

Model Evaluasi Kinerja Karyawan Berbasis Rekam Jejak Digital Menggunakan *Machine Learning*

Employee Performance Evaluation Model Based on Digital Footprints Using Machine Learning

Yan Yang Thanri^{a,1}, Juli Iriani^{b,2}, Angel Gowasa^{c,3}, Luthfi Zaidi^{d,4}

^{a,b,c,d}Universitas Potensi Utama, Jl. K.L Yos Sudarso, Km. 6,5, No. 3A, Medan, 20241, Indonesia
ythanri@gmail.com¹, juliiriani179@gmail.com²

ABSTRAK

Evaluasi kinerja karyawan merupakan aspek penting dalam manajemen sumber daya manusia, namun metode tradisional sering subjektif dan kurang efektif. Penelitian ini mengembangkan model evaluasi kinerja karyawan berbasis rekam jejak digital menggunakan algoritma machine learning untuk menghasilkan penilaian yang lebih objektif dan adil. Data dikumpulkan dari 2.435 karyawan, meliputi 16 variabel input terkait aktivitas kerja digital, data wearable & kesehatan, persepsi dan kepuasan kerja, kompetensi, serta informasi organisasi, dengan dua variabel output: *Performance_Score* dan *Performance_Label*. Tahapan penelitian mencakup seleksi fitur dengan ANOVA F-test, preprocessing data (imputasi missing value dan normalisasi dengan RobustScaler), penyeimbangan kelas menggunakan SMOTE, pembagian data latih (2.688 sampel) dan uji (672 sampel), serta pelatihan model dengan Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, KNN, SVM, dan Naïve Bayes. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, F1-score, serta confusion matrix. Hasil menunjukkan Random Forest memiliki performa terbaik dengan accuracy 0,872 dan F1-score 0,871, diikuti Gradient Boosting (accuracy 0,830, F1-score 0,827) dan Decision Tree (accuracy 0,826, F1-score 0,822). KNN dan SVM menunjukkan performa menengah, sedangkan Logistic Regression dan Naïve Bayes memiliki hasil paling rendah. Fitur paling berpengaruh adalah Organization Support Score, Tech Skills Score, Physical Activity Steps, Daily Working Hours, dan Overtime Hours Monthly. Model berbasis machine learning, khususnya Random Forest, efektif untuk evaluasi kinerja karyawan berbasis rekam jejak digital, memungkinkan penilaian yang lebih objektif, adil, dan akurat.

Kata Kunci : Evaluasi Kinerja Karyawan, Rekam Jejak Digital, Machine Learning, Random Forest, SMOTE, Feature Selection, Prediksi Kinerja, Data HRM

ABSTRACT

Employee performance evaluation is a crucial aspect of human resource management, yet traditional methods are often subjective and less effective. This study develops an employee performance evaluation model based on digital footprint using machine learning algorithms to provide a more objective and fair assessment. Data were collected from 2,435 employees, including 16 input variables related to digital work activities, wearable and health data, work perception and satisfaction, competencies, and organizational information, with two output variables: *Performance_Score* and *Performance_Label*. The research stages included feature selection using ANOVA F-test, data preprocessing (missing value imputation and normalization with RobustScaler), class balancing using SMOTE, splitting the dataset into training (2,688 samples) and testing (672 samples) sets, and training models with Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, KNN, SVM, and Naïve Bayes. Model evaluation was performed using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results show that Random Forest achieved the best performance with an accuracy of 0.872 and F1-score of 0.871, followed by Gradient Boosting (accuracy 0.830, F1-score 0.827) and Decision Tree (accuracy 0.826, F1-score 0.822). KNN and SVM exhibited moderate performance, while Logistic Regression and Naïve Bayes had the lowest results. The most influential features were Organization Support Score, Tech Skills Score, Physical Activity Steps, Daily Working Hours, and Overtime Hours Monthly. The machine learning-based model, particularly Random Forest, is effective for employee performance evaluation based on digital footprints, enabling more objective, fair, and accurate assessments.

Keywords : Employee Performance Evaluation, Digital Footprint, Machine Learning, Random Forest, SMOTE, Feature Selection, Performance Prediction, HRM Data

Info Artikel :

Disubmit:

Direview:

Diterima :

Copyright © 2025 – CSRID Journal. All rights reserved.

1. PENDAHULUAN

Dalam dunia bisnis yang kompetitif dan dinamis, evaluasi kinerja karyawan merupakan elemen kunci dalam manajemen sumber daya manusia (HRM)[1][2]. Evaluasi ini berperan penting dalam mengidentifikasi, mengembangkan, dan mempertahankan karyawan berkinerja tinggi serta memberikan umpan balik konstruktif bagi yang membutuhkan perbaikan [3]. Namun, proses evaluasi seringkali tidak sistematis dan menghadapi berbagai tantangan, seperti subjektivitas, ketidakjelasan kriteria penilaian, serta kurangnya data yang akurat [4]. Akibatnya, metode evaluasi menjadi kurang efektif. Secara tradisional, evaluasi dilakukan secara manual dan subjektif, sehingga rentan terhadap bias[5]. Rekam jejak karyawan dalam konteks HRM merujuk pada dokumentasi kinerja selama periode tertentu, meliputi pencapaian target, proyek yang diselesaikan, kontribusi terhadap tim, serta aspek kehadiran dan kedisiplinan[6]. Evaluasi yang berbasis rekam jejak sangat terkait dengan efektivitas dan objektivitas proses penilaian[7]. Penggunaan rekam jejak memungkinkan evaluasi menjadi lebih terstruktur, objektif, dan berbasis data konkret. Informasi terdokumentasi seperti pencapaian target, kontribusi proyek, dan umpan balik dari rekan kerja memungkinkan penilaian yang lebih holistik dan faktual [8]. Integrasi rekam jejak dalam evaluasi membantu mengatasi tantangan subjektivitas serta memperkuat pengambilan keputusan yang didasarkan pada data terukur. Kemajuan kecerdasan buatan dan machine learning membuka peluang untuk menjadikan proses evaluasi kinerja lebih objektif, efisien, akurat, dan sistematis [9]. Menurut Harvard Business Review (Antonio, 2018), AI memberikan dampak signifikan pada fungsi organisasi, khususnya dalam prediksi dan manajemen kinerja[10]. Dalam konteks HRM, AI berperan penting dalam prediksi dan pengambilan keputusan, seperti memprediksi tingkat pergantian karyawan, kinerja masa depan, dan kepuasan kerja [11][12]. Selain itu, faktor eksternal yang memengaruhi kehidupan karyawan turut berkontribusi terhadap kinerja dan perlu dipertimbangkan. Sayangnya, banyak penelitian masih mengabaikan faktor dinamis ini, yang menimbulkan kekhawatiran terkait keadilan dalam penilaian [13][14]. Rodgers dkk. (2023) mengembangkan kerangka kerja yang mempertimbangkan faktor sosial, ekonomi, dan fisik dalam pengambilan keputusan HRM. Namun, pendekatan manual menjadi kurang praktis mengingat banyaknya aturan dan preferensi yang beragam dalam penilaian[15]. Kerangka kerja berbasis machine learning memungkinkan organisasi melakukan evaluasi kinerja secara cepat dan bebas bias [12][16][17][18]. Kerangka ini dapat memprediksi dan mengambil keputusan berdasarkan faktor fisik, sosial/perilaku, ekonomi, organisasi, dan teknologi.

