J@TI JURNAL TEKNIK INDUSTRI UNIVERSITAS DIPONEGORO

Vol. 21, No. 1, Januari 2026

STUDI KOMPARATIF MODEL MACHINE LEARNING DALAM MEMPREDIKSI KETERLAMBATAN PEGAWAI: LOGISTIC REGRESSION, SVM, DAN RANDOM FOREST

Inggrid Nindia Aprila Palupi*[1], M. Fariz Fadillah Mardianto[2], Imam Yuadi[3], Budiyan Mariyadi[4]

[1]Program Studi Pengembangan Sumber Daya Manusia, Sekolah Pascasarjana Universitas Airlangga, Jl. Airlangga 4-6, Surabaya, Indonesia 60286
[2]Program Studi Statistika, Jurusan Matematika, Fakultas Sains & Teknologi, Universitas Airlangga, Jl. Dr. Ir. H. Soekarno, Surabaya, Indonesia, 60115
[3]Departemen Ilmu Informasi dan Perpustakaan, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Airlangga Jl. Dharmawangsa Dalam, Airlangga, Kec. Gubeng, Surabaya, Jawa Timur 60286
[4]Program Studi Teknik Industri Universitas Muhammadiyah Bandung, Jl. Soekarno-Hatta No.752, Cipadung Kidul, Kec. Panyileukan, Kota Bandung, Jawa Barat 40614

Abstrak

Keterlambatan karyawan adalah salah satu jenis pelanggaran terhadap disiplin kerja yang dapat berdampak pada produktivitas dan efektivitas organisasi. Penelitian ini bertujuan untuk mengembangkan serta membandingkan performa dari tiga algoritma machine learning Regresi Logistik, SVM, dan Random Forest dalam memprediksi keterlambatan pegawai dengan menggunakan data keterlambatan dan karakteristik individu. Dataset yang digunakan terdiri dari 1902 data, yang dibagi 80% data training dan 20% data testing dengan enam variabel, mencakup usia, lama bekerja, status pernikahan, jarak tempat tinggal ke kantor, jenis kendaraan yang digunakan, dan gaya hidup. Hasil analisis menunjukkan bahwa Random Forest memberikan kinerja prediktif yang paling baik dalam mengenali pegawai yang memiliki potensi untuk terlambat, dengan nilai akurasi tertinggi sebesar 0.82, presisi sebesar 0.93, recall sebesar 0.84, dan F1-score sebesar 0.88. Model ini terbukti dapat menunjukkan kemampuan klasifikasi yang andal dan seimbang. Analisis feature importance mengidentifikasi usia dan masa kerja sebagai faktor paling berpengaruh terhadap prediksi keterlambatan. Temuan ini tidak hanya memberikan wawasan baru dalam pengelolaan kedisiplinan pegawai, tetapi juga membuka peluang implementasi sistem peringatan dini yang dapat diintegrasikan ke dalam sistem kehadiran digital organisasi. Penelitian ini merekomendasikan perluasan variabel untuk studi lanjutan dan pemanfaatan hasil model sebagai dasar penyusunan kebijakan SDM yang lebih adaptif dan berbasis data.

Kata kunci: *keterlambatan pegawai; machine learning; random forest; logistic regression; SVM; Prediksi SDM*

Abstract

[Comparative Study of Machine Learning Models in Predicting Employee Delay: Logistic Regression, SVM, and Random Forest] *Employee tardiness is one type of violation of work discipline that can impact organizational productivity and effectiveness. This study aims to develop and compare the performance of three machine learning algorithms Logistic Regression, SVM, and Random Forest in predicting employee tardiness using tardiness data and individual characteristics. The dataset used consists of 1902 data, which is divided into 80% training data and 20% with six variables, including age, length of service, last education level, marital status, distance from residence to office, type of vehicle used, and lifestyle. The results of the analysis show that Random Forest provides the best predictive performance in identifying employees who have the potential to be late, with the highest accuracy value of 0.82, precision of 0.93, recall of 0.84, and F1-score of 0.88. This model is proven to be able to demonstrate reliable and balanced classification capabilities. Feature importance analysis identifies age and length of service as the most influential factors in predicting tardiness. These findings not only provide new insights into employee discipline management but also open up opportunities for the implementation of an early warning system that can be integrated*

into the organization's digital attendance system. This study recommends expanding the variables for further studies and utilizing the model results as a basis for formulating more adaptive and data-based HR policies.

Keywords: *sustainability industry; developing strategy; MCDM*

1. Pendahuluan

Kehadiran yang tepat waktu adalah salah satu tanda utama dalam penilaian disiplin kerja karyawan dan memberi kontribusi besar pada efisiensi serta produktivitas organisasi (Mutiara & Candra, 2024) (Ishiaka, A. A., 2025). Di berbagai bidang industri, keterlambatan karyawan secara terus-menerus dihubungkan dengan berkurangnya hasil operasional, meningkatnya tekanan kerja rekan-rekan, serta merosotnya budaya kerja secara keseluruhan (Beauty & Thomas, 2021). Oleh karena itu, keterlambatan karyawan tidak hanya merupakan masalah administratif, tetapi juga mempunyai dampak operasional yang meliputi penurunan produktivitas, peningkatan biaya lembur, risiko terhadap pemenuhan SLA (*Service Level Agreement)*, serta dampak negatif pada kesehatan mental dan kepuasan kerja (Glöckner & Lödding, 2019)(Ko & Choi, 2019)(Yu & Leka, 2022)(Shetty & Prabhu, 2024).

