

Optimasi Model Evaluasi Kinerja Karyawan Berbasis Rekam Jejak Digital Menggunakan PCA dan Algoritma Machine Learning

Yan Yang Thanri*, Juli Iriani, Angel Gowasa, Luthfi Zaidi

Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Potensi Utama, Medan, Indonesia

Email: ^{1,*}ythanri@gmail.com, ²juliiriani179@gmail.com, ³angel_gowasa@gmail.com, ⁴luthfi_zaidi@gmail.com

Email Penulis Korespondensi: ythanri@gmail.com

Submitted: 23/09/2025; Accepted: 30/06/2026; Published: 30/06/2026

Abstrak—Di era transformasi digital, organisasi menghadapi tantangan untuk mengevaluasi kinerja karyawan secara objektif dan berbasis data. Sistem penilaian kinerja tradisional sering kali mengandung subjektivitas dan keterbatasan dalam integrasi data, sehingga kurang efektif dalam lingkungan kerja yang dinamis. Penelitian ini bertujuan mengembangkan model evaluasi kinerja berbasis rekam jejak digital menggunakan pendekatan machine learning dan analisis multivariat. Rekam jejak digital mencakup data *aktivitas kerja* (jam kerja harian, waktu layar, pertemuan, dan email), *data wearable* (langkah fisik, durasi tidur, dan stres), *kepuasan* (keseimbangan kerja-hidup, dukungan organisasi), *kemampuan* (skor keterampilan, level jabatan, dan pelatihan), serta *data organisasi* (gaji, insentif, dan lembur). Teknik Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data dan mengidentifikasi indikator kinerja utama. Tiga algoritma machine learning, yaitu Decision Tree, Random Forest, dan Gradient Boosting, diterapkan untuk mengklasifikasikan tingkat kinerja karyawan ke dalam kategori Low, Average, Good, dan Excellent. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Gradient Boosting yang dikombinasikan dengan PCA menghasilkan performa terbaik dengan akurasi 0,887 dan F1-score 0,884. Penerapan PCA terbukti meningkatkan performa model klasifikasi secara signifikan dengan mengurangi noise dan multikolinearitas pada data berdimensi tinggi. Temuan ini menunjukkan potensi besar pemanfaatan data perilaku digital karyawan dalam membangun sistem evaluasi kinerja yang transparan dan adaptif. Penelitian ini memberikan kontribusi terhadap pengembangan manajemen SDM cerdas serta mendukung pengambilan keputusan berbasis data dalam organisasi modern.

Kata Kunci: Model Kinerja Karyawan; Rekam Jejak Digital; PCA; Decision Tree; Random Forest; Gradient Boosting; Analitik SDM

Abstract—In the era of digital transformation, organizations face challenges in evaluating employee performance objectively and based on data. Traditional performance appraisal systems often contain subjectivity and limitations in data integration, making them less effective in dynamic work environments. This study aims to develop a performance evaluation model based on digital footprints using machine learning and multivariate analysis. Digital footprints include work activity data (daily working hours, screen time, meetings, and emails), wearable data (physical steps, sleep duration, and stress levels), satisfaction (work-life balance, organizational support), capability (tech skills score, job level, and training), and organizational data (salary, incentives, and overtime). Principal Component Analysis (PCA) is used to reduce data dimensions and identify key performance indicators. Three machine learning algorithms—Decision Tree, Random Forest, and Gradient Boosting—are applied to classify employee performance into Low, Average, Good, and Excellent categories. Model evaluation is performed using accuracy, precision, recall, and F1-score metrics. The results show that the Gradient Boosting model combined with PCA delivers the best performance with an accuracy of 0.887 and an F1-score of 0.884. The application of PCA significantly improved classification model performance by reducing noise and multicollinearity in high-dimensional data. These findings highlight the great potential of leveraging employees' digital behavioral data to build a transparent and adaptive performance evaluation system. This study contributes to the development of intelligent HR management and supports data-driven decision-making in modern organizations.

Keywords: Employee Performance; Digital Footprint; Principal Component Analysis (PCA); Machine Learning; Gradient Boosting; Human Resource Analytics

1. PENDAHULUAN

Di era transformasi digital yang terus berkembang, organisasi menghadapi tantangan baru dalam mengelola sumber daya manusia (SDM) secara lebih efektif dan berbasis data. Salah satu aspek krusial dalam manajemen SDM adalah penilaian kinerja karyawan, yang menjadi dasar dalam pengambilan keputusan strategis seperti promosi, pengembangan karier, pemberian insentif, hingga perencanaan pelatihan [1], [2]. Namun, proses penilaian kinerja secara tradisional masih banyak menghadapi hambatan, seperti subjektivitas penilai, keterbatasan pengamatan langsung, serta kurangnya integrasi data kerja yang komprehensif [3], [4].

Seiring meningkatnya penggunaan teknologi digital dalam dunia kerja, seperti sistem absensi elektronik, platform manajemen proyek, komunikasi daring, hingga perangkat wearable, muncul peluang baru untuk memanfaatkan rekam jejak digital sebagai indikator kinerja yang lebih obyektif [5]. Rekam jejak digital ini mencakup data aktivitas karyawan yang secara otomatis terekam, seperti jumlah email yang dikirim, waktu login dan logout, partisipasi dalam rapat virtual, dan durasi waktu aktif harian. Bahkan, beberapa studi juga mengintegrasikan faktor fisik dan psikologis, seperti data tidur, tingkat stres, dan kesehatan mental, sebagai bagian dari indikator produktivitas [1], [6].

Namun, meskipun data tersebut tersedia secara melimpah, tantangan utama terletak pada bagaimana menganalisis data multivariat tersebut secara sistematis dan bermakna. Metode tradisional cenderung tidak mampu menangkap hubungan non-linear antara faktor-faktor penentu kinerja, sementara pendekatan statistik klasik seperti regresi linier memiliki keterbatasan dalam menangani data besar dan kompleks [7], [8]. Oleh karena itu, pendekatan

berbasis pembelajaran mesin (machine learning) menjadi relevan untuk dikembangkan karena memiliki kemampuan dalam menangani data yang beragam, besar, dan non-linear [9].

Beberapa studi sebelumnya telah mencoba membangun model evaluasi kinerja berbasis teknologi. [10] mengembangkan model regresi untuk mengevaluasi efisiensi kerja menggunakan data log aktivitas. [11] menggunakan random forest untuk mengidentifikasi variabel dominan yang memengaruhi kinerja tim. Di sisi lain, PCA (Principal Component Analysis) [12] dan analisis faktor telah digunakan dalam reduksi dimensi data kinerja untuk menghasilkan interpretasi yang lebih ringkas namun informatif [13], [14]. Meskipun demikian, pendekatan tersebut seringkali hanya memanfaatkan satu atau dua jenis data, sehingga belum mampu menggabungkan berbagai aspek kinerja secara holistik. Penelitian ini berbeda dengan penelitian terdahulu karena mengintegrasikan berbagai jenis data digital karyawan dalam satu model evaluasi yang komprehensif. Model ini tidak hanya mengandalkan data aktivitas kerja, tetapi juga mempertimbangkan faktor-faktor fisik dan psikologis melalui data wearable, serta pengaruh budaya kerja dan dukungan organisasi. Pendekatan ini menggabungkan faktor-faktor yang lebih luas dalam evaluasi kinerja, yang sering kali terpisah dalam penelitian sebelumnya.

Selain itu, meskipun telah ada penggunaan pembelajaran mesin dalam evaluasi kinerja, penelitian ini mengembangkan pendekatan multivariat dengan mengintegrasikan *Principal Component Analysis (PCA)* untuk mereduksi dimensi data dan mengidentifikasi indikator kinerja utama. Penggunaan teknik ini memungkinkan model untuk mengurangi noise dan multikolinearitas dalam data berdimensi tinggi, yang meningkatkan akurasi dan efektivitas klasifikasi kinerja. Sebelumnya, penelitian cenderung menggunakan teknik regresi atau klasifikasi yang terbatas pada satu atau dua jenis data, sedangkan penelitian ini menawarkan pendekatan yang lebih holistik dan adaptif terhadap perubahan dalam kinerja karyawan [15], [16].

