi data sumber daya manusia[22]. Data dibagi ke dalam data pelatihan dan data pengujian dengan beberapa skenario pembagian untuk memperoleh konfigurasi model yang optimal dalam mengklasifikasikan kinerja pegawai ke dalam kategori *Low*, *Medium*, dan *High*[23].

### **Evaluasi Model**

Evaluasi model dilakukan untuk menilai kualitas hasil klasifikasi kinerja pegawai menggunakan beberapa metrik evaluasi, yaitu akurasi, presisi, recall, dan f1-score. Akurasi digunakan untuk memberikan gambaran umum terhadap tingkat ketepatan prediksi model, namun karena data kinerja pegawai berpotensi memiliki distribusi kelas yang tidak seimbang, metrik presisi, recall, dan f1-score turut digunakan untuk menilai performa model secara lebih komprehensif pada masing-masing kelas. Selain itu, confusion matrix digunakan untuk menganalisis kesesuaian antara hasil prediksi dan data aktual, sehingga dapat diketahui pola kesalahan klasifikasi yang terjadi. Hasil evaluasi ini menjadi dasar dalam menilai efektivitas model dalam mengklasifikasikan kinerja pegawai secara objektif dan andal[24].

## Interpretasi dan Penarikan Kesimpulan

Tahap akhir penelitian berfokus pada interpretasi hasil klasifikasi untuk mengaitkan performa model dengan faktor-faktor penentu kinerja pegawai melalui analisis *feature importance*. Interpretasi ini digunakan untuk menilai pencapaian tujuan penelitian sekaligus memberikan dasar analitis bagi pengambilan keputusan manajerial, seperti penentuan kebijakan pengembangan pegawai dan sistem reward. Kesimpulan disusun berdasarkan kontribusi ilmiah dan implikasi praktis tersebut serta dilengkapi dengan saran untuk penelitian selanjutnya.



**Gambar 2.** Alur Proses Data Mining pada Penelitian Klasifikasi Kinerja Pegawai

Berdasarkan Gambar 2, penelitian ini dilaksanakan melalui beberapa tahapan utama yang mengikuti proses data mining untuk mengklasifikasikan kinerja pegawai menggunakan algoritma Random Forest. Setiap tahapan dijelaskan sebagai berikut.

### 1. Dataset INX Future Inc. Employee Performance

Dataset *INX Future Inc. Employee Performance* digunakan sebagai sumber data dalam penelitian ini. Dataset tersebut diperoleh dari platform publik Kaggle dan terdiri dari 1200 data pegawai dengan berbagai atribut yang merepresentasikan karakteristik individu, pekerjaan, serta lingkungan organisasi. Data ini digunakan sebagai dasar analisis untuk mengidentifikasi dan mengklasifikasikan tingkat kinerja pegawai.

### 2. Data Selection

Tahap pemilihan data bertujuan untuk menentukan data dan atribut yang sesuai dengan tujuan penelitian. Pada proses ini, dilakukan pemilihan atribut yang terkait dengan faktor-

faktor yang mempengaruhi kinerja pegawai, seperti kondisi lingkungan kerja, pengalaman kerja, insentif, dan faktor pribadi. Dalam penelitian ini, digunakan teknik total sampling, sehingga seluruh data yang tersedia dalam dataset dipakai tanpa melakukan pengurangan atau penghapusan data di tahap awal seleksi.

### **3. Data Preprocessing**

Proses preprocessing data dilakukan untuk memperbaiki kualitas data sebelum digunakan dalam tahap pemodelan. Langkah-langkahnya meliputi membersihkan data dari nilai yang tidak konsisten, mengatasi nilai kosong (missing values), serta mengubah atribut kategorikal menjadi bentuk numerik. Tujuan dari tahap ini adalah agar data berada dalam kondisi yang tepat sehingga dapat diproses secara optimal oleh algoritma machine learning.