Sebelumnya, cara menangani keterlambatan pegawai umumnya bersifat reaktif, misalnya memberikan sanksi atau teguran setelah terjadinya pelanggaran. Metode ini sering kali tidak berhasil dalam mencegah terjadinya kembali pola keterlambatan yang serupa (Rosett & Hagerty, 2021) (Angulakshmi et al., 2024). Dalam zaman digital yang kaya akan data pegawai, perusahaan memiliki kesempatan untuk berubah menjadi organisasi yang berdasarkan data dengan menggunakan *machine learning* (ML) untuk memprediksi perilaku kerja yang tidak diinginkan sebelum hal itu terjadi (Patil & Jadhav, 2023).

Machine learning memungkinkan analisis data karyawan dimasa lalu untuk meramalkan kemungkinan keterlambatan berdasarkan pola-pola yang telah terjadi sebelumnya. Metode ini telah diterapkan secara luas dalam bidang Sumber Daya Manusia untuk meramalkan tingkat keluar karyawan, performa, serta absensi (Kushwaha et al., 2023) (Zupančič & Panov, 2024). Dalam situasi keterlambatan, algoritma *machine learning* (ML) mampu mengolah berbagai variabel seperti jam kerja, jarak dari tempat tinggal, *shift*, dan riwayat keterlambatan, kemudian mengkategorikan kemungkinan seorang pegawai untuk terlambat pada hari tertentu (Angulakshmi et al., 2024).

Berbagai algoritma *machine learning* yang sering digunakan dalam prediksi klasifikasi biner meliputi *Logistic Regression*, *Random Forest*, dan *Support Vector Machine* (SVM). *Logistic Regression* terkenal karena kesederhanaannya dan kemampuannya untuk menghasilkan model yang dapat dengan mudah dipahami oleh pihak manajemen (Sujaini, 2020). Sebaliknya, *Random Forest* menawarkan tingkat akurasi yang tinggi berkat kemampuannya dalam mengelola data yang rumit dan memahami interaksi di antara variabel-variabel tersebut (Reznychenko et al., 2024). Sementara itu, SVM sangat berguna untuk mengatasi situasi data yang tidak linier dan kelas yang tidak seimbang, yang sering ditemukan dalam data absensi (Maharjan, 2022).

Namun, masih sedikit penelitian yang secara langsung membandingkan kinerja ketiga algoritma itu dalam konteks prediksi keterlambatan pegawai. Penelitian oleh Zupancic dan Panov pada tahun 2024, ditemukan bahwa model *tree-based machine learning* dengan pendekatan *predictive clustering trees* dapat secara efektif memprediksi cuti sakit dan cuti liburan pegawai dalam interval jangka pendek maupun panjang menggunakan variabel-variabel seperti profil absensi harian, agregasi absensi sebelumnya, dan data demografis pegawai (Zupančič & Panov, 2024). Sedangkan menurut Bingqing Hu dalam studinya tahun 2021 membuktikan bahwa penerapan *Random Forest* dan *AdaBoost* dengan variabel personal, demografis, dan perilaku karyawan mampu meningkatkan akurasi prediksi ketidakhadiran karyawan (Hu, 2021). Evaluasi yang membandingkan sangat krusial agar organisasi dapat menentukan model prediktif yang paling cocok dengan sifat data yang dimiliki (Mercara, 2020). Sehubungan dengan hal tersebut, penelitian ini memiliki tujuan untuk mengembangkan serta membandingkan kinerja model prediksi keterlambatan pegawai dengan menggunakan tiga pendekatan machine learning: *Logistic Regression*, *Random Forest*, dan SVM.

Hasil dari studi ini diharapkan dapat memperkuat proses pengambilan keputusan strategis dalam manajemen sumber daya manusia, terutama dalam merancang sistem peringatan dini yang mendukung manajemen risiko, peningkatan motivasi karyawan, peningkatan efisiensi operasional, pertumbuhan budaya organisasi yang positif dan pengambilan keputusan berbasis data (Walger et al., 2016)(Sutikno, 2019)(Yan et al., 2020)(Wang et al., 2022)(Arman et al., 2024). Dengan cara ini, organisasi mampu memberikan respons yang proaktif melalui kebijakan pencegahan yang didasarkan pada data (Odionu et al., 2024).

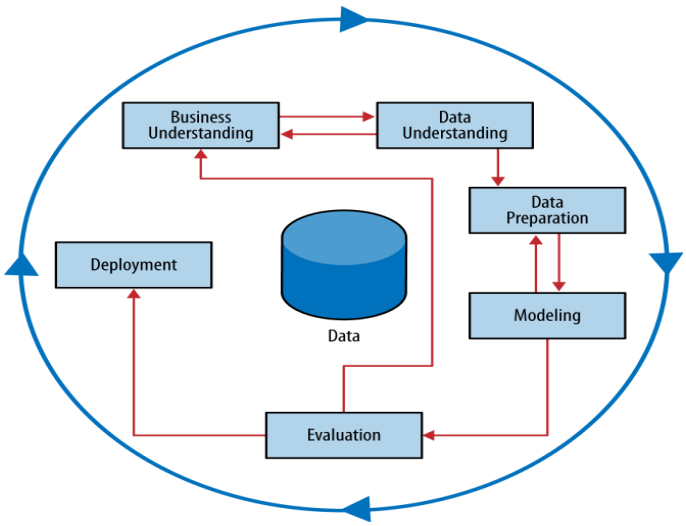
2. Metode Penelitian

2.1 Desain Penelitian

Penelitian ini menerapkan pendekatan kuantitatif eksperimen dengan metode komparatif dalam *data mining* untuk menciptakan dan menilai model prediksi keterlambatan karyawan. Studi ini dirancang untuk membandingkan kinerja tiga algoritma *machine learning*, yaitu *Logistic Regression*, *Random Forest*, dan *Support Vector*