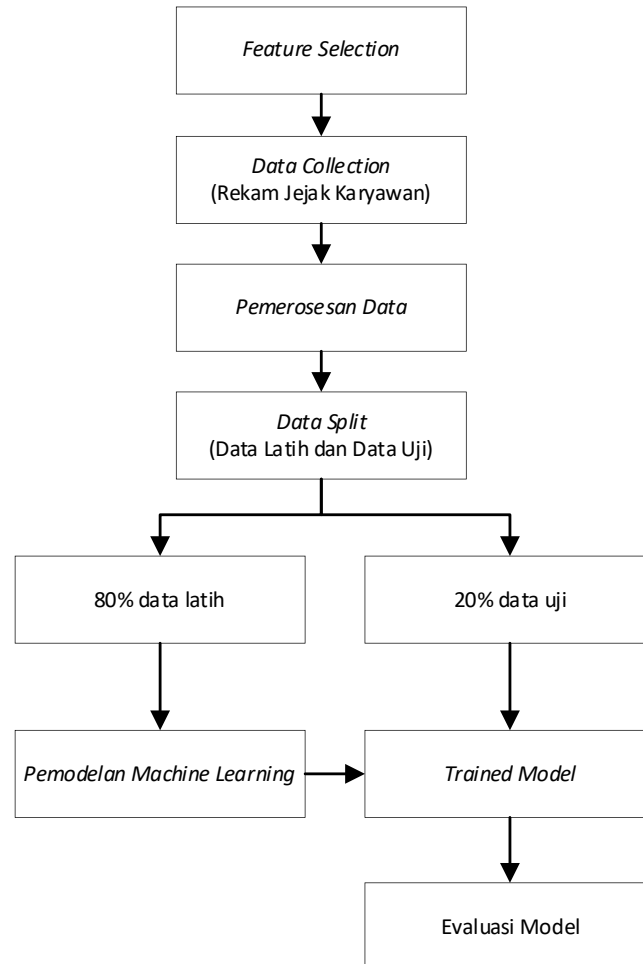
Berdasarkan perkembangan tersebut, penelitian ini menempati posisi penting dengan menawarkan model prediksi kinerja karyawan berbasis rekam jejak digital [19][20][21][22] yang secara komprehensif mengintegrasikan berbagai faktor fisik, sosial, ekonomi, organisasi, dan teknologi. Pendekatan ini tidak hanya mengikuti arah penelitian sebelumnya, tetapi juga menutup celah (gap) yang masih ada, yaitu kebutuhan akan model evaluasi yang lebih objektif, bebas bias, dan mempertimbangkan konteks lingkungan karyawan secara holistik. Model dikembangkan menggunakan beberapa algoritma machine learning seperti Logistic Regression, Gaussian Naive Bayes, Decision Tree, K-Nearest Neighbors, dan Support Vector Machine untuk mendapatkan prediksi yang paling akurat. Selain itu, keterlibatan berbagai stakeholder dari organisasi ‘for-profit’ dalam pengumpulan data memperkuat representativitas model yang dihasilkan.

Dengan demikian, state-of-the-art penelitian ini berada pada integrasi rekam jejak digital dan faktor lingkungan multidimensi dengan penerapan machine learning untuk menghasilkan evaluasi kinerja yang objektif dan komprehensif. Kebaruan penelitian terletak pada pendekatan holistik yang menggabungkan berbagai variabel yang selama ini kurang diperhatikan dalam penelitian terdahulu. Kontribusi penelitian ini diharapkan dapat memberikan solusi evaluasi kinerja yang lebih adil, akurat, dan mendukung pengambilan keputusan strategis dalam organisasi.

2. METODE PENELITIAN

A. Tahapan Penelitian

Tahapan penelitian model evaluasi kinerja karyawan berbasis rekam jejak menggunakan machine learning dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan awal dalam pengembangan model ini adalah mengidentifikasi variabel-variabel penting yang dapat memengaruhi kinerja karyawan berdasarkan rekam jejak digital. Variabel tersebut meliputi faktor-faktor seperti riwayat kehadiran, performa tugas, interaksi sistem (misalnya penggunaan aplikasi kerja atau platform kolaboratif), serta aspek sosial dan ekonomi lainnya. Feature selection bertujuan untuk memilih atribut yang paling relevan dan berpengaruh dalam proses evaluasi kinerja. Data Collection adalah data mengenai variabel-variabel tersebut dikumpulkan dari berbagai sumber, baik dari catatan internal perusahaan maupun dari kuesioner yang diberikan kepada karyawan. Pemrosesan Data, dimana data yang terkumpul kemudian diproses untuk membersihkan, mengubah, dan menyiapkan data agar siap digunakan dalam analisis. Langkah ini melibatkan pemilihan fitur (feature selection) untuk menentukan variabel yang paling berpengaruh. Pada tahap pemodelan Machine Learning, data yang telah melalui proses pra-pengolahan digunakan untuk melatih sejumlah algoritma machine learning, di antaranya Gradient Boosting, Gaussian Naive Bayes, K-Nearest Neighbors, Decision Tree, SVM, Logistic Regression, dan Random Forest. Selanjutnya dilakukan evaluasi model guna mengukur

performa masing-masing algoritma menggunakan metrik evaluasi yang relevan, seperti akurasi, presisi, recall, dan F1-score, untuk menentukan model dengan hasil terbaik.

B. Rekam Jejak Digital

Dalam penelitian ini, rekam jejak digital diartikan sebagai kumpulan data mengenai aktivitas karyawan yang tercatat secara otomatis melalui berbagai sistem digital yang digunakan dalam organisasi. Data ini mencerminkan pola kerja, kondisi fisik dan psikologis, serta faktor-faktor organisasi yang memengaruhi kinerja individu. Komponen-komponen rekam jejak digital yang dianalisis dalam penelitian ini disajikan pada tabel 1.

Tabel 1. Komponen Rekam Jejak Digital untuk Penilaian Kinerja Karyawan

Kategori Data	Fitur
Aktivitas Kerja Digital	Daily_Working_Hours, Screen_Time, Meetings_Attended, Emails_Sent
Data Wearable & Kesehatan	Physical_Activity_Steps, Sleep_Duration, Stress_Level
Persepsi & Kepuasan	Work_Life_Balance_Satisfaction, Organization_Support_Score
Kemampuan & Karier	Tech_Skills_Score, Job_Level, Tenure_Years, Trainings_Attended_YTD
Data Organisasi	Base_Salary, Incentive_Received_LastMonth, Overtime_Hours_Monthly
Output Kinerja	Performance_Score, Performance_Label (Low, Average, Good, Excellent)

Dataset rekam jejak digital menyajikan 16 variabel input dan 2 variabel output yang digunakan untuk menilai kinerja karyawan berdasarkan rekam jejak digital sebagaimana tertuang pada tabel 1. Data dikumpulkan dari 2.345 karyawan dan berasal dari berbagai sumber, termasuk aktivitas digital karyawan serta sistem informasi organisasi. Variabel-variabel tersebut kemudian dikelompokkan ke dalam enam kategori utama: aktivitas kerja digital, data wearable dan kesehatan, persepsi serta kepuasan, kompetensi dan pengembangan karier, informasi organisasi, serta hasil kinerja.