### **4. Data Transformation**

Setelah proses pra-pengolahan data selesai, dilakukan transformasi data untuk mengubah format data agar sesuai dan siap digunakan pada tahap pemodelan. Tahap ini meliputi penyesuaian struktur data serta representasi nilai atribut agar sesuai dengan persyaratan algoritma Random Forest yang diterapkan dalam penelitian[25].

### **5. Data Mining (*Modeling*)**

Proses tahap data mining (*modeling*) melibatkan pembuatan model klasifikasi menggunakan algoritma Random Forest. Model ini dibangun berdasarkan data pelatihan untuk mempelajari pola hubungan antara atribut-atribut yang ada dan tingkat kinerja pegawai. Dalam tahap ini, Random Forest dipilih karena kemampuannya dalam mengelola data dengan banyak atribut serta menghasilkan model yang stabil dan akurat. Model yang terbentuk kemudian digunakan untuk mengklasifikasikan kinerja pegawai ke dalam tiga kategori, yaitu Low, Medium, dan High[26].

### **6. Evaluasi Model**

Proses evaluasi model bertujuan untuk menilai kinerja dari model klasifikasi yang telah dibuat. Penilaian dilakukan dengan menggunakan berbagai metrik seperti accuracy, precision, recall, dan F1-score, yang berfungsi untuk mengukur seberapa baik model dalam mengklasifikasikan data uji. Selain itu, confusion matrix juga digunakan untuk melihat distribusi hasil prediksi model dibandingkan dengan data aktual pada setiap kelas kinerja pegawai [27].

### **7. Hasil Klasifikasi**

Tahap akhir dari proses penelitian ini adalah hasil klasifikasi, yaitu keluaran berupa tingkat kinerja pegawai berdasarkan hasil prediksi model Random Forest. Hasil klasifikasi ini digunakan sebagai dasar dalam analisis hasil penelitian serta penarikan kesimpulan mengenai kemampuan model dalam mengklasifikasikan kinerja pegawai sesuai dengan tujuan penelitian.

## HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil pengujian dan pembahasan model Random Forest yang diterapkan pada data kinerja pegawai. Analisis hasil dilakukan untuk mengevaluasi kemampuan model dalam mengklasifikasikan kinerja pegawai secara akurat dan konsisten ke dalam tiga kategori, yaitu Low, Medium, dan High. Evaluasi performa model menggunakan metrik accuracy, precision, recall, dan f1-score bertujuan untuk menilai efektivitas model yang diusulkan serta relevansinya dalam menjawab permasalahan penelitian yang telah dirumuskan pada bagian pendahuluan.

**Tabel 1.** Perbandingan Akurasi Model pada Berbagai Pembagian Data

Pembagian Data	Akurasi (%)
50%:50%	92.33
60%:40%	93.75
70%:30%	93.33
75%:25%	93.33
80%:20%	94.58

Berdasarkan Tabel 1, pembagian data latih dan data uji sebesar 80%:20% menghasilkan tingkat akurasi tertinggi, yaitu 94,58%, dibandingkan dengan skenario lainnya. Hal ini menunjukkan bahwa proporsi data latih yang lebih besar memungkinkan model Random Forest mempelajari pola kinerja pegawai secara lebih representatif, sehingga mampu menangkap hubungan kompleks antaratribut dan menghasilkan prediksi yang lebih stabil pada data uji. Temuan ini menegaskan keunggulan Random Forest dalam menangani klasifikasi kinerja pegawai yang bersifat multidimensi, sekaligus menunjukkan bahwa evaluasi menggunakan beberapa skenario pembagian data memberikan hasil yang lebih valid dan komprehensif dibandingkan pendekatan yang hanya mengandalkan satu skenario pembagian data.