* Penulis Korespondensi

E-mail: inggrid.sch@gmail.com



Gambar 1. Proses CRISP-DM

Tabel 1. Variabel yang Digunakan Dalam Pemodelan Keterlambatan Pegawai

No	Variabel	Tipe Data	Deskripsi
1.	Usia	Numerik	Usia pegawai (tahun)
2.	Masa Kerja	Numerik	Lama bekerja pegawai sejak tahun masuk (tahun)
3.	Status Pernikahan	Kategorik	Status pernikahan pegawai (menikah/lajang /duda/janda)
4.	Jarak Tempat Tinggal	Kategorik	Rentang jarak tempat tinggal ke kantor (kos/rumah/kontrakan) dengan kantor (0-5 km, 6-10 km, 11 – 25 km, >26 km)
5.	Kendaraan ke Kantor	Kategorik	Kendaraan yang digunakan untuk ke kantor (kendaraan pribadi (motor/mobil) atau kendaraan umum/transportasi *online*
6.	Gaya Hidup	Kategorik	Gaya hidup pegawai (seimbang/tidak seimbang)
7.	Status Keterlambatan	Biner (Label)	1 = Terlambat, 0 = Tidak Terlambat

Machine (SVM), dalam menentukan apakah seorang karyawan akan datang terlambat atau tidak dengan menggunakan data historis. Proses penelitian ini dirancang berdasarkan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang mencakup enam langkah utama (Shearer, 2000). Kerangka kerja CRISP-DM dapat dilihat pada **Gambar 1**.

2.2 Sumber dan Jenis Data

Penelitian ini memanfaatkan data sekunder yang diambil dari sistem informasi kepegawaian perusahaan, yang telah mencatat berbagai informasi personal dan profesional pegawai secara terstruktur. Penggunaan data sekunder dalam konteks penelitian SDM dinilai efisien karena memungkinkan analisis berbasis data historis tanpa perlu melalui proses pengumpulan data primer yang memakan waktu dan biaya (Johnston, 2017). Selain itu, sistem informasi kepegawaian umumnya menyajikan data yang konsisten dan terstandarisasi, sehingga cocok untuk keperluan analisis prediktif (Suryani et al., 2020).

Dataset yang digunakan mencakup 1902 data, di mana masing-masing baris mewakili satu individu pegawai. Setiap data mengandung delapan fitur utama yang signifikan berkaitan dengan kecenderungan keterlambatan. Fitur-fitur tersebut mencakup:

1. Usia pegawai (dalam tahun) (Douglas & Roberts, 2020).
2. Masa kerja (lama bekerja di perusahaan) (Puji Lestari & Sinambela, 2021),
3. Status pernikahan (menikah/lajang/duda/janda) (Padmanabhan & Magesh, 2016),
4. Jarak tempat tinggal ke kantor (dalam kilometer) (Borzikowsky et al., 2023) (Flynn, 2017),
5. Jenis kendaraan yang digunakan (mobil, motor, angkutan umum) (Shahikhaneh et al., 2020) (Le & Trinh, 2016) (Salazar-Serna et al., 2023),
6. Gaya hidup pegawai (seimbang/tidak seimbang) (Valery et al., 2023) (Mendis & Weerakkody, 2014),
7. Serta status keterlambatan, yang menjadi variabel target dalam analisis klasifikasi.

Struktur data ini terdiri atas gabungan variabel kategorik dan numerik. Misalnya, status pernikahan dan jenis kendaraan tergolong ke dalam variabel kategorik nominal, sedangkan usia, masa kerja, dan jarak tempat tinggal termasuk dalam variabel numerik kontinu. Karakteristik ini menuntut pendekatan pemrosesan data yang tepat, seperti *one-hot encoding* untuk variabel kategorik dan normalisasi atau standarisasi untuk variabel numerik (Witten et al., 2016).

Secara umum, struktur dan kualitas data yang digunakan telah memenuhi syarat untuk diterapkan dalam analisis klasifikasi menggunakan metode machine learning. Kombinasi antara fitur-fitur demografis, perilaku, dan logistik diyakini memberikan gambaran yang lebih menyeluruh terhadap faktor-faktor yang berperan dalam keterlambatan kerja pegawai (Luksyte et al., 2013) (Mercara, 2020) (De Clercq et al., 2024). **Tabel 1**

Tabel 2. *Dataset* Keterlambatan Pegawai (Sepuluh Data Awal)

No	NIP	Nama	Usia	Masa Kerja	Status Nikah	Jarak Tempat Tinggal	Kendaraan ke Kantor	Gaya Hidup	Status Terlambat
1	1218600XX	Fulan 1	22	2	Lajang	6-10 KM	Kendaraan umum/ transportasi *online*	Seimbang	Terlambat
2	1228600XX	Fulan 2	22	1	Lajang	0-5 KM	Kendaraan pribadi - motor	Seimbang	Tidak terlambat
3	218600XX	Fulan 3	23	2	Lajang	0-5 KM	Kendaraan pribadi - motor	Tidak seimbang	Terlambat
4	218601XX	Fulan 4	23	2	Lajang	0-5 KM	Kendaraan pribadi - motor	Tidak seimbang	Terlambat
5	218602XX	Fulan 5	23	2	Lajang	11-25 KM	Kendaraan pribadi - motor	Seimbang	Terlambat
6	218603XX	Fulan 6	23	2	Lajang	0-5 KM	Kendaraan umum/ transportasi *online*	Tidak seimbang	Terlambat
7	218604XX	Fulan 7	23	2	Lajang	6-10 KM	Kendaraan pribadi - motor	Seimbang	Terlambat
8	218605XX	Fulan 8	23	2	Lajang	11-25 KM	Kendaraan pribadi - motor	Tidak seimbang	Terlambat
9	218606XX	Fulan 9	23	2	Lajang	11-25 KM	Kendaraan pribadi - motor	Tidak seimbang	Terlambat
10	218607XX	Fulan 10	23	2	Lajang	11-25 KM	Kendaraan pribadi - motor	Tidak seimbang	Terlambat

menunjukkan jenis dan deskripsi dari masing-masing variabel yang dianalisis dalam model.