Fitur-fitur ini mencakup aspek-aspek penting seperti produktivitas, kesejahteraan, persepsi terhadap pekerjaan, kemampuan, kompensasi, dan pencapaian kinerja aktual. Dengan menggabungkan data dari berbagai dimensi yang memengaruhi kinerja, pendekatan ini memungkinkan evaluasi yang lebih menyeluruh, objektif, dan mengurangi potensi bias.

Selain itu, distribusi data Performance_Label menunjukkan ketidakseimbangan antar kelas, sebagaimana dirangkum pada tabel 2.

Tabel 2. Distribusi data

Performance_Label	Kategori	Jumlah Data	Persentase (%)
0	Average	840	34,5
2	Good	628	25,8
1	Excellent	511	21
3	Low	456	18,7
Total	-	2.435	100

Tabel 2 menunjukkan bahwa kelas Average memiliki proporsi data tertinggi, sementara kelas Low memiliki jumlah data paling sedikit, sehingga perlu diperhatikan teknik penyeimbangan data (SMOTE) sebelum digunakan pada model prediksi kinerja.

3. HASIL DAN PEMBAHASAN

A. Feature Selection

Untuk meningkatkan performa model prediksi kinerja karyawan, dilakukan seleksi fitur dari 16 variabel input. Variabel target Performance_Label dikodekan secara numerik, sedangkan variabel kategorikal diubah menjadi representasi numerik melalui one-hot encoding. Seleksi fitur dilakukan menggunakan univariate feature selection dengan ANOVA F-test, menghasilkan 12 fitur terbaik yang paling berpengaruh terhadap klasifikasi kinerja terlihat pada gambar 2. Fitur-fitur ini digunakan sebagai input utama model machine learning, sehingga meningkatkan akurasi prediksi sekaligus menyederhanakan kompleksitas data.



Gambar 2. Hasil seleksi 12 fitur terbaik berdasarkan ANOVA F-test

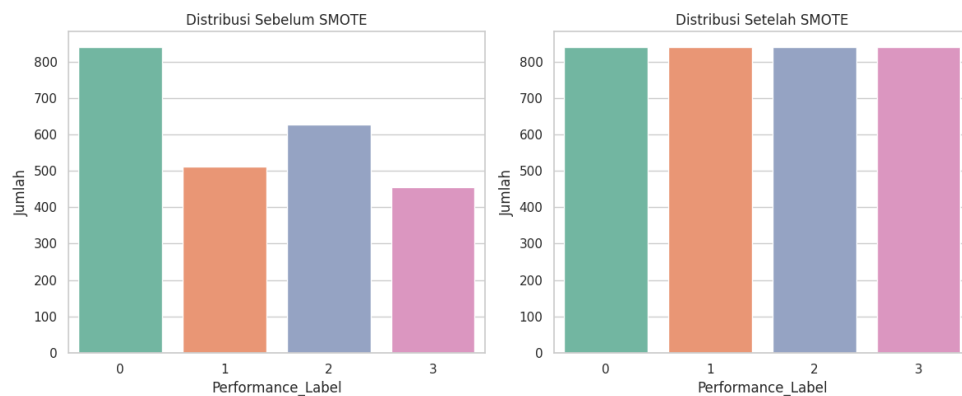
Hasil seleksi fitur pada gambar 2 menggunakan metode univariate feature selection dengan ANOVA F-test, yang menghasilkan 12 variabel dengan kontribusi paling signifikan terhadap klasifikasi kinerja karyawan. Dari 16 variabel input awal, fitur Organization Support Score dan Tech Skills Score menempati urutan tertinggi dengan skor ANOVA terbesar, diikuti oleh Physical_Activity_Steps, Daily Working_Hours, dan Overtime_Hours_Monthly. Sementara itu, fitur seperti Emails Sent, Screen Time, dan Base Salary menunjukkan pengaruh yang relatif lebih rendah.

Hasil ini menegaskan bahwa aspek dukungan organisasi, keterampilan teknis, aktivitas fisik, serta beban kerja harian merupakan indikator penting dalam menentukan performa karyawan. Dengan demikian, fitur-fitur terpilih tersebut digunakan sebagai input utama dalam pemodelan machine learning, sehingga model yang dibangun dapat lebih akurat sekaligus efisien.

B. Distribusi Variabel Target

Berdasarkan Tabel 3, variabel target *Performance_Label* terdiri dari empat kategori, yaitu *Average*, *Good*, *Excellent*, dan *Low*, dengan total 2.435 data karyawan. Distribusi awal menunjukkan adanya ketidakseimbangan antar kelas. Kategori *Average* mendominasi dengan 840 data (34,5%), diikuti oleh *Good* sebanyak 628 data (25,8%), *Excellent* sebanyak 511 data (21%), serta *Low* yang hanya memiliki 456 data (18,7%)

Ketidakseimbangan distribusi ini berpotensi menyebabkan model prediksi lebih condong dalam mengenali kelas mayoritas (*Average* dan *Good*), sementara kelas minoritas (*Low*) kurang terwakili. Oleh karena itu, pada tahap selanjutnya dilakukan penyeimbangan data menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) agar setiap kategori memiliki proporsi yang lebih setara, sehingga model *machine learning* dapat belajar secara lebih adil dan menghasilkan performa prediksi yang optimal.



Gambar 3. Grafik Disktribusi kelas sebelum dan sesudah dilakukan SMOTE

Grafik pada gambar 3 menunjukkan distribusi kelas *Performance_Label* sebelum dan sesudah dilakukan SMOTE memperlihatkan adanya perbedaan signifikan pada jumlah data tiap kelas. Sebelum dilakukan SMOTE, data tidak seimbang, di mana kategori *Average* memiliki jumlah terbanyak yaitu 840 data (34,5%), disusul kategori *Good* sebanyak 628 data (25,8%), kategori *Excellent* sebanyak 511 data (21%), dan kategori *Low* yang paling sedikit yaitu 456 data (18,7%). Ketidakseimbangan ini dapat memengaruhi kinerja model karena algoritma cenderung lebih fokus pada kelas mayoritas. Setelah proses SMOTE, jumlah data tiap kelas berhasil diseimbangkan, masing-masing kelas memiliki 840 data. Dengan demikian, distribusi yang merata ini membuat model pembelajaran mesin mampu belajar lebih adil terhadap seluruh kelas, sehingga meningkatkan kemampuan prediksi, khususnya pada kelas yang sebelumnya minoritas. Tabel 3 memperjelas perubahan distribusi jumlah data pada setiap kelas sebelum dan sesudah dilakukan SMOTE.