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model evaluasi kinerja karyawan berbasis machine learning dan analisis multivariat, dengan memanfaatkan rekam jejak digital sebagai sumber data utama. Model ini akan mengintegrasikan berbagai faktor yang memengaruhi kinerja meliputi durasi kerja dan produktivitas harian, tingkat penggunaan perangkat lunak manajemen tugas, kualitas tidur dan tingkat stres berdasarkan wearable, keterampilan digital dan adaptasi teknologi dan dukungan organisasi dan budaya kerja digital

Dengan menerapkan pendekatan supervised machine learning seperti Decision Tree, Random Forest, dan Gradient Boosting, serta evaluasi multivariat menggunakan PCA dan klasifikasi adaptif, model ini diharapkan mampu tidak hanya memprediksi skor kinerja karyawan secara kuantitatif, tetapi juga mengklasifikasikan tingkat kinerja (Low, Average, Good, Excellent) dengan obyektif [15], [16].

Signifikansi dari penelitian ini terletak pada kontribusinya terhadap praktik manajemen SDM modern. Model yang dikembangkan berpotensi mengurangi bias subjektif dalam proses evaluasi dan memberikan dasar yang kuat bagi organisasi untuk mengadopsi pendekatan berbasis data (data-driven decision making) dalam pengambilan keputusan SDM. Selain itu, penelitian ini memperluas wacana ilmiah dalam integrasi teknologi digital karyawan dengan model evaluasi kinerja yang adaptif dan berkelanjutan, sebagaimana diangkat dalam studi-studi kontemporer [17], [18].

Dengan demikian, penelitian ini memberikan kontribusi teoritis, metodologis, dan praktis bagi pengembangan sistem evaluasi kinerja di era digital, serta menjadi pijakan dalam membangun manajemen SDM yang lebih cerdas, responsif, dan berkelanjutan.

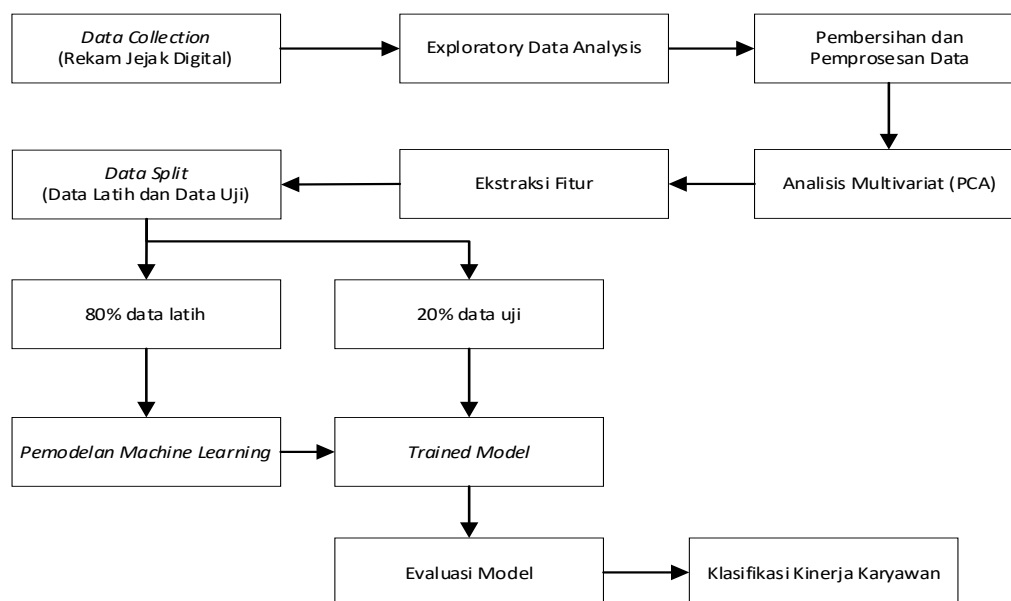
2. METODOLOGI PENELITIAN

2.1 Framework Model

Penelitian ini menggunakan pendekatan kuantitatif berbasis supervised machine learning untuk membangun model penilaian kinerja karyawan berdasarkan rekam jejak digital. Data dikumpulkan dari 2.345 karyawan yang mencakup aktivitas kerja, komunikasi digital, data wearable, kemampuan teknis, kepuasan kerja, dan data organisasi. Tahap penelitian meliputi pengumpulan data, preprocessing, normalisasi menggunakan StandardScaler, serta exploratory data analysis untuk memahami distribusi dan hubungan antarvariabel. Selanjutnya, data kinerja diklasifikasikan ke dalam empat kategori, yaitu Low, Average, Good, dan Excellent.

Tahap berikutnya adalah analisis multivariat menggunakan Principal Component Analysis (PCA) untuk mereduksi dimensi data dan memilih fitur yang paling berpengaruh. Hasil PCA digunakan sebagai input pada model Decision Tree, Random Forest, dan Gradient Boosting dengan pembagian data latih dan data uji sebesar 80:20. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score, di mana kombinasi PCA dan Gradient Boosting memberikan performa terbaik. Seluruh tahapan penelitian disajikan pada Gambar 1.

Tahapan sistematis dalam membangun model penilaian kinerja karyawan berbasis rekam jejak digital yang dianalisis melalui pendekatan pembelajaran mesin dan analisis multivariat terlihat pada Gambar 1. Proses dimulai dengan pengumpulan data rekam jejak digital yang diperoleh dari berbagai sumber aktivitas karyawan, seperti jam kerja harian, jumlah email, partisipasi rapat, penggunaan perangkat lunak manajemen tugas, data wearable (misalnya durasi tidur dan tingkat stres), serta variabel organisasi lainnya. Data yang terkumpul kemudian melalui tahap pemrosesan data yang mencakup pembersihan (data cleaning), normalisasi, penanganan nilai hilang, serta encoding fitur kategorikal agar siap digunakan untuk analisis lebih lanjut. Selanjutnya dilakukan eksplorasi data awal (EDA) berupa analisis deskriptif dan visualisasi guna memahami pola umum, distribusi data, serta hubungan antarvariabel.



Gambar 1. Tahapan Penelitian

Tahap berikutnya adalah analisis multivariat menggunakan Principal Component Analysis (PCA) yang berfungsi mereduksi dimensi data tanpa menghilangkan informasi penting sekaligus mengidentifikasi faktor utama yang memengaruhi kinerja. Dari hasil tersebut dilakukan ekstraksi fitur penting yang berkontribusi besar terhadap variabilitas kinerja untuk dijadikan input ke dalam model pembelajaran mesin. Data kemudian dibagi ke dalam data latih dan data uji dengan rasio perbandingan adalah 80: 20 untuk menjamin evaluasi model yang objektif. Pada tahap pelatihan dan evaluasi model, tiga algoritma pembelajaran mesin, yaitu Decision Tree, Random Forest, dan Gradient Boosting, dilatih menggunakan data latih, kemudian dievaluasi dengan metrik akurasi, precision, recall, dan F1-score. Berdasarkan hasil pelatihan, model digunakan untuk klasifikasi kinerja karyawan ke dalam empat kategori, yakni Low, Average, Good, dan Excellent. Tahap akhir adalah validasi model dan interpretasi hasil, di mana performa model diuji pada data uji untuk menilai kemampuan generalisasi, kemudian hasil analisis digunakan untuk menghasilkan insight yang dapat mendukung pengambilan keputusan dan kebijakan SDM berbasis data.

2.2. Rekam Jejak Digital

Rekam jejak digital dalam penelitian ini merujuk pada kumpulan data aktivitas karyawan yang terdokumentasi secara otomatis melalui berbagai sistem digital dalam organisasi. Data ini mencerminkan perilaku kerja, kondisi fisik dan psikologis, serta faktor organisasi yang berkontribusi terhadap kinerja individu. Komponen rekam jejak digital yang digunakan dalam penelitian ini dapat dilihat 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)