2.3 Tahapan CRISP-DM

Berikut ini adalah tahapan CRISP-DM:

1. *Business Understanding*

Langkah pertama adalah memahami bahwa keterlambatan pegawai tidak hanya sebagai permasalahan suatu individu, tetapi juga merupakan isu strategis yang dapat mempengaruhi kinerja tim dan organisasi. Adapun data yang digunakan adalah data absensi pegawai per hari kerja. Cakupan waktu data yang dianalisis selama 1 tahun, dengan kriteria inklusi berupa pegawai aktif yang memiliki catatan kehadiran lengkap, sedangkan pegawai dengan status kontrak jangka pendek atau data kehadiran yang tidak konsisten dikecualikan dari analisis. Oleh karena itu, tujuan penelitian ini adalah untuk mengembangkan model guna memprediksi keterlambatan pegawai agar Bidang *Human Resource* (HR) dapat melakukan tindakan pencegahan.

2. *Data Understanding*

Eksplorasi data dilakukan untuk memahami pola distribusi dari variabel, mengatasi nilai yang hilang, nilai *outlier*, dan ketidakseimbangan dalam data. Beberapa catatan kehadiran ditemukan memiliki *missing value*, misalnya pada kolom jam masuk atau alasan keterlambatan yang tidak tercatat. Nilai yang hilang ini ditangani dengan dua pendekatan, yakni imputasi menggunakan nilai modus untuk variabel kategorik (misalnya alasan keterlambatan), serta penghapusan *entri* jika data yang kosong dianggap krusial dan tidak dapat direpresentasikan secara tepat.

Selain itu, juga teridentifikasi *outlier* pada variabel jarak tempat tinggal ke kantor, dimana sebagian kecil pegawai tercatat menempuh jarak lebih dari 50 km, jauh di atas rata-rata populasi. Data ekstrim seperti ini dievaluasi lebih lanjut, apabila masih relevan dengan kondisi nyata maka tetap dipertahankan, tetapi jika dinilai mengganggu kestabilan model maka dilakukan *winsorizing* atau penghapusan agar distribusi data lebih proporsional.

Terkait *shift* kerja, variabel ini tidak dimasukkan dalam analisis karena mayoritas pegawai bekerja pada jam reguler, sementara proporsi pegawai dengan pola *shift* relatif kecil dan tidak representatif. Dengan demikian, fokus penelitian diarahkan pada keterlambatan dalam jam kerja normal agar hasil model prediksi lebih konsisten, dapat digeneralisasi, dan relevan dengan mayoritas populasi pegawai.

3. *Data Preparation*

Pada tahap ini dilakukan pembersihan data, transformasi data dan normalisasi data. Nilai yang hilang diisi menggunakan median atau modus untuk variabel kategorik. Variabel kategorik dikonversi ke bentuk numerik dengan menggunakan *One Hot Encoding* atau *Label Encoding*. *Dataset* kemudian dibagi menjadi dua bagian, yaitu data untuk pelatihan (80%) dan data untuk pengujian (20%), sementara variabel numerik dilakukan standarisasi agar sesuai dengan karakteristik algoritma SVM dan *Logistic Regression*.

Dalam penelitian ini, definisi target keterlambatan didefinisikan secara biner, yaitu *1 = terlambat* dan *0 = tidak terlambat*. Ambang batas keterlambatan ditentukan lebih dari jam kerja yang telah ditetapkan, yaitu mengacu pada kebijakan absensi dalam perusahaan. Target ini bersifat kejadian tunggal per hari kerja, dan setiap baris data merepresentasikan satu catatan kumulatif kehadiran pegawai. Penelitian ini menggunakan 1.902 data keterlambatan pegawai. Sebagai ilustrasi, **Tabel 2** menampilkan 10 data awal.

Tabel 3. *Dataset* Keterlambatan Pegawai *Encoding* (Sepuluh Data Awal)

No	NIP	Nama	Usia	Masa Kerja	Status Nikah	Jarak Tempat Tinggal	Kendaraan ke Kantor	Gaya Hidup	Status Terlambat
1	1218600XX	Fulan 1	22	2	1	2	3	1	1
2	1228600XX	Fulan 2	22	1	1	1	1	1	0
3	218600XX	Fulan 3	23	2	1	1	1	2	1
4	218601XX	Fulan 4	23	2	1	1	1	2	1
5	218602XX	Fulan 5	23	2	1	3	1	1	1
6	218603XX	Fulan 6	23	2	1	1	3	2	1
7	218604XX	Fulan 7	23	2	1	2	1	1	1
8	218605XX	Fulan 8	23	2	1	3	1	2	1
9	218606XX	Fulan 9	23	2	1	3	1	2	1
10	218607XX	Fulan 10	23	2	1	3	1	2	1

Keterangan:
1. Status Pernikahan: Lajang = 1, Kawin = 2, Duda/Janda = 3
2. Jarak Tempat Tinggal: 0-5 km = 1, 6-10 km = 2, 11-25 km = 3, >26 km = 4
3. Kendaraan ke Kantor: Motor = 1, Mobil = 2, Umum/*Online* = 3
4. Gaya Hidup: Seimbang = 1, Tidak Seimbang = 2
5. Status Keterlambatan: Tidak Terlambat = 0, Terlambat = 1

Dataset tersebut tidak mengandung *missing value*, dan selanjutnya dilakukan pelabelan pada data kategorik. *Dataset* yang sudah dilakukan pelabelan melalui MS. Excel yang ditunjukkan pada **Tabel 3**.