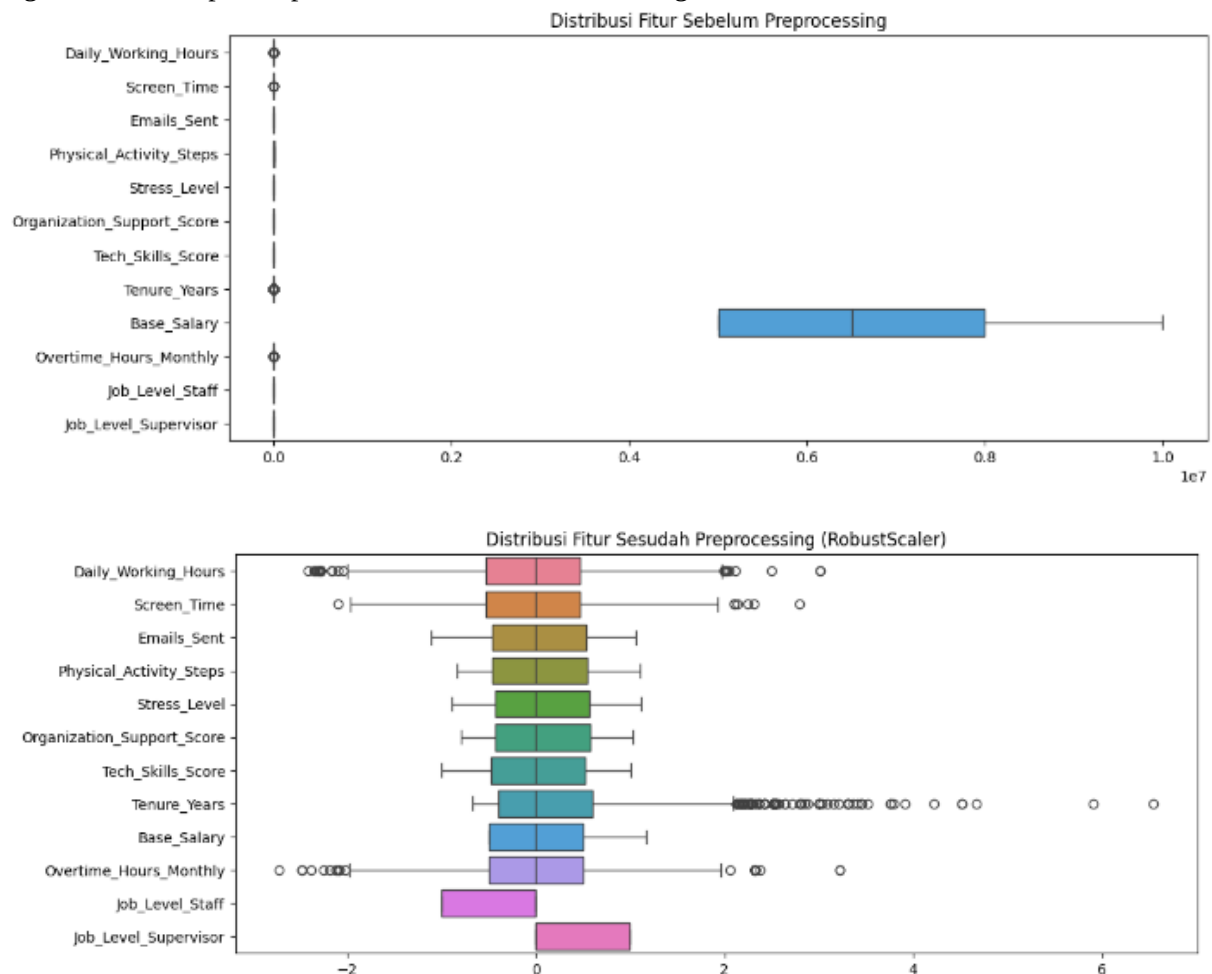
Tabel 3. Distribusi jumlah data pada setiap kelas sebelum dan sesudah dilakukan SMOTE

Kelas	Jumlah Awal	Jumlah Sesudah SMOTE
Excellent (1)	511	840
Average (0)	840	840
Good (2)	628	840
Low (3)	456	840

Tabel 3 menunjukkan bahwa sebelum SMOTE terdapat perbedaan cukup besar antar kelas, dengan selisih tertinggi antara kelas Average (840) dan Low (456) mencapai 384 data. Setelah SMOTE, seluruh kelas memiliki jumlah yang sama yaitu 840 data, sehingga distribusi menjadi seimbang. Perubahan ini memastikan tidak ada kelas yang mendominasi data latih, sehingga meningkatkan keadilan (fairness) model dalam mempelajari pola pada setiap kelas.

C. Preprocessing Data

Pada tahap Preprocessing Data, dilakukan serangkaian langkah untuk memastikan kualitas data sebelum digunakan dalam pemodelan machine learning. Pertama, proses imputasi missing value dilakukan menggunakan metode KNNImputer dengan lima tetangga terdekat, di mana nilai yang hilang diprediksi berdasarkan kemiripan data dengan sampel lain. Hal ini bertujuan agar dataset tetap lengkap dan tidak kehilangan informasi penting akibat pembuangan data yang hilang. Selanjutnya, data hasil imputasi dinormalisasi menggunakan RobustScaler, sebuah metode normalisasi yang tahan terhadap outlier karena memanfaatkan median dan interquartile range (IQR) sebagai dasar transformasi. Dengan demikian, skala data antar fitur menjadi lebih seimbang dan tidak bias akibat adanya nilai ekstrim. Hasil dari tahap preprocessing pada gambar 4 adalah dataset yang lebih bersih, terstandarisasi, dan siap digunakan dalam proses pelatihan model machine learning.



Gambar 4. Grfaik Distribusi fitur sebelum dan sesudah preprocessing

Grafik distribusi fitur sebelum dan sesudah preprocessing pada gambar 4 menunjukkan perbedaan yang signifikan dalam skala dan penyebaran data. Pada grafik sebelum preprocessing, terlihat bahwa sebagian besar fitur memiliki skala yang relatif kecil, sementara fitur *Base_Salary* mendominasi dengan nilai yang jauh lebih besar dibandingkan fitur lainnya. Hal ini menyebabkan ketidakseimbangan distribusi, sehingga fitur dengan skala kecil tidak terlihat jelas perbedaannya dan berpotensi memengaruhi kinerja model pembelajaran mesin. Selain itu, beberapa fitur juga menunjukkan adanya outlier yang cukup ekstrem.

Setelah dilakukan preprocessing menggunakan *RobustScaler*, grafik distribusi menunjukkan perubahan yang lebih seimbang. Semua fitur berada pada skala yang serupa, dengan median yang cenderung terpusat di sekitar nol. *RobustScaler* juga membantu mengurangi pengaruh outlier, meskipun masih terlihat beberapa titik ekstrem pada fitur tertentu seperti *Tenure_Years* dan *Overtime_Hours_Monthly*. Namun, secara keseluruhan distribusi menjadi lebih proporsional sehingga setiap fitur memiliki kontribusi yang lebih adil dalam proses pembelajaran model. Dengan demikian, preprocessing berhasil mengatasi masalah perbedaan skala antar fitur dan membuat dataset lebih siap untuk digunakan pada tahap pemodelan.