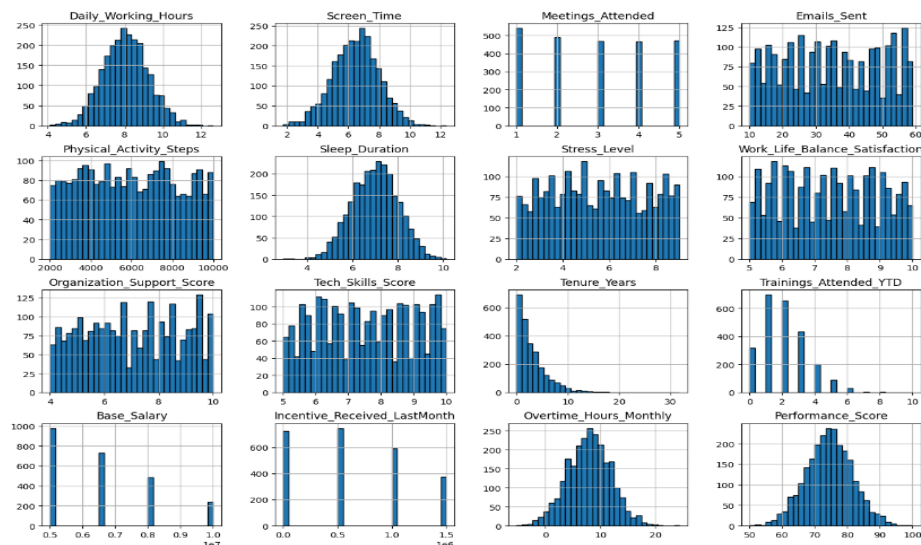
Tabel 1 menyajikan 16 fitur input dan 2 output yang digunakan dalam penilaian kinerja karyawan berbasis rekam jejak digital. Data dikumpulkan dari 2.345 karyawan, berasal dari berbagai sumber, baik melalui aktivitas digital karyawan maupun sistem informasi organisasi, kemudian diklasifikasikan ke dalam enam kategori utama: aktivitas kerja digital, data wearable & kesehatan, persepsi & kepuasan, kemampuan & karier, data organisasi, serta output kinerja.

Fitur-fitur tersebut mencakup berbagai aspek penting seperti produktivitas, kesejahteraan, persepsi kerja, kompetensi, kompensasi, dan hasil kinerja aktual. Pendekatan ini memungkinkan proses evaluasi yang lebih komprehensif, objektif, dan minim bias, karena mengintegrasikan data terukur dari beragam dimensi yang memengaruhi kinerja karyawan.

2.3. Exploratory Data Analysis (EDA)

Analisis Data Eksploratif (EDA) bertujuan untuk memahami karakteristik dataset kinerja karyawan berbasis rekam jejak digital.

Distribusi Fitur Numerik

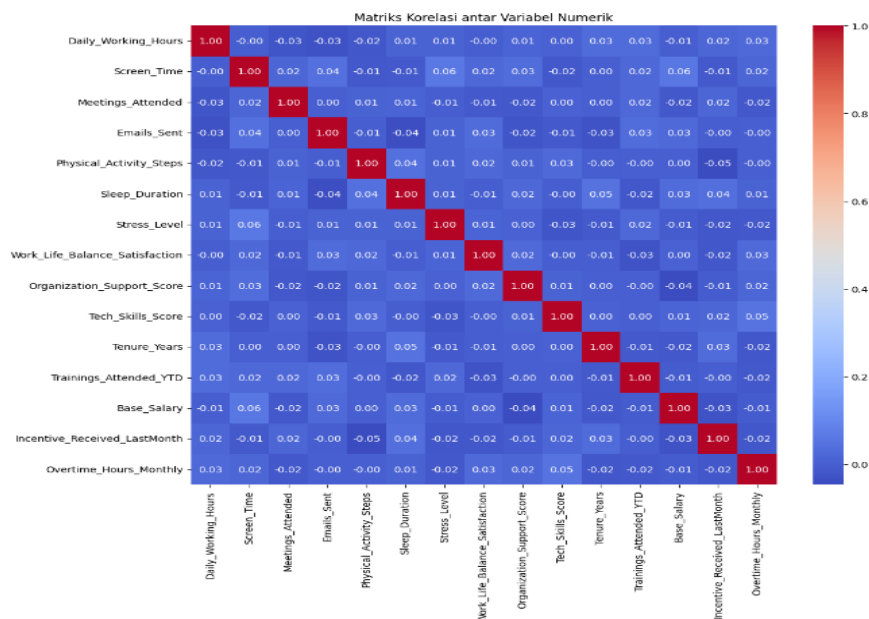


Gambar 2. Grafik distribusi fitur numerik pada kinerja karyawan

Hasil distribusi setiap fitur pada gambar 2 adalah distribusi fitur Daily Working Hours dan Screen Time menunjukkan pola distribusi normal yang simetris, menandakan bahwa mayoritas karyawan bekerja dan menggunakan layar dalam jumlah jam yang mendekati rata-rata, dengan penyebaran yang merata di sekitar nilai tengah. Fitur Meetings Attended dan Emails Sent menunjukkan distribusi kategori atau diskrit yang relatif merata, menandakan variasi besar dalam frekuensi partisipasi rapat dan pengiriman email. Physical Activity Steps dan Sleep Duration juga menunjukkan distribusi mendekati normal, yang berarti kebiasaan fisik dan tidur karyawan cenderung terkonsentrasi di nilai tengah dengan sedikit penyimpangan.

Stress Level, Work-Life Balance Satisfaction, dan Organization Support Score memiliki distribusi yang relatif seragam, mengindikasikan persebaran tingkat stres, kepuasan keseimbangan hidup, dan dukungan organisasi yang beragam di seluruh populasi. Sementara Tech Skills Score dan Trainings Attended YTD memperlihatkan distribusi yang bervariasi namun dengan kecenderungan ke nilai-nilai tertentu. Fitur Tenure Years memiliki distribusi yang condong ke kiri, menunjukkan bahwa sebagian besar karyawan adalah pegawai baru dengan masa kerja pendek.

Base Salary dan Incentive Received LastMonth terlihat sebagai fitur diskrit atau kategori, dengan Base Salary menunjukkan jumlah level gaji tertentu yang sering muncul. Fitur Overtime Hours Monthly dan Performance Score menunjukkan distribusi normal, mengindikasikan bahwa sebagian besar karyawan memiliki jumlah lembur dan skor performa yang berada di sekitar rata-rata. Secara keseluruhan, sebagian besar fitur numerik memiliki distribusi mendekati normal atau seragam, meskipun beberapa fitur seperti Tenure Years dan Trainings Attended YTD menunjukkan kemiringan distribusi.

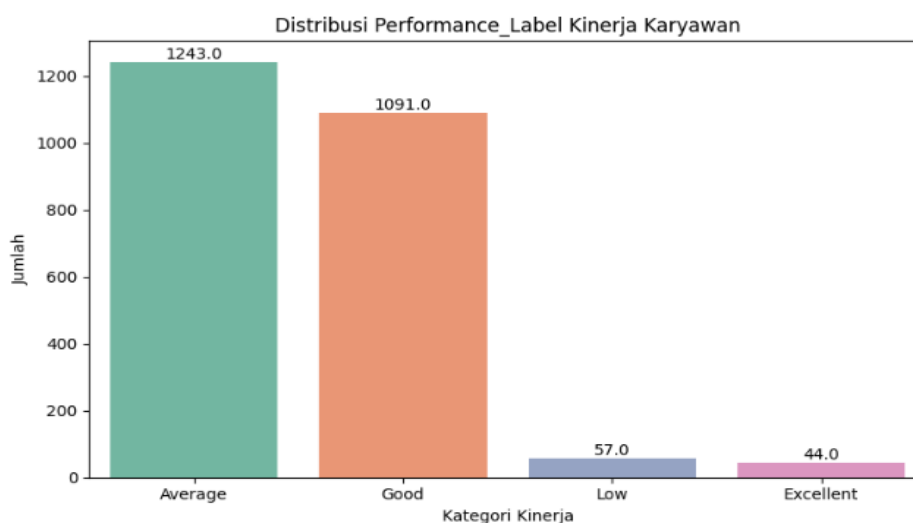


Gambar 3. Matriks Korelasi antar Variabel Numerik pada Kinerja Karyawan

Matriks korelasi antar variabel numerik pada kinerja karyawan yang ditunjukkan pada gambar 3 mengungkapkan hubungan linier di antara berbagai fitur yang memengaruhi performa kerja. Korelasi yang ditampilkan berkisar antara -1 hingga 1, di mana nilai positif menunjukkan hubungan searah, sedangkan nilai negatif menunjukkan hubungan berlawanan. Nilai mendekati nol menandakan korelasi yang lemah atau tidak signifikan. Dari matriks ini, terlihat bahwa sebagian besar variabel memiliki korelasi yang lemah, menunjukkan bahwa setiap fitur memiliki kontribusi yang relatif independen terhadap kinerja karyawan. Misalnya, fitur Daily Working Hours dan Screen Time menunjukkan korelasi positif yang sangat rendah, menandakan bahwa peningkatan jam kerja harian tidak secara langsung berhubungan dengan waktu yang dihabiskan di depan layar.