4. *Modeling*

Pada tahap *modeling* menggunakan tiga algoritma *machine learning* yang diterapkan adalah:

a. *Logistic Regression*: metode statistik untuk memprediksi probabilitas dari kelas biner berdasarkan variabel *input*. *Logistic regression* memodelkan probabilitas bahwa *output* (target) y=1 sebagai fungsi *sigmoid* dari kombinasi linier fitur:

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}} \quad (1)$$

Dimana:
$P(y=1|x)$: probabilitas bahwa *output* bernilai 1 dengan kondisi *input* x.
e : bilangan eksponensial.
β_0 : konstanta (*intercept).*
$\beta_1, \beta_2, \ldots, \beta_n$: koefisien regresi dari variabel bebas.
$x_1, x_2, \ldots, x_n$: variabel independen.

Fungsi *sigmoid* dari kombinasi linier fitur, dan koefisien model dapat langsung diambil setelah pelatihan untuk menuliskan persamaan regresi logistik secara eksplisit.

b. *Random Forest*: metode *ensemble learning* yang didasarkan pada pohon keputusan *(decision tree)*, yang membangun banyak pohon pada *subset* data yang berbeda dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*. Untuk regresi, prediksi akhir dihitung dengan rata-rata dari semua prediksi pohon:

$$\check{y} = \frac{1}{B}\sum_{b=1}^{B} h_b(x) \quad (2)$$

Dimana:
$\check{y}$: prediksi akhir dari model *Random Forest*.
B : jumlah pohon (*trees*) dalam hutan (*forest*).
$h_b(x)$: prediksi dari pohon keputusan ke- b.
x : variabel *input*.

Jika $h_1(x)$, $h_2(x)$,...,$h_B(x)$ adalah prediksi dari masing-masing pohon ke-b maka maka prediksi akhir $\check{y}$ adalah hasil voting mayoritas dari semua pohon $\check{y}$ = mode{$h_1(x)$,$h_2(x)$,…,$h_B(x)$}

c. *Support Vector Machine* (SVM): algoritma klasifikasi yang bekerja dengan mencari *hyperplane* terbaik yang memisahkan dua kelas dalam ruang fitur. Fungsi Keputusan model SVM

$$f(x) = \omega^T x + b \quad (3)$$

Dimana:
$f(x)$: skor keputusan hasil pemisahan kelas.
ω : vektor bobot (normal terhadap *hyperplane*).
x : vektor fitur *input*.
b : bias (*intercept*) yang menentukan posisi *hyperplane*.
Jika $f(x) > 0$, maka data diprediksi kelas **+1**, sedangkan jika $f(x) < 0$ maka diprediksi kelas **-1**.

Dengan f(x) sebagai skor keputusan, tanda positif = kelas 1 dan tanda negatif = kelas -1. Tujuannya adalah memaksimalkan margin antara dua kelas. Semakin kecil $\|\omega\|$, semakin besar margin.

Tabel 4. Data Keterlambatan Pegawai untuk Pengujian (Sepuluh Data Pengujian)

No	NIP	Nama	Usia	Masa Kerja	Status Nikah	Jarak Tempat Tinggal	Kendaraan ke Kantor	Gaya Hidup	Status Terlambat
1	91098616XX	Fulan 756	32	14	2	4	2	1	
2	93128602XX	Fulan 924	30	11	2	4	2	1	
3	91108616XX	Fulan 763	32	13	2	4	2	1	
4	95158600XX	Fulan 1141	28	8	2	4	2	1	
5	89088600XX	Fulan 561	34	15	2	4	2	1	
6	95148620XX	Fulan 1138	28	9	2	4	2	1	
7	95148615XX	Fulan 1133	28	9	2	4	2	1	
8	70928601XX	Fulan 69	53	31	2	1	2	1	
9	94198605XX	Fulan 1112	29	4	2	1	2	1	
10	90168606XX	Fulan 748	33	7	2	1	2	1	

Keterangan:

1. Status Pernikahan: Lajang = 1, Kawin = 2, Duda/Janda = 3
2. Jarak Tempat Tinggal: 0-5 km = 1, 6-10 km = 2, 11-25 km = 3, >26 km = 4
3. Kendaraan ke Kantor: Motor = 1, Mobil = 2, Umum/*Online* = 3
4. Gaya Hidup: Seimbang = 1, Tidak Seimbang = 2
5. Status Keterlambatan: Tidak Terlambat = 0, Terlambat = 1

$$\min_{w,b} \frac{1}{2} ||\omega||^2 \quad (4)$$

Dimana:

$||\omega||$: norma Euclidean dari vektor bobot.

Tujuan dari optimasi ini adalah meminimalkan norma bobot sehingga margin antara dua kelas menjadi maksimal. Semakin kecil $||\omega||$, semakin lebar margin pemisahan kelas yang dihasilkan oleh SVM.

Logistic Regression dipilih karena sering digunakan sebagai *baseline* model dalam prediksi perilaku karyawan dan menunjukkan performa yang kompetitif di berbagai studi klasifikasi SDM (Mohbey, 2020) (Nayak & Palai, 2023). *Random Forest* digunakan karena terbukti menjadi algoritma terbaik dalam memprediksi perilaku pegawai seperti attrition dan keterlambatan dengan akurasi dan F1-score yang tinggi (Sujatha & Dhivya, 2021) (Murzaeva et al., 2024). SVM digunakan karena menunjukkan performa kompetitif terutama pada *dataset* yang tidak linear (Gandhi et al., 2020) (Aditya et al., 2024).