D. Split Data

Pada tahap split data, proses pembagian dataset dilakukan menggunakan fungsi *train_test_split* dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data yang digunakan merupakan hasil preprocessing, yaitu setelah melalui proses imputasi, normalisasi, serta penyeimbangan kelas menggunakan *SMOTE*. Agar distribusi kelas tetap seimbang antara data latih dan data uji, digunakan parameter *stratify=y_resampled*, sementara *random_state=42* ditetapkan untuk memastikan hasil pembagian konsisten. Hasil pembagian menunjukkan bahwa data latih terdiri dari 2.688 sampel dengan 12 fitur, sedangkan data uji terdiri dari 672 sampel dengan 12 fitur. Pembagian ini bertujuan agar model dapat mempelajari pola dari sebagian besar data (training set), kemudian diuji pada data yang belum pernah dilihat sebelumnya (testing set) untuk mengukur kinerja model secara objektif.

E. Pemodelan dengan Machine Learning

Pada tahap pemodelan, digunakan beberapa algoritma *machine learning* yang berbeda untuk membandingkan performa dalam mengklasifikasikan kinerja karyawan berdasarkan fitur-fitur terpilih. Pemilihan beragam algoritma ini bertujuan agar diperoleh gambaran menyeluruh mengenai model mana yang paling sesuai dengan karakteristik data. Secara umum, algoritma yang digunakan terdiri atas model linear, pohon keputusan, metode ensemble, serta algoritma berbasis probabilistik dan jarak. Rincian algoritma yang digunakan dapat dilihat pada tabel 4 berikut.

Tabel 4. Daftar algoritma machine learning

No	Algoritma	Parameter Utama	Keterangan
1	Logistic Regression	max_iter=1000, random_state=42	Digunakan sebagai model baseline linear
2	Decision Tree	random_state=42	Model pohon keputusan, mudah diinterpretasi
3	Random Forest	n_estimators=200, random_state=42	Ensemble dari banyak decision tree untuk meningkatkan akurasi
4	Gradient Boosting	n_estimators=200, random_state=42	Ensemble boosting yang fokus memperbaiki kesalahan model sebelumnya
5	K-Nearest Neighbors	Default parameter	Klasifikasi berbasis kemiripan tetangga terdekat
6	Support Vector Machine	probability=True, random_state=42	Membuat hyperplane optimal untuk memisahkan kelas
7	Naïve Bayes	Default parameter (GaussianNB)	Probabilistik sederhana, asumsi independensi antar fitur

Berdasarkan model machine learning pada tabel 4 menunjukkan bahwa *Logistic Regression* dipilih sebagai model baseline linear, sedangkan *Decision Tree* digunakan karena sifatnya yang mudah diinterpretasi. Model *Random Forest* dan *Gradient Boosting* mewakili pendekatan *ensemble* yang mampu meningkatkan performa prediksi dengan menggabungkan banyak pohon keputusan. *K-Nearest Neighbors (KNN)* digunakan untuk mengukur kesamaan antar data berdasarkan kedekatan jarak, sementara *Support Vector Machine (SVM)* membangun *hyperplane* optimal untuk memisahkan kelas dengan margin maksimal. Terakhir, *Naïve Bayes* mewakili model probabilistik sederhana yang efektif pada data berdimensi tinggi dengan asumsi independensi antar fitur.

Dengan kombinasi model yang beragam, mulai dari pendekatan linear, *tree-based*, *ensemble*, hingga *probabilistic*, penelitian ini memberikan ruang pembandingan yang komprehensif terhadap performa masing-masing algoritma. Hal ini penting untuk mengidentifikasi algoritma terbaik yang mampu melakukan evaluasi kinerja karyawan berbasis rekam jejak digital secara lebih akurat dan reliabel. Pendekatan ini juga memastikan bahwa aspek interpretabilitas, akurasi, serta kemampuan generalisasi dapat diperhitungkan secara seimbang dalam membangun model prediktif yang robust.

F. Evaluasi Model

Untuk menilai efektivitas algoritma dalam mengklasifikasikan kinerja karyawan, dilakukan proses evaluasi menggunakan empat metrik utama yaitu Accuracy, Precision, Recall, dan F1-score. Metrik-metrik ini dipilih karena mampu memberikan gambaran menyeluruh mengenai kemampuan model, tidak hanya dari sisi ketepatan prediksi, tetapi juga dari keseimbangan antara prediksi benar positif dan negatif. Hasil perbandingan kinerja seluruh model machine learning yang diuji dapat dilihat pada tabel 5 berikut.

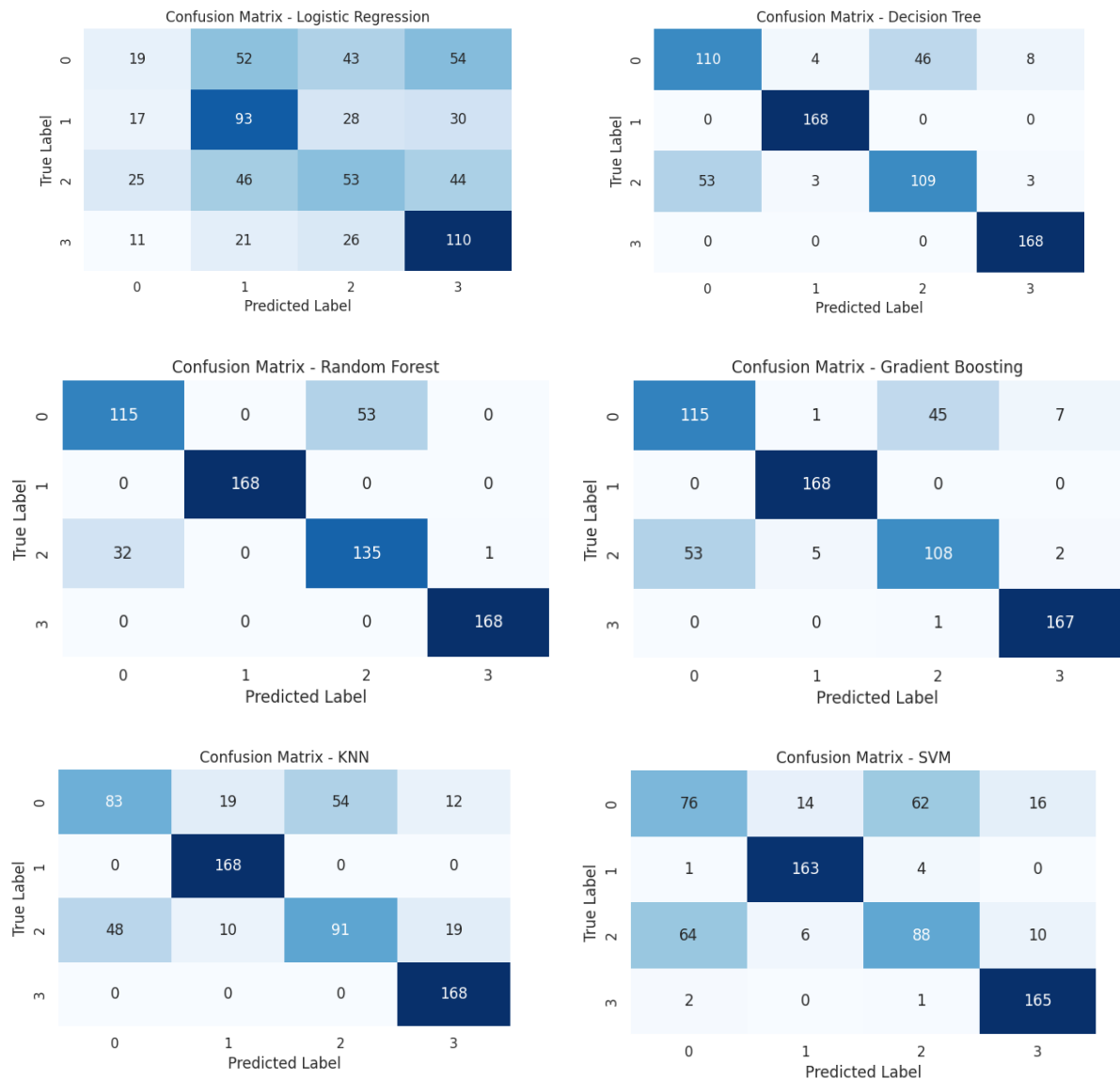
Tabel 5. Perbandingan Hasil Evaluasi Model Sebelum dan Sesudah SMOTE