Sementara itu, Base Salary menunjukkan sedikit korelasi positif dengan Tenure Years, yang logis mengingat pengalaman kerja yang lebih lama sering kali dikaitkan dengan gaji yang lebih tinggi. Namun, korelasi yang rendah pada fitur seperti Stress Level dan Work-Life Balance Satisfaction menunjukkan bahwa faktor psikologis dan keseimbangan hidup tidak memiliki pengaruh langsung yang signifikan terhadap skor kinerja yang terukur.

Dengan demikian, hasil analisis korelasi ini mengindikasikan bahwa model prediktif untuk menilai kinerja karyawan perlu mempertimbangkan kombinasi fitur yang kompleks dan multivariat, mengingat rendahnya korelasi langsung antar fitur yang dianalisis.



Gambar 4. Distribusi Label Kinerja Karyawan

Distribusi label kinerja karyawan pada Gambar 4 yang menunjukkan dominasi kategori Average (1243 karyawan) dan Good (1091 karyawan), serta jumlah yang lebih sedikit pada kategori Low (57 karyawan) dan Excellent (44 karyawan), sejalan dengan temuan statistik deskriptif.

- Waktu Kerja Harian dan Waktu Layar yang menunjukkan distribusi normal menandakan bahwa sebagian besar karyawan bekerja dalam durasi yang wajar, yang mungkin mendukung kinerja rata-rata dan baik.
- Langkah Aktivitas Fisik dan Durasi Tidur yang juga terdistribusi secara normal menunjukkan keseimbangan gaya hidup yang relatif stabil, yang berkontribusi pada stabilitas kinerja.
- Tingkat Stres yang bervariasi dan Kepuasan Keseimbangan Hidup-Kerja yang cukup tinggi menunjukkan bahwa sebagian besar karyawan dapat mengelola beban kerja mereka dengan baik, namun ada beberapa yang mengalami stres tinggi, yang mungkin berkontribusi pada kinerja rendah.
- Dukungan Organisasi yang positif dan Skor Keterampilan Teknis yang tinggi juga mendukung performa karyawan di tingkat yang baik.
- Masa Kerja yang bervariasi menunjukkan bahwa karyawan dengan pengalaman yang lebih lama cenderung memiliki performa yang stabil, sementara karyawan baru masih beradaptasi.

Kesenjangan yang signifikan pada kategori Low dan Excellent dapat dihubungkan dengan perbedaan dalam akses ke pelatihan, insentif, dan dukungan organisasi. Dengan analisis multivariat dan pembelajaran mesin, faktor-faktor kunci yang memengaruhi kinerja dapat diidentifikasi untuk meningkatkan dukungan bagi karyawan dengan performa rendah dan mendorong lebih banyak karyawan mencapai performa Excellent.

2.4. Data preprocessing

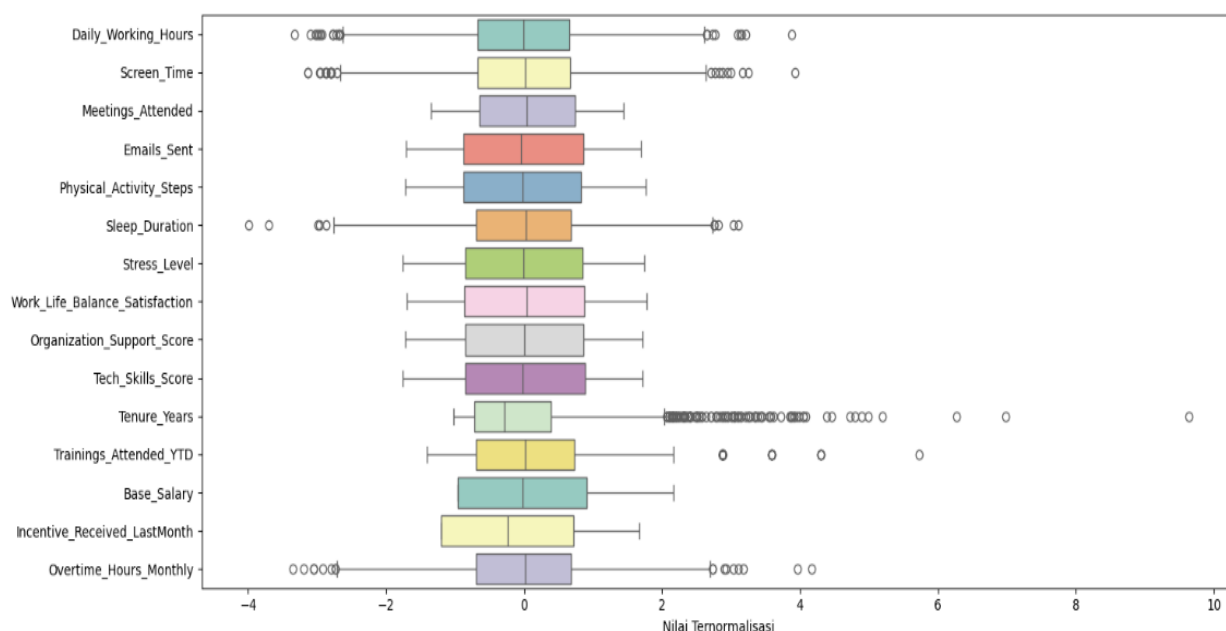
Tahapan data preprocessing dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses analisis multivariat dan pelatihan model pembelajaran mesin. Dataset yang digunakan dalam penelitian ini terdiri dari 2.435 entri dan 19 fitur yang mencerminkan rekam jejak digital karyawan, seperti jam kerja harian, waktu layar, jumlah email, langkah aktivitas fisik, hingga skor kinerja (*Performance Score*)[19][20].

Langkah awal preprocessing adalah pemilihan fitur numerik dengan mengecualikan kolom non-numerik dan label target *Performance_Label*. Fitur numerik diproses menggunakan metode normalisasi *StandardScaler* dengan persamaan (1), yang mengubah setiap fitur sehingga memiliki nilai rata-rata 0 dan standar deviasi 1 terlihat pada

gambar 5. Normalisasi ini penting untuk memastikan bahwa setiap fitur memiliki skala kontribusi yang sebanding dalam analisis Principal Component Analysis (PCA) dan pelatihan model klasifikasi. StandardScaler mengubah setiap nilai fitur x menjadi nilai baru z [21], [19] dengan persamaan (1):

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Dalam proses normalisasi data menggunakan metode z-score, x merupakan nilai asli dari suatu fitur, sedangkan μ menunjukkan nilai rata-rata (*mean*) dari fitur tersebut. Selanjutnya, σ merepresentasikan standar deviasi yang menggambarkan tingkat penyebaran data terhadap nilai rata-ratanya. Adapun z merupakan nilai fitur setelah melalui proses normalisasi, yang disebut sebagai nilai standar dan menunjukkan posisi relatif suatu data terhadap rata-rata berdasarkan satuan standar deviasi. Persamaan (1) menunjukkan proses normalisasi data menggunakan metode StandardScaler, yaitu mengubah setiap nilai fitur menjadi skala standar dengan rata-rata nol dan standar deviasi satu. Proses ini dilakukan dengan cara mengurangi nilai asli fitur (x) dengan nilai rata-rata (μ), kemudian hasilnya dibagi dengan standar deviasi (σ). Transformasi ini bertujuan untuk menghilangkan perbedaan skala antar fitur sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses pembelajaran model. Tanpa normalisasi, fitur dengan rentang nilai yang lebih besar dapat mendominasi model dan menyebabkan bias dalam proses pelatihan. Oleh karena itu, hasil transformasi berupa nilai z merepresentasikan posisi relatif suatu data terhadap distribusinya, sehingga meningkatkan stabilitas dan kinerja algoritma machine learning, terutama pada metode yang sensitif terhadap skala data seperti PCA dan algoritma berbasis jarak.