Seluruh model dilatih dengan menggunakan *data training* yang sebelumnya telah dilakukan penyeimbangan data menggunakan teknik *SMOTE* untuk selanjutnya dilakukan uji dengan menggunakan *data testing.* Parameter awal diterapkan sesuai dengan standar literatur (Pedregosa et al., 2011), dan penyesuaian dilakukan menggunakan metode *Grid Search* atau *Randomized Search* pada data pelatihan dengan teknik validasi silang *5-fold* untuk meningkatkan performa dan menghindari *overfitting*. Adapun *tools* yang digunakan adalah Python 3.13.5.

5. *Evaluation*

Evaluasi terhadap model dilaksanakan dengan memanfaatkan metrik klasifikasi, yaitu akurasi, presisi, *recall,* F1 *score,* dan ROC-AUC. Akurasi digunakan untuk melihat seberapa banyak prediksi yang benar. Presisi dan *recall* digunakan untuk menilai kinerja model dalam menangani kelas minoritas (pegawai yang terlambat). F1-*score* digunakan untuk mengukur keseimbangan antara presisi dan *recall* dalam sebuah model klasifikasi. ROC-AUC digunakan untuk mengukur kemampuan model klasifikasi dalam membedakan antara dua kelas, Model terbaik dipilih berdasarkan keseimbangan metrik performa, bukan hanya akurasi tertinggi saja. Hal ini dilakukan untuk memastikan bahwa hasil klasifikasi relevan secara praktis bagi SDM dalam mendeteksi risiko keterlambatan secara akurat.

6. *Implementation*

Pada penelitian ini hanya dilakukan sampai pada tahap evaluasi. Untuk penelitian selanjutnya, dapat dilakukan implementasi model ke dalam sistem SDM organisasi. Model yang optimal dipersiapkan untuk diintegrasikan ke dalam sistem absensi atau dasbor bidang SDM sebagai elemen dari sistem peringatan dini. Rekomendasi disusun untuk mendukung tindakan manajerial yang bersifat pencegahan dan berdasarkan data.

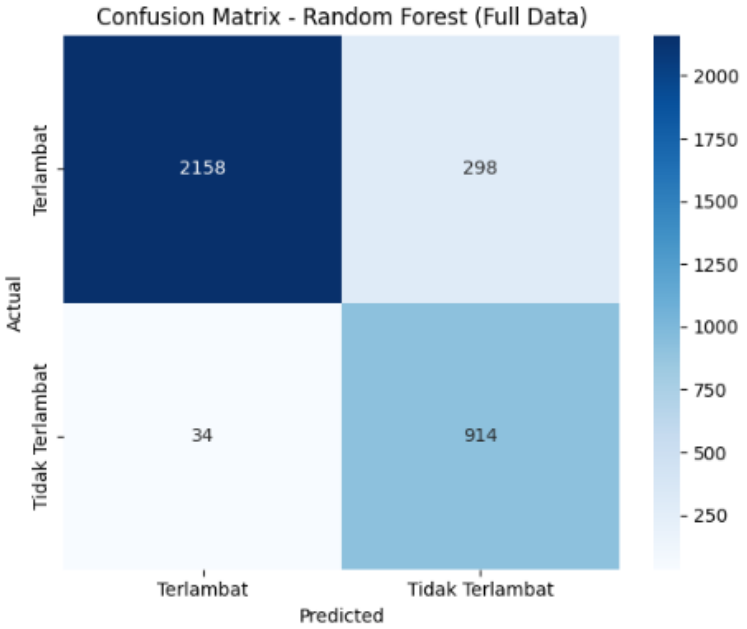
3. Hasil dan Pembahasan

Setelah melaksanakan pelatihan model dengan menggunakan data keterlambatan pegawai berjumlah 1902 data, diperoleh hasil evaluasi kinerja dari setiap algoritma yang didasarkan pada lima metrik utama: akurasi, presisi, *recall*, F1-*score*, dan ROC-AUC. Sebagai ilustrasi, berikut 10 data yang digunakan untuk pengujian, yang dapat dilihat pada **Tabel 4**. Selain itu data kinerja setiap model dapat dilihat pada **Gambar 3**.

Studi ini mengevaluasi efektivitas tiga algoritma model *machine learning* yaitu, *Logistic Regression*, *Random Forest,* dan *Support Vector Machine* (SVM), dalam memprediksi keterlambatan pegawai. Hasil dari evaluasi awal menunjukkan bahwa

```
=== Logistic Regression ===
              precision    recall  f1-score   support

           0       0.95      0.72      0.82       265
           1       0.47      0.86      0.60        76

    accuracy                           0.75       341
   macro avg       0.71      0.79      0.71       341
weighted avg       0.84      0.75      0.77       341

=== SVM ===
              precision    recall  f1-score   support

           0       0.94      0.74      0.82       265
           1       0.47      0.83      0.60        76

    accuracy                           0.76       341
   macro avg       0.71      0.78      0.71       341
weighted avg       0.83      0.76      0.78       341

=== Random Forest ===
              precision    recall  f1-score   support

           0       0.93      0.84      0.88       265
           1       0.58      0.76      0.66        76

    accuracy                           0.82       341
   macro avg       0.75      0.80      0.77       341
weighted avg       0.85      0.82      0.83       341
```