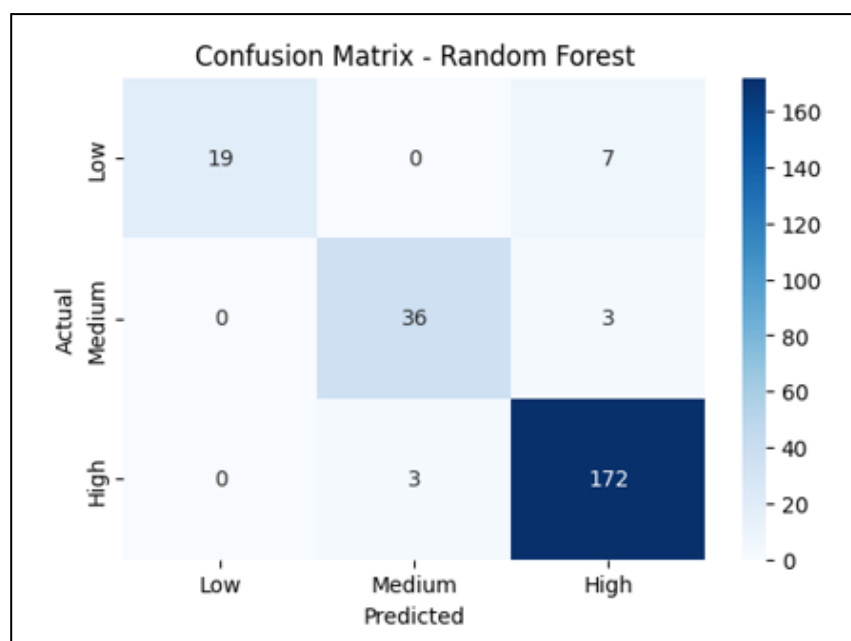
**Tabel 2.** Hasil Evaluasi Model Random Forest

Kelas	Precision	Recall	F1-score	Support
Low	1.000	0.731	0.844	26
Medium	0.923	0.923	0.923	39
High	0.945	0.983	0.964	175
Accuracy			0.946	240
Macro Average	0.956	0.879	0.910	240
Weighted Average	0.947	0.946	0.944	240

Berdasarkan Tabel 2, model Random Forest mencapai nilai akurasi sebesar 0,946 yang menunjukkan kemampuan klasifikasi yang tinggi. Namun, capaian ini perlu dipahami dengan mempertimbangkan karakteristik dataset yang memiliki distribusi kelas tidak seimbang, di mana kelas *High* mendominasi jumlah data dibandingkan kelas *Low* dan *Medium*. Oleh karena itu, penggunaan metrik *weighted average* menjadi lebih relevan dibandingkan *macro average*, karena metrik ini mempertimbangkan proporsi masing-masing kelas sehingga memberikan gambaran kinerja model yang lebih representatif terhadap

kondisi data sebenarnya. Nilai weighted average pada precision, recall, dan f1-score yang masing-masing sebesar 0,947, 0,946, dan 0,944 menunjukkan bahwa model tetap konsisten dalam mengklasifikasikan data meskipun terdapat ketidakseimbangan jumlah sampel antar kelas, sejalan dengan tujuan penelitian untuk menghasilkan model evaluasi kinerja yang robust dan andal.