Gambar 5. Boxplot Hasil Normalisasi Fitur Numerik pada Kinerja Karyawan

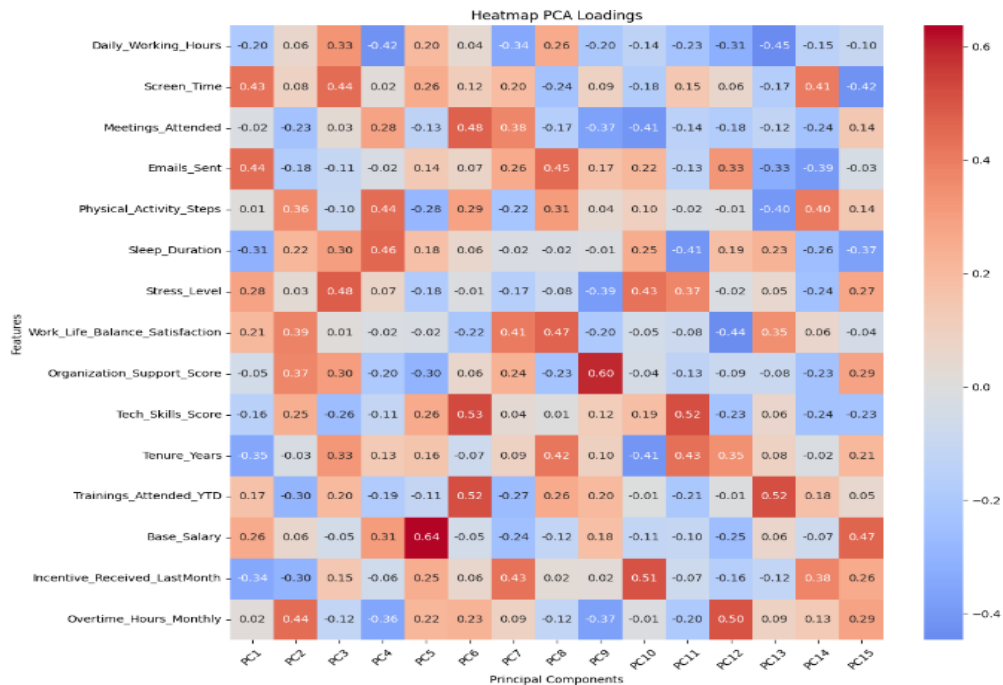
Boxplot Hasil Normalisasi Fitur Numerik pada Kinerja Karyawan pada gambar 5 menunjukkan distribusi data yang telah dinormalisasi menggunakan StandardScaler. Mayoritas fitur, seperti *Daily_Working_Hours* dan *Screen_Time*, memiliki distribusi simetris, sedangkan fitur seperti *Tenure_Years* dan *Base_Salary* mengandung outlier signifikan. Variasi besar terlihat pada *Overtime_Hours_Monthly* dan *Physical_Activity_Steps*, sementara fitur seperti *Work_Life_Balance_Satisfaction* memiliki rentang yang lebih sempit.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Multivariat Rekam Jejak Digital

Analisis multivariat dilakukan untuk mengevaluasi secara komprehensif keterkaitan antar berbagai fitur numerik yang membentuk rekam jejak digital karyawan. Dalam penelitian ini, pendekatan utama yang digunakan adalah Principal Component Analysis (PCA) untuk mereduksi dimensi data dan mengidentifikasi faktor-faktor utama yang berkontribusi terhadap variasi kinerja karyawan. Dengan PCA, variabel yang awalnya berjumlah banyak disederhanakan menjadi beberapa komponen utama (principal components) yang tetap mampu mewakili informasi terpenting dari data. Heatmap pada Gambar 6 menunjukkan beban faktor PCA untuk rekam jejak digital karyawan, yang mengungkapkan kontribusi berbagai fitur terhadap komponen utama yang terbentuk. Warna merah menunjukkan kontribusi positif yang kuat, sedangkan warna biru menunjukkan kontribusi negatif. Fitur seperti *Base Salary* dan *Incentive Received LastMonth* memiliki pengaruh dominan pada PC5, mencerminkan pentingnya faktor finansial dalam kinerja karyawan. *Screen Time* dan *Emails Sent* menunjukkan kontribusi signifikan pada PC1, yang

mengindikasikan hubungan erat antara aktivitas digital dan produktivitas. Selain itu, *Work-Life Balance Satisfaction* berkontribusi positif pada PC8, sementara *Stress Level* berkontribusi negatif pada beberapa komponen, menandakan bahwa keseimbangan kerja-hidup dan tingkat stres memainkan peran penting dalam performa. Distribusi beragam dari fitur seperti *Physical Activity Steps* dan *Sleep Duration* menunjukkan bahwa aspek gaya hidup juga memengaruhi kinerja secara multivariat.

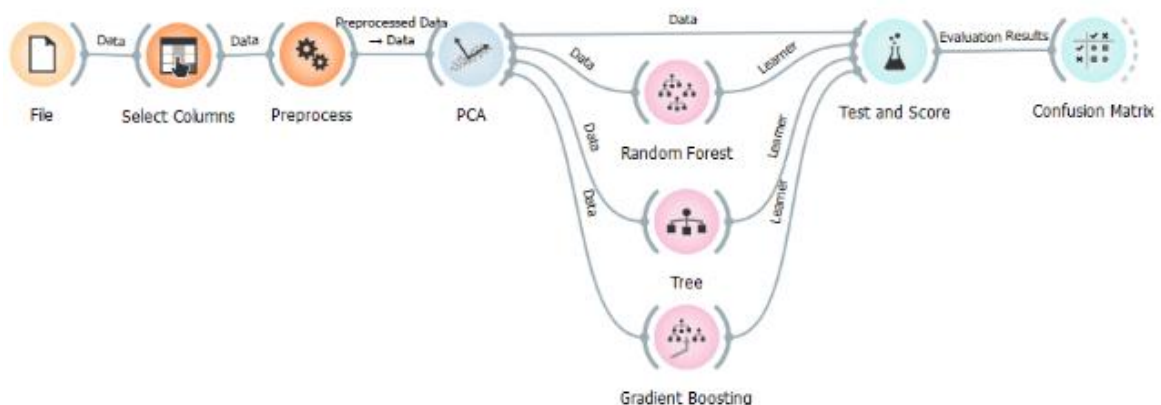


Gambar 6. Heatmap Beban Faktor PCA untuk Rekam Jejak Digital Karyawan

Dengan demikian, hasil heatmap ini menegaskan bahwa PCA berhasil melakukan reduksi fitur dengan menyaring faktor-faktor utama yang paling berpengaruh, sehingga analisis lebih lanjut dapat difokuskan pada komponen yang memiliki kontribusi terbesar terhadap variasi kinerja karyawan.

3.2 Modeling

Untuk membangun model klasifikasi kinerja karyawan berbasis data rekam jejak digital, penelitian ini menggunakan beberapa algoritma machine learning, yaitu Decision Tree, Random Forest, dan Gradient Boosting. Sebelum model dilatih, dilakukan proses reduksi dimensi menggunakan Principal Component Analysis (PCA) terlihat pada gambar 7 guna mengurangi kompleksitas data, meningkatkan efisiensi komputasi, serta membantu meminimalkan overfitting. Setiap model dievaluasi berdasarkan metrik akurasi, presisi, recall, dan f1-score untuk menentukan performa terbaik dalam memprediksi label kinerja karyawan.



Gambar 7. Model Penilaian Kinerja Karyawan berbasis Machine Learning

Arsitektur model penilaian kinerja karyawan pada gambar 7, dikembangkan dengan mengintegrasikan metode pembelajaran mesin dan analisis multivariat berbasis rekam jejak digital. Proses dimulai dari tahap pengumpulan data, di mana data diperoleh dari berbagai sumber digital seperti sistem absensi, log aktivitas pekerjaan, komunikasi digital (email, chat), hingga sistem manajemen kinerja. Selanjutnya, data yang terkumpul melalui tahap preprocessing, yang

mencakup pembersihan data, transformasi variabel, normalisasi, dan pengkodean fitur agar dapat digunakan dalam model pembelajaran mesin. Setelah itu, data yang telah diproses digunakan dalam pemodelan machine learning, seperti Decision Tree, Random Forest, atau Support Vector Machine, untuk memprediksi atau mengklasifikasikan tingkat kinerja karyawan. Hasil dari model ini kemudian dianalisis lebih lanjut menggunakan analisis multivariat, seperti Principal Component Analysis (PCA) atau analisis faktor, guna mengidentifikasi dimensi utama yang memengaruhi performa. Akhirnya, sistem memberikan output penilaian kinerja dalam bentuk skor, kategori, atau rekomendasi, yang dapat digunakan oleh manajer SDM untuk pengambilan keputusan berbasis data secara objektif dan transparan.