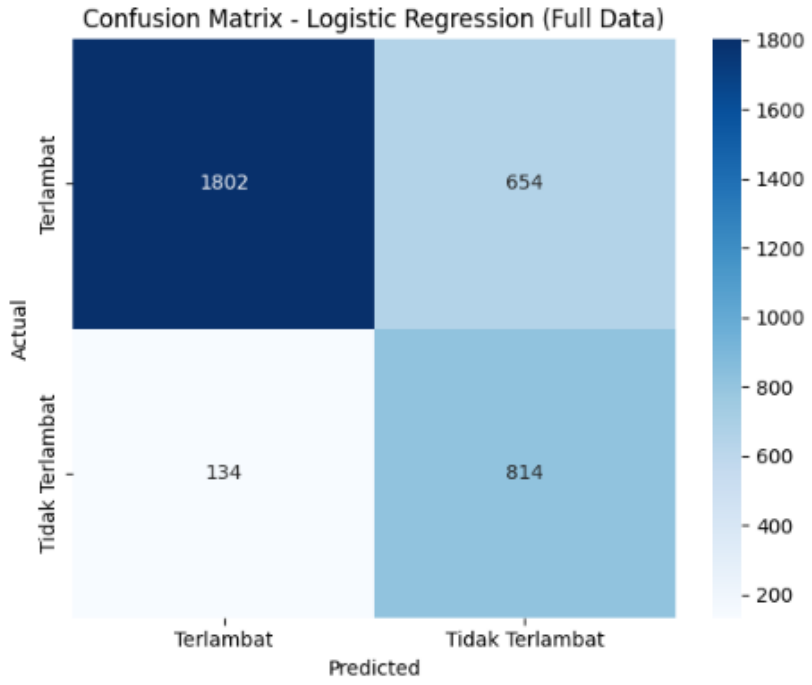
Gambar 3. Hasil Evaluasi Kinerja Setiap Model



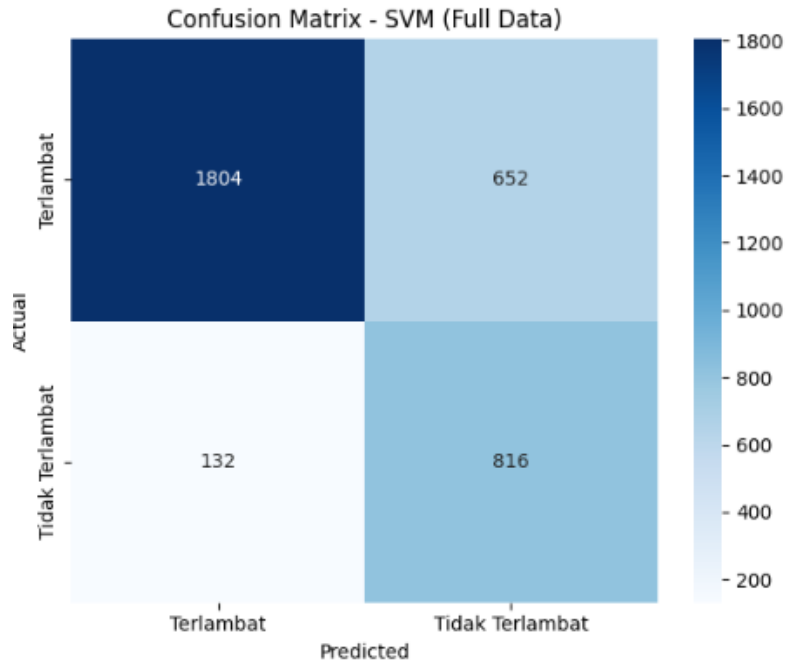
Gambar 4. *Confusion Matrix* Model *Random Forest*

Random Forest merupakan model dengan performa terbaik secara keseluruhan. *Model Random Forest* memberikan hasil terbaik di antara ketiga model, dengan akurasi tertinggi sebesar 82%. Hal ini menunjukkan bahwa model ini tidak hanya signifikan dalam mendeteksi keterlambatan, tetapi juga lebih bijak dalam membedakan pegawai yang benar-benar hadir tepat waktu. Pada model ini, untuk kelas "tidak terlambat" menunjukkan *precision* sebesar 0.58, *recall* sebesar 0.76, dan F1-score sebesar 0.66. Performa ini menunjukkan bahwa *Random Forest* mampu mengidentifikasi pegawai yang tidak terlambat secara lebih akurat dan seimbang dibanding dua model lainnya. *Random Forest* tidak hanya mengenali lebih banyak pegawai tidak terlambat, tetapi juga lebih tepat dalam memprediksi siapa saja mereka, hal ini juga ditunjukan oleh *confusion matrix* pada **Gambar 4**.

Model *Logistic Regression* menunjukkan performa akurasi sebesar 75%, dengan kemampuan mengenali kelas "tidak terlambat" cukup baik ditunjukkan oleh nilai recall yang tinggi sebesar 0.86. Namun, model ini memiliki *precision* rendah sebesar 0.47, yang berarti banyak prediksi "tidak terlambat" yang ternyata salah. Hal ini menyebabkan *F1-score*



Gambar 5. *Confusion Matrix* Model *Logistic Regression*



Gambar 6. *Confusion Matrix* Model SVM

untuk kelas tersebut hanya mencapai 0.60, yang menunjukkan adanya *trade-off* antara sensitivitas dan ketepatan, sesuai dengan yang ditunjukkan *confusion matrix* pada **Gambar 5**.