Model	Tanpa SMOTE				SMOTE			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.388	0.394	0.388	0.356	0.409	0.380	0.409	0.381
Decision Tree	0.817	0.810	0.817	0.813	0.826	0.819	0.826	0.822
Random Forest	0.811	0.809	0.811	0.804	0.872	0.874	0.872	0.871
Gradient Boosting	0.782	0.777	0.782	0.778	0.830	0.825	0.830	0.827
KNN	0.667	0.641	0.667	0.646	0.759	0.740	0.759	0.743
SVM	0.671	0.661	0.671	0.655	0.732	0.713	0.732	0.720
Naïve Bayes	0.390	0.387	0.390	0.372	0.399	0.387	0.399	0.380

Hasil pada tabel 5 menunjukkan adanya peningkatan performa model setelah dilakukan penyeimbangan kelas dengan SMOTE. Secara umum, seluruh algoritma mengalami perbaikan pada nilai evaluasi, khususnya pada metrik *accuracy* dan *F1-score* yang lebih merepresentasikan keseimbangan antara *precision* dan *recall*. Peningkatan paling signifikan terlihat pada model Random Forest yang sebelumnya memiliki *accuracy* sebesar 0,811 dan meningkat menjadi 0,872 setelah SMOTE, dengan nilai *F1-score* yang juga melonjak dari 0,804 menjadi 0,871. Model Gradient Boosting dan Decision Tree juga menunjukkan peningkatan konsisten pada semua metrik.

Sementara itu, model sederhana seperti Logistic Regression dan Naïve Bayes tidak mengalami perbaikan berarti, bahkan Logistic Regression sedikit menurun pada nilai *precision*. Hal ini menunjukkan bahwa algoritma kompleks berbasis ensemble lebih mampu memanfaatkan distribusi data yang seimbang dibanding model linear atau probabilistik sederhana. Dengan demikian, dapat disimpulkan bahwa proses SMOTE berperan penting dalam meningkatkan kinerja model, terutama untuk algoritma yang mengandalkan representasi data yang lebih kaya seperti Random Forest dan Gradient Boosting.

Untuk memperoleh gambaran yang lebih detail mengenai kinerja setiap model pada masing-masing kelas, dilakukan evaluasi *menggunakan confusion matrix*. Confusion matrix memberikan informasi terkait jumlah prediksi yang benar maupun salah dari setiap kategori label kinerja karyawan, sehingga dapat diketahui kelas mana yang masih sulit diprediksi oleh model. Gambar 5 menampilkan *confusion matrix* dari seluruh algoritma yang digunakan, mulai dari Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, KNN, SVM, hingga Naïve Bayes.



Confusion Matrix - Naive Bayes

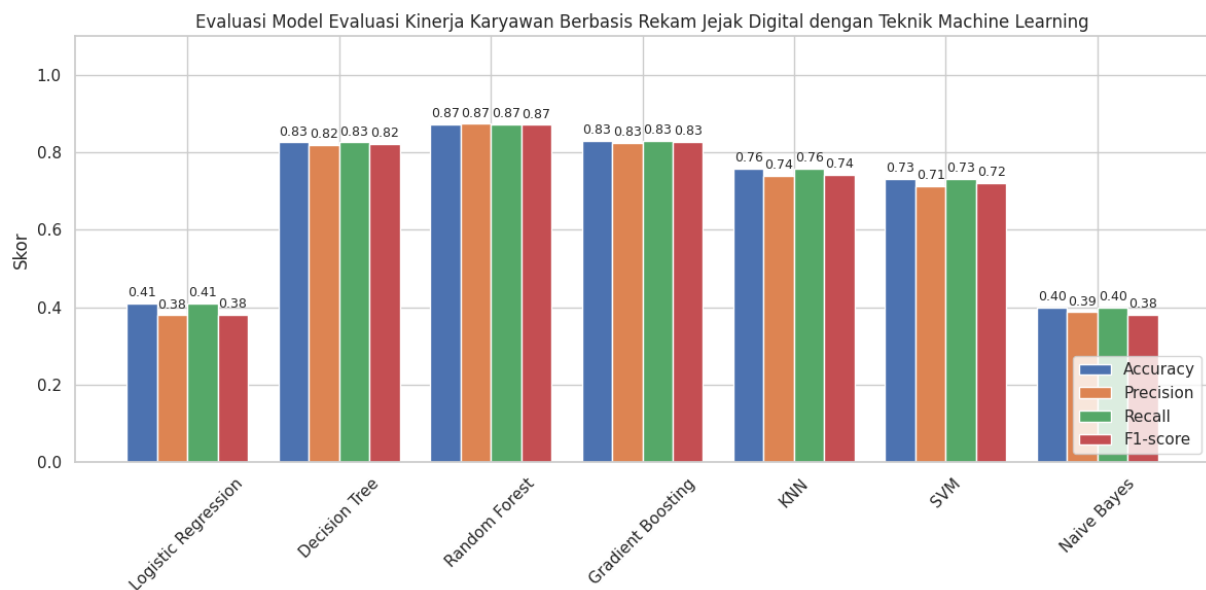
True Label \ Predicted Label	0	1	2	3
0	29	53	35	51
1	10	97	20	41
2	27	54	50	37
3	23	32	21	92