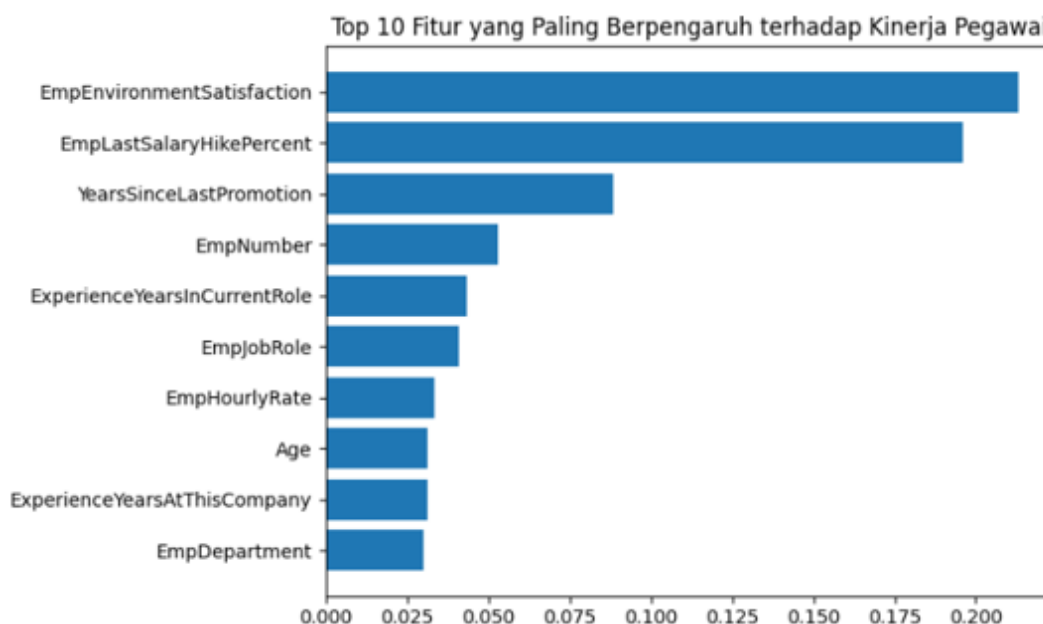
Ditinjau dari performa per kelas, kategori *High* memiliki nilai recall tertinggi (0,983) dan f1-score sebesar 0,964, yang mengindikasikan bahwa pegawai berkinerja tinggi lebih mudah dikenali oleh model. Hal ini diduga disebabkan oleh karakteristik atribut pada kelas *High* yang lebih konsisten dan menonjol, seperti tingkat kepuasan kerja, penghargaan, serta stabilitas kinerja yang relatif lebih baik dibandingkan kelas lainnya. Sebaliknya, pada kelas *Low*, meskipun nilai precision mencapai 1,000, nilai recall yang lebih rendah menunjukkan bahwa tidak seluruh pegawai berkinerja rendah berhasil dikenali oleh model dan masih diklasifikasikan ke kelas lain yang lebih tinggi. Kondisi ini berpotensi menimbulkan implikasi manajerial, seperti tertundanya intervensi atau program pembinaan bagi pegawai yang sebenarnya membutuhkan perhatian khusus. Oleh karena itu, hasil ini menegaskan bahwa selain memiliki performa yang baik secara statistik, interpretasi hasil klasifikasi juga berperan penting dalam mendukung pengambilan keputusan manajerial yang lebih tepat dan berbasis data.



**Gambar 3.** Confusion Matrix Klasifikasi Kinerja Pegawai Menggunakan Random Forest

Berdasarkan Gambar 3, confusion matrix memberikan gambaran pola kesalahan klasifikasi yang tidak terlihat secara langsung pada tabel metrik evaluasi. Kesalahan klasifikasi terutama terjadi antar kelas yang berdekatan, yaitu antara *Low–High* dan *Medium–High*, sementara hampir tidak ditemukan kesalahan prediksi langsung dari kelas *Low* ke *Medium* maupun sebaliknya. Pola ini mengindikasikan adanya tumpang tindih karakteristik atribut kinerja antara pegawai dengan kinerja sedang dan tinggi, serta sebagian kecil pegawai berkinerja rendah yang memiliki atribut mendekati kelas tinggi. Temuan ini menunjukkan bahwa model Random Forest cenderung lebih sensitif dalam mengenali kelas *High*, namun masih menghadapi tantangan dalam membedakan batas transisi antar tingkat kinerja,

sehingga analisis lanjutan terhadap atribut pembeda antar kelas menjadi penting untuk meningkatkan ketepatan klasifikasi.