3.3 Evaluasi Model

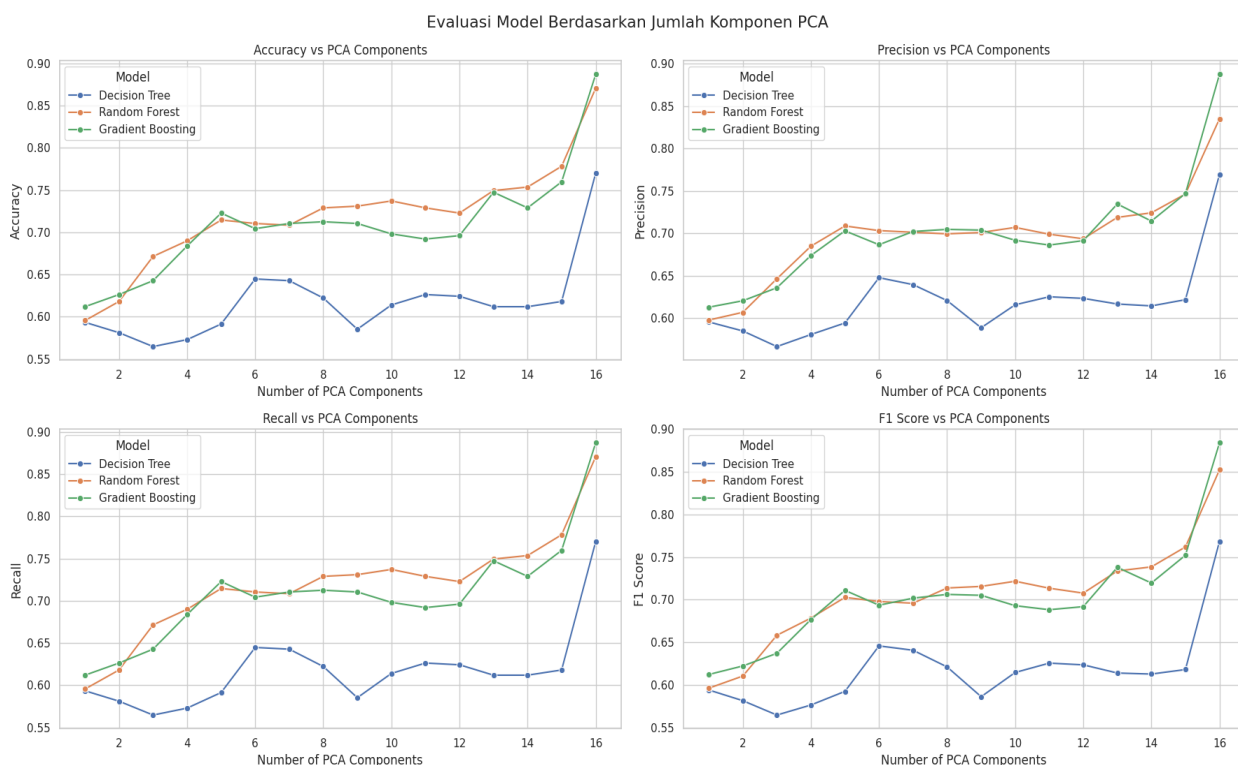
Untuk mengetahui tingkat kinerja dari model pembelajaran mesin yang digunakan dalam menilai kinerja karyawan, dilakukan proses evaluasi menggunakan beberapa metrik pengukuran. Evaluasi ini bertujuan untuk mengukur seberapa baik model dalam mengklasifikasikan data secara akurat berdasarkan rekam jejak digital yang tersedia. Beberapa metrik yang digunakan meliputi akurasi, presisi, recall, dan F1-score. Hasil evaluasi disajikan dalam bentuk tabel 2 dan visualisasi grafik terlihat pada gambar 8 untuk mempermudah interpretasi dan analisis performa dari setiap model yang diimplementasikan.

Tabel 2. Hasil Evaluasi Model Machine Learning berdasarkan Principal Component Analysis (PCA)

PCA	Model	Accuracy	Precision	Recall	F1 Score
1	Decision Tree	0,593	0,595	0,593	0,594
	Random Forest	0,595	0,598	0,595	0,596
	Gradient Boosting	0,612	0,613	0,612	0,612
2	Decision Tree	0,581	0,585	0,581	0,582
	Random Forest	0,618	0,607	0,618	0,610
	Gradient Boosting	0,626	0,620	0,626	0,622
3	Decision Tree	0,565	0,566	0,565	0,565
	Random Forest	0,671	0,646	0,671	0,658
	Gradient Boosting	0,643	0,636	0,643	0,637
4	Decision Tree	0,573	0,581	0,573	0,576
	Random Forest	0,690	0,685	0,690	0,678
	Gradient Boosting	0,684	0,674	0,684	0,677
5	Decision Tree	0,591	0,594	0,591	0,592
	Random Forest	0,715	0,709	0,715	0,703
	Gradient Boosting	0,723	0,703	0,723	0,711
6	Decision Tree	0,645	0,648	0,645	0,646
	Random Forest	0,710	0,703	0,710	0,698
	Gradient Boosting	0,704	0,686	0,704	0,694
7	Decision Tree	0,643	0,639	0,643	0,641
	Random Forest	0,708	0,701	0,708	0,696
	Gradient Boosting	0,710	0,702	0,710	0,702
8	Decision Tree	0,622	0,621	0,622	0,621
	Random Forest	0,729	0,699	0,729	0,714
	Gradient Boosting	0,713	0,705	0,713	0,706
9	Decision Tree	0,585	0,589	0,585	0,586
	Random Forest	0,731	0,701	0,731	0,716
	Gradient Boosting	0,710	0,704	0,710	0,705
10	Decision Tree	0,614	0,616	0,614	0,615
	Random Forest	0,737	0,707	0,737	0,722
	Gradient Boosting	0,698	0,692	0,698	0,693
11	Decision Tree	0,626	0,625	0,626	0,626
	Random Forest	0,729	0,699	0,729	0,714
	Gradient Boosting	0,692	0,686	0,692	0,688
12	Decision Tree	0,624	0,623	0,624	0,624
	Random Forest	0,723	0,693	0,723	0,708
	Gradient Boosting	0,696	0,691	0,696	0,692
13	Decision Tree	0,612	0,617	0,612	0,614
	Random Forest	0,749	0,719	0,749	0,734
	Gradient Boosting	0,747	0,734	0,747	0,738
14	Decision Tree	0,612	0,614	0,612	0,613
	Random Forest	0,754	0,724	0,754	0,739
	Gradient Boosting	0,729	0,714	0,729	0,720

PCA	Model	Accuracy	Precision	Recall	F1 Score
15	Decision Tree	0,618	0,622	0,618	0,618
	Random Forest	0,778	0,746	0,778	0,762
	Gradient Boosting	0,760	0,747	0,760	0,752
16	Decision Tree	0,770	0,769	0,770	0,768
	Random Forest	0,871	0,835	0,871	0,852
	Gradient Boosting	0,887	0,887	0,887	0,884

Hasil evaluasi pada Tabel 2 disajikan dalam bentuk grafik untuk memperjelas peningkatan akurasi pada setiap jumlah komponen PCA, seperti yang ditampilkan pada Gambar 9.



Gambar 9. Grafik evaluasi model

Berdasarkan Tabel 2 dan Gambar 9 grafik evaluasi performa model berdasarkan jumlah komponen PCA, dapat disimpulkan bahwa pemilihan jumlah komponen utama (Principal Component) dalam PCA secara signifikan mempengaruhi performa model klasifikasi. Tiga algoritma yang diuji Decision Tree, Random Forest, dan Gradient Boosting menunjukkan tren peningkatan performa seiring bertambahnya jumlah komponen PCA yang digunakan.

Pada tabel 2, performa masing-masing model dievaluasi berdasarkan empat metrik: Accuracy, Precision, Recall, dan F1 Score. Hasilnya menunjukkan bahwa model Gradient Boosting (Tuned) secara konsisten unggul dibandingkan dua model lainnya. Sebagai contoh, ketika jumlah komponen PCA mencapai 26, Gradient Boosting mencapai akurasi tertinggi sebesar 0.887, precision 0.887, recall 0.887, dan F1 score 0.884. Sementara itu, Decision Tree memiliki nilai tertinggi yang jauh lebih rendah, dengan akurasi hanya 0.77 pada titik yang sama.