Gambar 5. Confusion Matrix

Berdasarkan confusion matrix pada setiap model, terlihat bahwa kinerja model dalam memprediksi empat kelas kinerja karyawan menunjukkan variasi yang cukup jelas. Model linear seperti Logistic Regression menghasilkan tingkat kesalahan tertinggi (34,5%–88,7%). Hal ini dapat dipahami karena Logistic Regression hanya mampu memodelkan hubungan linear antar fitur dan tidak dapat menangkap interaksi yang lebih kompleks. Dataset kinerja karyawan cenderung memiliki hubungan non-linear, misalnya kombinasi antar variabel produktivitas, durasi kerja, dan tingkat stres yang saling berinteraksi. Ketidakmampuan Logistic Regression dalam mempelajari pola non-linear ini menyebabkan tingginya misclassification, terutama pada kelas minoritas yang memiliki representasi data lebih sedikit. Model Naïve Bayes menunjukkan pola yang mirip, walaupun berbasis probabilistik. Asumsi independensi antar fitur membuat model ini tidak dapat menangani fitur yang saling berkorelasi, seperti working hours, stress level, dan organization support yang dalam data nyata biasanya berhubungan erat. Hal ini menjelaskan mengapa model tersebut kesulitan memberikan prediksi yang stabil dan akurat. Pada algoritma KNN dan SVM, kinerja berada pada tingkat menengah. KNN memiliki tingkat error tinggi pada kelas 0 dan 2, yang kemungkinan disebabkan oleh distribusi data antar kelas yang saling berdekatan (overlapping). Pada KNN, proses prediksi sangat dipengaruhi oleh densitas dan jarak antar titik; sehingga ketika kelas tidak terpisah jelas, keputusan menjadi kurang akurat. Sementara itu, SVM bekerja lebih baik karena kemampuannya memaksimalkan margin antar kelas. Namun, jika kernel tidak sesuai kompleksitas data, SVM tetap mengalami kesalahan pada kelas yang boundary-nya tidak terlalu jelas. Berbeda dengan itu, Decision Tree, Random Forest, dan Gradient Boosting menunjukkan performa jauh lebih baik. Decision Tree mampu menangkap pola non-linear serta interaksi antar fitur tanpa asumsi khusus, sehingga hasil prediksinya lebih baik dibanding model linear. Namun, Decision Tree cenderung overfitting pada dataset kecil atau kompleks, sehingga fluktuasi performa masih terlihat. Model berbasis ensemble seperti Random Forest dan Gradient Boosting memberikan hasil terbaik dan paling konsisten. Random Forest menggabungkan banyak pohon keputusan sehingga mampu menurunkan variansi dan mengurangi risiko overfitting yang terdapat pada Decision Tree tunggal. Hal ini tercermin dari tingkat error yang sangat rendah, termasuk 0% error pada kelas 1 dan kelas 3, serta error hanya 19,6% pada kelas 2. Sementara itu, Gradient Boosting memperkuat kemampuan prediksi dengan memfokuskan pembelajaran pada kesalahan dari iterasi sebelumnya, membuatnya efektif dalam membedakan kelas-kelas yang memiliki pola halus. Untuk memastikan apakah perbedaan performa antar model hanya kebetulan atau benar-benar signifikan, dilakukan Friedman Test. Hasil uji menunjukkan nilai $p < 0.05$, menandakan terdapat perbedaan performa yang signifikan. Uji lanjutan Nemenyi Post-Hoc Test menunjukkan bahwa Random Forest dan Gradient Boosting secara signifikan lebih baik dibanding Logistic Regression dan KNN. Namun, tidak terdapat perbedaan signifikan antara Random Forest dan Gradient Boosting, yang menunjukkan bahwa keduanya berada pada kelompok model dengan kinerja terbaik. Secara keseluruhan, keunggulan model ensemble ini dapat dijelaskan oleh kemampuannya menangkap hubungan kompleks antar fitur, mengurangi variansi, dan menghasilkan prediksi lebih stabil pada dataset multi-kelas dengan pola non-linear.

Dalam penelitian ini, hasil evaluasi kinerja model pembelajaran mesin ditampilkan melalui perbandingan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik tersebut tidak hanya memberikan gambaran umum mengenai tingkat ketepatan prediksi model, tetapi juga mengungkap bagaimana keseimbangan antara kemampuan model dalam mengenali kelas positif secara benar (*recall*), ketepatan prediksi positif yang dihasilkan (*precision*), serta harmonisasi keduanya yang tercermin pada nilai *F1-score*. Dengan demikian, interpretasi hasil tidak hanya bertumpu pada satu ukuran tunggal, melainkan mempertimbangkan berbagai aspek yang saling melengkapi.

Perbandingan metrik ini sangat penting karena setiap algoritma pembelajaran mesin memiliki karakteristik dan kekuatan yang berbeda. Misalnya, ada algoritma yang unggul dalam akurasi secara keseluruhan, tetapi kurang baik dalam menangani data yang tidak seimbang, sementara algoritma lain mungkin lebih seimbang dalam *precision* dan *recall* meskipun akurasi totalnya sedikit lebih rendah. Analisis yang komprehensif melalui empat metrik ini memungkinkan peneliti untuk menilai performa model secara lebih objektif, sehingga dapat memilih pendekatan yang paling sesuai dengan tujuan penelitian. Hasil perbandingan metrik evaluasi kinerja model pembelajaran mesin dalam penelitian ini dapat dilihat pada gambar 6 berikut.



Gambar 6. Perbandingan Hasil Evaluasi Kinerja Model Machine Learning

Berdasarkan gambar 6, terlihat bahwa algoritma Random Forest menunjukkan kinerja terbaik dibandingkan dengan model lainnya, dengan nilai *accuracy* dan *F1-score* yang lebih tinggi serta lebih stabil. Hal ini mengindikasikan bahwa kombinasi beberapa pohon keputusan dalam Random Forest mampu meningkatkan ketepatan prediksi sekaligus mengurangi risiko overfitting. Selanjutnya, algoritma Decision Tree dan Gradient Boosting juga menunjukkan performa yang cukup baik, meskipun masih sedikit di bawah Random Forest. Decision Tree dan Gradient Boosting unggul dalam menangani interaksi antar fitur yang lebih kompleks, meskipun Gradient Boosting memiliki sedikit penurunan pada *recall* pada kelas tertentu.

Sementara itu, algoritma KNN dan SVM cenderung berada pada tingkat performa menengah, di mana nilai *recall* dan *F1-score* menunjukkan adanya keterbatasan dalam mengenali pola data yang lebih kompleks. KNN sangat dipengaruhi oleh jarak antar titik, dan pada dataset dengan distribusi kelas yang tumpang tindih, model ini kurang efektif. Sementara itu, SVM masih menghadapi kesulitan pada kelas yang boundary-nya tidak terlalu jelas.

Adapun Logistic Regression dan Naïve Bayes memiliki hasil yang relatif rendah, yang dapat dikaitkan dengan sifat data yang mungkin tidak sepenuhnya memenuhi asumsi linearitas maupun independensi antar fitur. Confusion matrix untuk kedua model ini menunjukkan kesalahan prediksi yang sangat tinggi pada kelas-kelas yang lebih sulit dikenali, terutama pada kelas minoritas.