**Gambar 4.** Visualisasi Feature Importance Model Random Forest

Berdasarkan Gambar 4, hasil *feature importance* menunjukkan bahwa *EmpEnvironmentSatisfaction* dan *EmpLastSalaryHikePercent* merupakan atribut dengan kontribusi relatif paling besar dalam klasifikasi kinerja pegawai, yang mengindikasikan bahwa kepuasan lingkungan kerja dan sistem penghargaan berperan dominan dalam membedakan tingkat kinerja. Temuan ini selaras dengan teori manajemen sumber daya manusia yang menempatkan motivasi, kepuasan kerja, dan penghargaan sebagai faktor kunci dalam peningkatan kinerja. Selain itu, atribut *YearsSinceLastPromotion* dan fitur pengalaman kerja juga memberikan kontribusi yang cukup signifikan, menunjukkan pentingnya pengembangan karier dan akumulasi pengalaman dalam evaluasi kinerja. Sebaliknya, atribut struktural dan demografis seperti departemen, jabatan, usia, dan masa kerja memiliki pengaruh relatif lebih kecil, sehingga berperan sebagai faktor pendukung. Dengan demikian, hasil ini menegaskan bahwa model Random Forest dalam penelitian ini lebih menekankan faktor psikologis dan pengembangan karier dibandingkan faktor administratif, sesuai dengan pola yang terbentuk pada dataset yang digunakan.

## SIMPULAN

Hasil penelitian menunjukkan bahwa model Random Forest mampu mengklasifikasikan kinerja pegawai secara andal dengan performa yang stabil, ditunjukkan oleh akurasi tertinggi sebesar 94,58% serta keseimbangan nilai precision, recall, dan f1-score pada pembagian data 80%:20%. Capaian ini menegaskan kemampuan algoritma ensemble dalam menangani kompleksitas serta ketidakseimbangan distribusi kelas pada data kinerja pegawai. Selain itu, analisis *feature importance* mengungkap bahwa kepuasan lingkungan kerja, sistem penghargaan, pengembangan karier, dan pengalaman kerja merupakan faktor dominan dalam membedakan tingkat kinerja, sehingga menegaskan pentingnya faktor psikologis dan pengembangan individu dalam evaluasi kinerja berbasis data.

Secara ilmiah, penelitian ini memberikan kontribusi metodologis melalui pengembangan kerangka evaluasi kinerja pegawai berbasis machine learning yang tidak hanya menekankan performa klasifikasi, tetapi juga interpretabilitas model. Secara praktis, temuan ini dapat mendukung pengambilan keputusan manajerial yang lebih objektif dan terukur, khususnya dalam mengidentifikasi kebutuhan intervensi dan pengembangan pegawai. Adapun keterbatasan penelitian ini meliputi ketidakseimbangan jumlah data antar kelas, tidak digunakannya algoritma pembandingan, serta belum dilakukannya optimasi hiperparameter. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data, membandingkan beberapa algoritma klasifikasi, serta melakukan tuning hiperparameter guna meningkatkan kinerja dan generalisasi model.