Grafik pada gambar 8 yang menyertai tabel 2 memperjelas tren performa tersebut. Setiap grafik menunjukkan hubungan antara jumlah komponen PCA (sumbu X) dengan masing-masing metrik evaluasi (sumbu Y). Dari keempat grafik (Accuracy, Precision, Recall, dan F1 Score), terlihat bahwa Gradient Boosting (garis hijau) menunjukkan tren peningkatan yang stabil dan signifikan sejak awal hingga komponen ke-16. Random Forest (garis merah) berada di posisi kedua dalam hal performa, dengan kenaikan yang stabil namun tidak sekuat Gradient Boosting. Sebaliknya, Decision Tree (garis biru) menunjukkan performa yang fluktuatif dan cenderung stagnan, serta jauh di bawah dua model lainnya, menunjukkan bahwa model ini kurang cocok untuk dataset setelah diterapkan PCA.

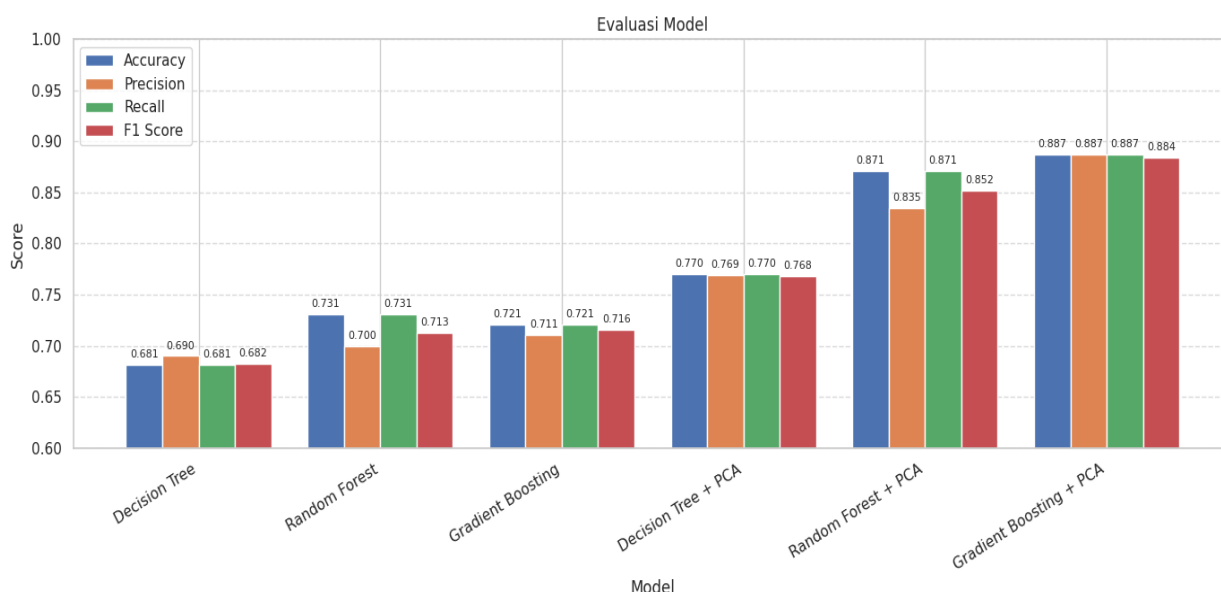
Untuk mengevaluasi pengaruh penerapan Principal Component Analysis (PCA) terhadap performa model machine learning, dilakukan perbandingan antara hasil klasifikasi dari model Decision Tree, Random Forest, dan Gradient Boosting baik dengan maupun tanpa PCA. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil perbandingan dapat dilihat pada Tabel 3 berikut.

Tabel 3. Perbandingan Hasil Evaluasi Model Machine Learning dengan dan tanpa PCA

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	0,681	0,690	0,681	0,682

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0,731	0,700	0,731	0,713
Gradient Boosting	0,721	0,711	0,721	0,716
Decision Tree+PCA	0,770	0,769	0,770	0,768
Random Forest+PCA	0,871	0,835	0,871	0,852
Gradient Boosting+PCA	0,887	0,887	0,887	0,884

Untuk memberikan gambaran yang lebih jelas mengenai perbandingan performa masing-masing model, dilakukan visualisasi dalam bentuk diagram batang. Grafik ini menunjukkan perbedaan nilai akurasi, presisi, recall, dan F1-score antara model dengan PCA dan tanpa PCA, seperti ditampilkan pada Gambar 10 berikut.



Gambar 10. Grafik Perbandingan Evaluasi Model dengan dan tanpa PCA

Berdasarkan Tabel 6 dan Gambar 5, terlihat bahwa penerapan PCA memberikan dampak positif terhadap performa model machine learning yang digunakan dalam klasifikasi penyakit. Secara umum, nilai akurasi, presisi, recall, dan F1-score pada ketiga model (Decision Tree, Random Forest, dan Gradient Boosting) mengalami peningkatan setelah dilakukan reduksi dimensi menggunakan PCA. Sebaliknya, performa model tanpa PCA cenderung lebih rendah, yang mengindikasikan bahwa adanya fitur yang tidak relevan atau redundan dapat memengaruhi kualitas klasifikasi.

Model Gradient Boosting menunjukkan kinerja paling konsisten dan tinggi setelah penerapan PCA, terutama pada nilai F1-score yang mencapai angka tertinggi dibanding model lainnya. Hal ini menunjukkan bahwa PCA membantu model dalam menangkap pola yang lebih representatif dari data utama, dengan mengurangi noise dan kolinearitas antar fitur. Selain itu, model Random Forest juga menunjukkan peningkatan yang signifikan dalam presisi dan akurasi setelah menggunakan PCA, yang memperkuat peran PCA dalam membantu proses pemodelan.

Sebaliknya, model tanpa PCA menunjukkan ketidakseimbangan antara nilai presisi dan recall. Hal ini menandakan bahwa model tersebut mengalami kesulitan dalam membedakan antara kelas target secara optimal. F1-score yang relatif lebih rendah juga menunjukkan bahwa keseimbangan antara presisi dan recall tidak tercapai dengan baik pada model tanpa PCA. Maka dari itu, dapat disimpulkan bahwa penerapan PCA mampu meningkatkan kinerja model klasifikasi, terutama dalam konteks data berdimensi tinggi seperti pada kasus ini.

3.4 Pembahasan

Dalam upaya meningkatkan efektivitas evaluasi kinerja karyawan berbasis data, berbagai penelitian telah mengembangkan dan mengevaluasi model menggunakan pendekatan yang berbeda. Tabel berikut menyajikan perbandingan akurasi dari beberapa model terkini dengan model yang diusulkan dalam studi ini. Perbandingan ini bertujuan untuk menunjukkan posisi pendekatan yang digunakan dalam studi ini dibandingkan dengan metode yang telah ada, serta memberikan gambaran menyeluruh mengenai performa berbagai kombinasi algoritma yang digunakan. Tabel 5 menggambarkan posisi studi ini dibandingkan dengan literatur terkini.

Tabel 4. Perbandingan studi ini dengan literatur terkini

Model	Accuracy	Referensi
Logistic Regression	0,75	[22]
Logistic Regression	0,86	[15]
Random Forest	0,87	[23]

Model	Accuracy	Referensi
Decision Tree+PCA	0,77	Studi ini
Random Forest+PCA	0,87	Studi ini
Gradient Boosting+PCA	0,89	Studi ini

Tabel 5 membandingkan performa berbagai model yang digunakan untuk evaluasi kinerja karyawan dengan akurasi sebagai metrik evaluasi utama. Studi ini menunjukkan bahwa kombinasi Gradient Boosting + PCA memberikan akurasi tertinggi sebesar 88,70%, mengungguli seluruh model pembandingan yang dilaporkan pada literatur terkini. Model Random Forest + PCA dalam studi ini (87,10%) juga sebanding dan sedikit lebih tinggi dibandingkan Random Forest pada penelitian Singh dan Khuteja [22] (87,00%), meskipun keduanya menggunakan dataset dan fitur yang berbeda.

Dibandingkan dengan literatur terkini, model dari studi ini mengungguli metode lain seperti Logistic Regression/Gradient Boosting pada Alqahtani dan Almaleh[15] (86,00%) dan model Logistic Regression pada studi atrisi karyawan oleh Setiawan dkk. (75,00%). Model AdaBoost, SVM, dan Random Forest pada studi prediksi turnover karyawan mencapai akurasi maksimum 88,00% [23], mendekati performa Gradient Boosting + PCA dalam studi ini, namun tetap berada di bawahnya.