Dengan demikian, analisis grafik ini memperlihatkan bahwa model ensemble, khususnya Random Forest, lebih andal dalam mengklasifikasikan kinerja karyawan berbasis rekam jejak digital dibandingkan model lainnya.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma machine learning mampu melakukan klasifikasi kinerja karyawan berbasis rekam jejak digital dengan hasil yang bervariasi pada setiap model. Algoritma Random Forest terbukti memberikan performa terbaik dengan nilai evaluasi tertinggi pada accuracy, precision, recall, dan F1-score, sementara Decision Tree dan Gradient Boosting juga menunjukkan hasil yang kompetitif meskipun sedikit lebih rendah. Sebaliknya, algoritma KNN dan SVM menghasilkan performa sedang yang mengindikasikan keterbatasan dalam menangani data kompleks, sedangkan Logistic Regression dan Naïve Bayes cenderung memberikan hasil terendah karena kurang mampu menangkap pola non-linear dan dimensi data yang tinggi. Kelebihan penelitian ini terletak pada penerapan SMOTE yang berhasil mengatasi ketidakseimbangan kelas sehingga model dapat belajar lebih adil terhadap seluruh kategori kinerja karyawan. Namun demikian, potensi overfitting akibat penggunaan data sintetik menjadi kelemahan yang perlu diperhatikan, sehingga diperlukan pengujian lebih lanjut menggunakan dataset nyata yang lebih besar dan bervariasi. Ke depan, penelitian serupa dapat dikembangkan dengan algoritma yang lebih canggih seperti XGBoost, LightGBM, maupun Deep Learning, serta dilengkapi dengan metode interpretabilitas (Explainable AI) agar dapat memberikan pemahaman lebih mendalam mengenai faktor dominan yang memengaruhi kinerja karyawan.

Dengan demikian, dapat disimpulkan bahwa Random Forest merupakan algoritma paling andal dalam penelitian ini dan berpotensi digunakan sebagai model utama untuk penilaian kinerja karyawan berbasis rekam jejak digital.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia atas dukungan kebijakan dan pendanaan yang telah mendorong pengembangan riset di lingkungan perguruan tinggi. Apresiasi juga ditujukan kepada Universitas Potensi Utama yang telah memfasilitasi penyediaan sarana, bimbingan akademik, serta kesempatan bagi penulis untuk melaksanakan penelitian ini. Dukungan dari kedua pihak tersebut sangat berarti dalam terselesaikannya penelitian ini.

REFERENSI

- [1] M. Barry and A. Wilkinson, 'Employee voice, psychologisation and human resource management (HRM)', *Hum. Resour. Manag. J.*, vol. 32, no. 3, 2022, doi: 10.1111/1748-8583.12415.
- [2] T. D. N. Vuong and L. T. Nguyen, 'The Key Strategies for Measuring Employee Performance in Companies: A Systematic Review', *Sustainability*, vol. 14, no. 21, 2022, doi: 10.3390/su142114017.
- [3] K. M. Omar, 'Factors Motivating Human Resources Management (HRM) in the Public and Private Sectors', *Open J. Bus. Manag.*, vol. 09, no. 02, 2021, doi: 10.4236/ojbm.2021.92036.

- [4] Joel Rodriguez and Kelley Walters, 'The Importance of Training and Development in Employee Performance and Evaluation', *Int. J. Peer Rev. J. Ref. J. Index. J. UGC Approv. J. Impact Factor*, vol. 3, no. 10, 2017.
- [5] A. Fitriana *et al.*, 'Decision support system of employee performance evaluation', *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special Issue 2, 2019, doi: 10.35940/ijeat.F1307.0886S219.
- [6] Mutinda Teguh Widayanto, Yuliana, Ismawati, and Nurdin Natan, 'Implementation Of Performance Assessment To Determine Employee Performance', *Int. J. Sci. Technol. Manag.*, vol. 2, no. 5, 2021, doi: 10.46729/ijstm.v2i5.302.
- [7] J. Flanagan and R. K. Burns, 'The Employee Performance Record: a New Appraisal and Development Tool', *Harv. Bus. Rev.*, vol. 79, no. July, 2020.
- [8] P. C. Agu, E. E. Njoku, S. Umaru, N. I. Eleke, I. H. Nwokoma, and O. T. Bashiru, 'Records Management and Organizational Performance', *ARRUS J. Soc. Sci. Humanit.*, vol. 2, no. 2, 2022, doi: 10.35877/soshum738.
- [9] Y. Pan and F. J. Froese, 'An interdisciplinary review of AI and HRM: Challenges and future directions', *Hum. Resour. Manag. Rev.*, vol. 33, no. 1, 2023, doi: 10.1016/j.hrmr.2022.100924.
- [10] S. Lada, B. Chekima, M. Rahimie Abdul Karim, N. F. Fabeil, and M. Salleh Ayub, 'Determining factors related to artificial intelligence (AI) adoption among Malaysia's small and medium-sized businesses', *J. Open Innov. Technol. Mark. Complex.*, vol. 9, no. 4, Dec. 2023.
- [11] D. Gélina, A. Sadreddin, and R. Vahidov, 'Artificial intelligence in human resources management: A review and research agenda', *Pacific Asia J. Assoc. Inf. Syst.*, vol. 14, no. 6, 2022, doi: 10.17705/1pais.14601.
- [12] H. Li, Q. Wang, J. Liu, and D. Zhao, 'A Prediction Model of Human Resources Recruitment Demand Based on Convolutional Collaborative BP Neural Network', *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3620312.
- [13] K. Patel *et al.*, 'RanKer: An AI-Based Employee-Performance Classification Scheme to Rank and Identify Low Performers', *Mathematics*, vol. 10, no. 19, 2022, doi: 10.3390/math10193714.
- [14] P. Sujatha and R. S. Dhivya, 'Ensemble Learning Framework to Predict the Employee Performance', 2022, doi: 10.1109/ICPC2T53885.2022.9777078.
- [15] I. Ahmed, I. Sultana, S. K. Paul, and A. Azeem, 'Employee performance evaluation: A fuzzy approach', *Int. J. Product. Perform. Manag.*, vol. 62, no. 7, 2013, doi: 10.1108/IJPPM-01-2013-0013.
- [16] R. Lv, 'Research on human resource management performance evaluation based on BP algorithm', *Appl. Math. Nonlinear Sci.*, vol. 8, no. 2, 2023, doi: 10.2478/amns.2021.2.00302.
- [17] A. S. Lather, R. Malhotra, P. Saloni, P. Singh, and S. Mittal, 'Prediction of employee performance using machine learning techniques', 2019, doi: 10.1145/3373477.3373696.
- [18] M. Siahaan, 'An Analysis of Contract Employee Performance Assessment Using Machine Learning', *J. INFORMATICS Telecommun. Eng.*, vol. 5, no. 1, 2021, doi: 10.31289/jite.v5i1.5357.
- [19] Safrizal, L. Tanti, I. F. Rahmad, and Y. Thanri, 'Monitoring and Evaluation of Flight Instructor Performance with PROMETHEE Method', 2019, doi: 10.1109/CITSM.2018.8674261.
- [20] Y. Y. Thanri and F. P. Nasution, '... Sistem Pendukung Keputusan Dalam Menentukan Operator Caroline Officer Inbond 188 Terbaik dengan Metode AHP Dan SMART Berdasarkan Rekam Jejak Kinerja ...', *J. Media ...*, vol. 4, 2020.

- [21] Y. Y. Thanri, L. Sipahutar, and F. Putri, Ade, 'PENENTUAN NILAI KINERJA PADA OPERATOR MENGGUNAKAN DECISION SUPPORT SYSTEM DENGAN METODE COPRAS', *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 13, no. 3a, pp. 296–305, 2021.
- [22] Y. Y. Thanri, 'Rekam Jejak Kinerja Crew Sebagai Model Sistem Pendukung Keputusan Pengangkatan Leader', *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 1, 2020, doi: 10.30865/mib.v4i1.1785.