## DAFTAR PUSTAKA

- [1] F. Halida, "Penilaian Kinerja Karyawan PT . Muda Kreatif Menggunakan Metode Simple Additive Weighting ( SAW ) Berbasis Web," *Infotech J. Teknol. dan Inov. Digit.*, vol. 1, no. 02, pp. 49–57, 2025, doi: 10.65369/ys711h91.
- [2] A. Sistem, I. Manajemen, K. Dosen, D. Metode, and S. A. Weighting, "Analisis Sistem Informasi Manajemen Penilaian Kinerja Dosen dengan Metode Simple Additive Weighting," *Nobel Manag. Rev.*, vol. 6, no. 4, pp. 374–382, 2025, doi: 10.37476/nmar.v6i4.5543.
- [3] S. Ismail, M. Iqbal, and R. F. Wijaya, "Analisis Pola Peningkatan Indeks Profesionalitas ASN Kabupaten Asahan Menggunakan Random Forest Dan Gradient Boosting Machines," *Jatilima J. Multimed. Dan Teknol. Inf.*, vol. 07, no. 02, pp. 141–149, 2025, doi: 10.54209/jatilima.v7i02.1504.
- [4] A. A. Prayoga, M. Hasanuddin, S. Khodijah, and C. A. Rizki, "Analisis Penerapan Machine Learning dalam Sistem Prediksi dan Pengambilan Keputusan," *J. Electr. Eng. Res.*, vol. 1, no. 3, pp. 84–90, 2025.
- [5] I. Meiriska, A. S. Wati, and L. Rahmi, "Pemanfaatan Algoritma Decision Tree Pada Machine Learning Dalam Penentuan Klasifikasi Kinerja Karyawan Pada CV Duta Media," *Jupiter (Jurnal Penelit. Ilmu dan Teknol. Komputer)*, vol. 17, no. 1, pp. 169–178, 2025.
- [6] M. A. Tenriadjeng, I. As'ad, and A. Faradibah, "Implementasi Naïve Bayes untuk Evaluasi dan Klasifikasi Beban Kerja Pegawai di Badan Kepegawaian Kabupaten Barru," *Linier Lit. Inform. dan Komput.*, vol. 2, no. 3, pp. 286–293, 2025, doi: 10.33096/linier.v2i3.3140.
- [7] R. Sani, S. Defit, and G. W. Nurcahyo, "Metode Rough Set dan SVM untuk Penilaian Kinerja Kepegawaian serta Rekomendasi Jabatan," *J. KomtekInfo*, vol. 11, no. 4, pp. 342–353, 2024, doi: 10.35134/komtekinf.v11i4.581.
- [8] F. Nuraeni, D. Kurniadi, and M. H. Diazki, "Algoritma K-Nearest Neighbor Pada Kasus Dataset Imbalanced Untuk Klasifikasi Kinerja Karyawan Perusahaan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 557–568, 2024, doi: 10.25126/jtiik.938144.
- [9] S. Sufina and L. Wati, "Implementation of Random Forest Algorithm in Lecturer Performance Evaluation System," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 2, pp. 790–798, 2025, doi: 10.30811/jaise.v5i2.6975.
- [10] S. Sitohang and V. Karnadi, "Prediksi Harga Saham Menggunakan Generalize Fuzzy