Secara keseluruhan, hasil dalam tabel ini menggarisbawahi keunggulan kombinasi PCA dengan algoritma ensemble seperti Random Forest dan Gradient Boosting dalam meningkatkan akurasi klasifikasi kinerja karyawan, serta menegaskan kelebihan pendekatan studi ini dibandingkan dengan literatur yang ada.

4. KESIMPULAN

Penelitian ini telah berhasil mengevaluasi dan membandingkan performa tiga algoritma machine learning yakni Decision Tree, Random Forest, dan Gradient Boosting dalam mengklasifikasikan kinerja karyawan, baik tanpa maupun dengan penerapan teknik reduksi dimensi Principal Component Analysis (PCA). Hasil pengujian menunjukkan bahwa penerapan PCA secara signifikan meningkatkan performa dari semua algoritma yang diuji. Tanpa PCA, model Random Forest dan Gradient Boosting menunjukkan akurasi dan nilai evaluasi yang cukup baik, masing-masing dengan akurasi 0,731 dan 0,721. Namun, setelah penerapan PCA, nilai akurasi meningkat secara signifikan, khususnya pada Gradient Boosting+PCA yang mencapai akurasi tertinggi sebesar 0,887 disertai dengan nilai precision, recall, dan F1 Score yang seimbang. Penemuan ini menegaskan bahwa integrasi PCA dapat memperbaiki efektivitas model klasifikasi dengan mengurangi dimensi data tanpa kehilangan informasi penting, sehingga dapat mengurangi kompleksitas model sekaligus meningkatkan kinerjanya. Kontribusi utama penelitian ini adalah membuktikan bahwa kombinasi antara PCA dan algoritma ensemble seperti Random Forest dan Gradient Boosting sangat efektif untuk menangani data klasifikasi kinerja karyawan. Dari segi implementasi praktis, temuan ini dapat diaplikasikan dalam sistem evaluasi sumber daya manusia untuk mendukung pengambilan keputusan terkait promosi, pelatihan, atau intervensi kinerja. Adapun keterbatasan dalam penelitian ini adalah masih terbatas pada satu sumber dataset dan penggunaan algoritma dengan parameter yang belum dieksplorasi secara menyeluruh. Untuk penelitian selanjutnya, disarankan untuk menguji model pada data yang lebih beragam untuk memastikan generalisasi model yang lebih baik. Selain itu, integrasi data tambahan seperti data kepribadian atau rekam jejak pelatihan juga dapat meningkatkan akurasi klasifikasi. Kesimpulannya, PCA terbukti sebagai teknik yang efektif dalam meningkatkan performa model klasifikasi kinerja karyawan.

UCAPAN TERIMA KASIH

Penulis mengucapkan 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 disampaikan kepada Universitas Potensi Utama yang telah memfasilitasi penyediaan sarana, bimbingan akademik, serta kesempatan untuk melaksanakan penelitian ini.

REFERENCES

- [1] Z. Wang, S. Ren, and L. Meng, 'High-performance work systems and thriving at work: the role of cognitive appraisal and servant leadership', *Pers. Rev.*, vol. 51, no. 7, 2022, doi: 10.1108/PR-10-2019-0561.
- [2] S. S. Kamble, A. Gunasekaran, A. Ghadge, and R. Raut, 'A performance measurement system for industry 4.0 enabled smart manufacturing system in SMMEs- A review and empirical investigation', *Int. J. Prod. Econ.*, vol. 229, 2020, doi: 10.1016/j.ijpe.2020.107853.
- [3] G. Balakayeva, M. Zhanuzakov, and G. Kalmenova, 'Development of a digital employee rating evaluation system (DERES) based on machine learning algorithms and 360-degree method', *J. Intell. Syst.*, vol. 32, no. 1, 2023, doi: 10.1515/jisys-2023-0008.
- [4] A. Ijadi Maghsoodi, G. Abouhamzeh, M. Khalilzadeh, and E. K. Zavadskas, 'Ranking and selecting the best performance appraisal method using the MULTIMOORA approach integrated Shannon's entropy', *Front. Bus. Res. China*, vol. 12, no. 1, 2018, doi: 10.1186/s11782-017-0022-6.
- [5] H. A. Richardson, D. H. Kluemper, and S. G. Taylor, 'Too little and too much authority sharing: Differential relationships

- with psychological empowerment and in-role and extra-role performance', *J. Organ. Behav.*, vol. 42, no. 8, 2021, doi: 10.1002/job.2548.
- [6] L. Yan, L. Wang, X. Shen, P. Li, and J. Guo, 'The Effect of Transformational Change on Performance: An Employee's Stress Appraisals Perspective', *Front. Psychol.*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.897769.
 - [7] I. Aliyu, T. W. Um, S. J. Lee, C. G. Lim, and J. Kim, 'Deep Learning for Multivariate Prediction of Building Energy Performance of Residential Buildings', *Comput. Mater. Contin.*, vol. 75, no. 3, 2023, doi: 10.32604/cmc.2023.037202.
 - [8] S. Dash, S. K. Giri, S. Mallik, S. K. Pani, M. A. Shah, and H. Qin, 'Predictive healthcare modeling for early pandemic assessment leveraging deep auto regressor neural prophet', *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-55973-y.
 - [9] A. Mahmoud and A. Mohammed, 'Leveraging Hybrid Deep Learning Models for Enhanced Multivariate Time Series Forecasting', *Neural Process. Lett.*, vol. 56, no. 5, Oct. 2024, doi: 10.1007/s11063-024-11656-3.
 - [10] Z. Nayem and M. A. Uddin, 'Unbiased employee performance evaluation using machine learning', *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 1, 2024, doi: 10.1016/j.joitmc.2024.100243.
 - [11] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, 'An improved random forest based on the classification accuracy and correlation measurement of decision trees', *Expert Syst. Appl.*, vol. 237, 2024, doi: 10.1016/j.eswa.2023.121549.
 - [12] M. Bilal and M. S. Hanif, 'Fast Anomaly Detection for Vision-Based Industrial Inspection Using Cascades of Null Subspace PCA Detectors', *Sensors*, vol. 25, no. 15, Aug. 2025, doi: 10.3390/s25154853.
 - [13] H. Li, 'Advancing Principal Component Analysis: Challenges and Innovations in Big Data Analysis', *Appl. Comput. Eng.*, vol. 118, pp. 48–54, Feb. 2025, doi: 10.54254/2755-2721/2025.20807.
 - [14] O. Dorabiala, A. Y. Aravkin, and J. N. Kutz, 'Ensemble Principal Component Analysis', *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3350984.
 - [15] F. A. Alqahtani and A. Almaleh, 'Analysis and Prediction of Employee Promotions Using Machine Learning', 2022, doi: 10.1109/DSIT55514.2022.9943959.
 - [16] V. Musanga, E. Tarambiwa, and K. Zvarevashe, 'A Supervised Machine Learning Model to Optimize Human Resources Analytics for Employee Churn Prediction', 2022, doi: 10.1109/ZCICT55726.2022.10045987.
 - [17] P. Zhao, 'Research on Enterprise Human Resource Performance Management Innovation in the Era of Big Data', *BCP Soc. Sci. Humanit.*, vol. 16, 2022, doi: 10.54691/bcpssh.v16i.463.
 - [18] Y. Jiaping, 'Enterprise Human Resource Management Model by Artificial Intelligence Digital Technology', *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/6186811.
 - [19] A. Demircioğlu, 'The effect of feature normalization methods in radiomics', *Insights Imaging*, vol. 15, no. 1, pp. 2–14, Jan. 2024, doi: 10.1186/s13244-023-01575-7.
 - [20] L. Wang, Y. Zhou, K. Sanders, J. H. Marler, and Y. Zou, 'Determinants of effective HR analytics Implementation: An In-Depth review and a dynamic framework for future research', *J. Bus. Res.*, vol. 170, 2024, doi: 10.1016/j.jbusres.2023.114312.
 - [21] L. Arend *et al.*, 'Systematic evaluation of normalization approaches in tandem mass tag and label-free protein quantification data using PRONE', *Brief. Bioinform.*, vol. 26, no. 3, May 2025, doi: 10.1093/bib/bbaf201.
 - [22] I. Setiawan, S. Suprihanto, A. C. Nugraha, and J. Hutahaean, 'HR analytics: Employee attrition analysis using logistic regression', in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 830, no. 3, doi: 10.1088/1757-899X/830/3/032001.