- Inference System (GENFIS3)," in *Seminar Nasional Ilmu Sosial dan Teknologi (SAINTEK)*, 2020, no. Februari, pp. 132–144.
- [11] I. A. Thaher, A. Septiariani, and N. Puspitasari, "Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means," *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 131–141, 2022, doi: 10.34010/komputika.v11i2.5518.
- [12] S. Rukmini, N. Nurchim, and D. Hartanti, "Pendekatan Naive Bayes dalam Klasifikasi Penilaian Kinerja Pegawai," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 2811–2818, 2025, doi: 10.36040/jati.v9i2.13124.
- [13] M. Erkamim, S. Suswandi, M. Z. Subarkah, and E. Widarti, "Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM," *J. Sist. Inf. Bisnis*, vol. 02, pp. 127–134, 2023, doi: 10.21456/vol13iss2pp127-134.
- [14] P. Alkhairi and Z. Situmorang, "Penerapan Data Mining Untuk Menganalisis Kepuasan Pegawai Terhadap Pelayanan Bidang SDM dengan Algoritma C4.5," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 7, no. 1, pp. 40–50, 2022.
- [15] B. M. Dinata and E. Ardianto, "Pemodelan Hybrid Support Vector Machine dan K-Nearest Neighbors Untuk Peramalan Permintaan Suku Cadang Otomotif: Pendekatan Machine Learning Terhadap Data Fluktuatif," *J. Educ. Dev. Inst. Pendidik. Tapanuli Selatan*, vol. 14, no. 1, pp. 178–188, 2026, doi: 10.37081/ed.v14i1.7642.
- [16] R. S. Siburian, O. Rozalia, P. Alpianita, and A. A. Dermawan, "Klasifikasi Ketidakhadiran Karyawan Menggunakan Metode Algoritma Decision Tree C4.5," *JT-IBSI (Jurnal Tek. Ibnu Sina)*, vol. 9, no. 01, pp. 61–73, 2024, doi: 10.36352/jt-ibsi.v9i01.887.
- [17] S. Sitohang and R. P. Tampubolon, "Pengaruh Lingkungan Kerja Terhadap Kinerja Pegawai Dinas Pemadam Kebakaran Kota Batam," *J. Mirai Manag. J. Mirai Manag.*, vol. 4, no. 2, pp. 233–242, 2019, doi: 10.37531/mirai.v4i2.504.
- [18] A. Pratama, S. Mujiyono, and U. P. Sanjaya, "Analisis Klasifikasi Resign Karyawan dengan Random Forest," *J. Unitek*, vol. 18, no. 1, pp. 2580–2582, 2025, doi: 10.52072/unitek.v18i1.1415.
- [19] S. Darma, Y. Yusman, and J. Hendrawan, "Analisis Data Tingkat Kehadiran Pegawai dengan Menggunakan Clustering K-Means Pada Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Langkat," *J. Minfo Polgan*, vol. 13, no. 1, pp. 1106–1116, 2024, doi: 10.33395/jmp.v13i1.13958.
- [20] M. A. Dinata, H. Yenni, W. Agustin, and A. Aguston, "Prediksi Jumlah Titik Ruang Terbuka Hijau (RTH) Menggunakan Regresi Linier Dan Model Random Forest," *Betrik Basemah Teknol. Inf. dan Komput.*, vol. 16, no. 02, pp. 144–156, 2025.
- [21] D. Kiswanto, F. Ramadhani, N. M. Surbakti, and N. A. Nasution, "Pengembangan dan Implementasi Sistem Deteksi Serangan DDoS Berbasis Algoritma Random Forest," *Bull. Inf. Technol.*, vol. 6, no. 3, pp. 247–256, 2025, doi: 10.47065/bit.v6i3.2203.
- [22] S. A. Putri and R. Rachmatika, "Penerapan Algoritma Random Forest dan SMOTE untuk Prediksi Risiko Putus Sekolah Siswa Sekolah Menengah Kejuruan," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 3, pp. 903–910, 2025, doi: 10.51454/decode.v5i3.1360.
- [23] M. Z. Abidin, M. A. Rosid, and A. Eviyanti, "Analisis Sentimen Pengguna Aplikasi Shopee Pada Google Play Store Menggunakan Metode Random Forest," *STORAGE – J. Ilm. Tek. dan Ilmu Komput.*, vol. 4, no. 3, pp. 265–276, 2025, doi: 10.55123/storage.v4i3.6027.
- [24] A. S. Lather, R. Malhotra, P. Saloni, P. Singh, and S. Mittal, "Prediction of Employee Performance using Machine Learning Techniques," *Proc. 1st Int. Conf. Adv. Inf. Sci.*

- Syst.*, no. November, 2019, doi: 10.1145/3373477.3373696.
- [25] T. Tugino, "Analisis Data Mining Menggunakan Teknik Single Linkage Untuk Klasterisasi Key Performance (Studi Kasus Di Stmik Triguna Dharma Medan)," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 3, no. 2, pp. 114–122, 2020.
- [26] P. Simanjuntak, C. E. Suharyanto, S. Sitohang, and K. Handoko, "Data Mining Untuk Klasifikasi Status Pandemi Covid 19," *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 327, 2022, doi: 10.37600/tekinkom.v5i2.620.
- [27] D. Mayangsari and T. Wahyudi, "Classification of New Employee Selection Using the C4.5," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. January, pp. 26–34, 2025, doi: 10.57152/malcom.v5i1.1525.