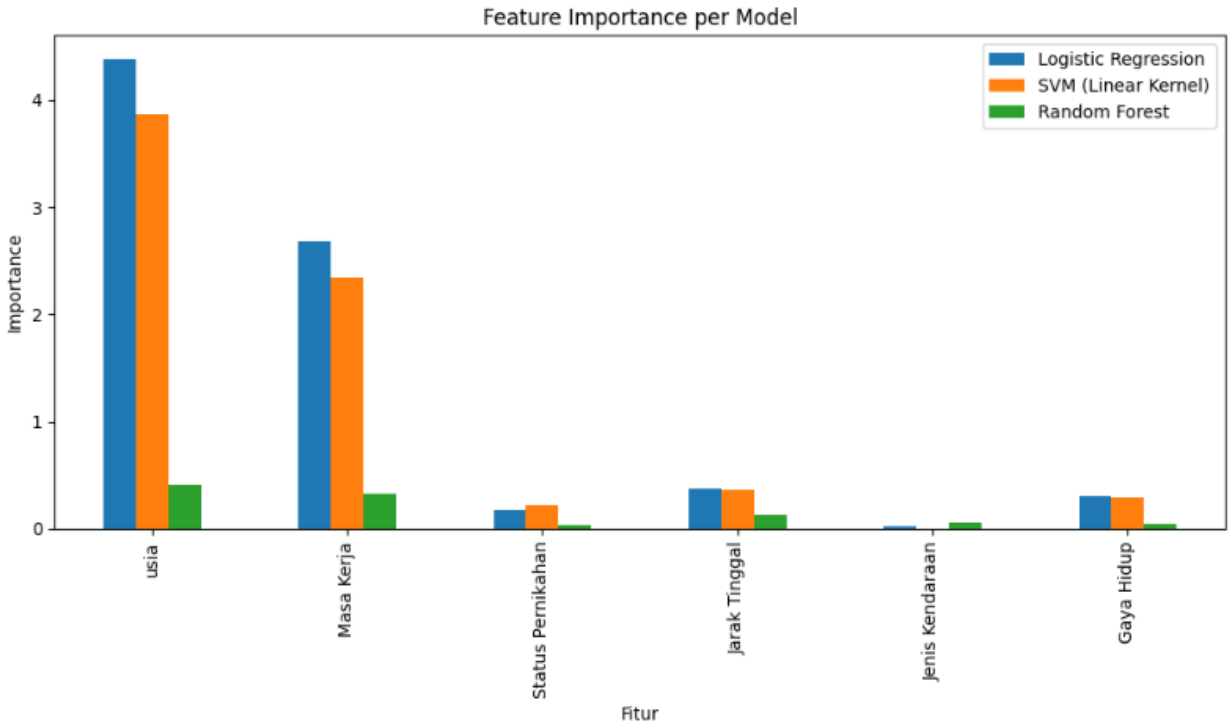
Model SVM memiliki akurasi sedikit lebih tinggi yaitu 76%**.** Performa SVM hampir identik dengan *Logistic Regression*, dengan *recall* 0.83 dan *precision* 0.47 pada kelas "tidak terlambat", menghasilkan F1-score sebesar 0.60. Ini menunjukkan bahwa SVM memiliki kecenderungan serupa, yaitu sensitif terhadap kelas minoritas namun belum akurat sepenuhnya. Hal ini dapat terlihat pada *confusion matrix* SVM pada **Gambar 6**.

Secara keseluruhan, evaluasi menunjukkan bahwa *Random Forest* paling unggul dari sisi akurasi dan kekuatan klasifikasi, sementara *Logistic Regression* menawarkan kinerja yang kompeten dengan kelebihan interpretabilitas yang tinggi. SVM, dalam konteks ini, tidak direkomendasikan karena performanya yang jauh di bawah dua model lainnya. Sehingga, dalam penelitian ini model *Random Forest* dipilih sebagai model terbaik untuk memprediksi keterlambatan pegawai.

Selanjutnya dilakukan analisis lebih lanjut untuk mengetahui variabel yang paling berpengaruh dalam memprediksi keterlambatan pegawai *(feature importance)*. Berikut ini merupakan hasil *feature*

	Fitur	Logistic Regression	SVM (Linear Kernel)	Random Forest
0	usia	4.383879	3.864333	0.408222
1	Masa Kerja	2.684140	2.350010	0.327595
2	Status Pernikahan	0.177629	0.216661	0.034008
3	Jarak Tinggal	0.373952	0.359701	0.130045
4	Jenis Kendaraan	0.024118	0.000012	0.057755
5	Gaya Hidup	0.303017	0.289546	0.042376

Gambar 7. Hasil *Feature Importance*



Gambar 8. Visualisasi *Feature Importance*

importance yang ditunjukkan dalam **Gambar 7** dan **Gambar 8.**

Sehingga, dapat diketahui bahwa usia pegawai merupakan faktor yang paling berpengaruh dalam model. Variabel ini secara konsisten memperoleh skor tertinggi di seluruh metode, dengan nilai *Logistik Regression* sebesar 4.38, SVM 3.86 dan skor *Random Forest* mencapai 0.488. Hal ini mengindikasikan bahwa usia memiliki hubungan yang kuat terhadap kecenderungan keterlambatan. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa usia berkorelasi dengan stabilitas kerja, ritme hidup, dan kecenderungan terhadap disiplin kerja (Islam et al., 2018) (Gao et al., 2019).

Selanjutnya, masa kerja menempati urutan kedua dengan skor *Random Forest* sebesar 0.327, menunjukkan bahwa masa kerja juga berperan penting dalam membentuk pola keterlambatan. Pegawai dengan masa kerja yang lebih singkat atau terlalu lama kemungkinan memiliki dinamika motivasi dan adaptasi yang berbeda, yang kemudian mempengaruhi perilaku kehadiran. Masa kerja yang lebih lama dapat mencerminkan loyalitas dan adaptasi terhadap sistem organisasi, namun juga bisa menimbulkan kejenuhan yang berpengaruh terhadap motivasi dan kedisiplinan. Studi oleh (Sari & Lhaksmana, 2022) mendukung pentingnya masa kerja dalam memodelkan perilaku ketenagakerjaan, termasuk keterlambatan atau *turnover*.

Fitur-fitur lain seperti jarak rumah ke kantor dan jenis kendaraan menunjukkan kontribusi sedang. Jarak rumah ke kantor memperoleh skor *Random Forest* sebesar 0.13, sedangkan jenis kendaraan mencatat nilai 0.05. Meskipun tidak sebesar dua variabel sebelumnya, keduanya tetap relevan secara praktis karena terkait dengan aksesibilitas dan kemungkinan gangguan eksternal yang dapat mempengaruhi ketepatan waktu pegawai. Studi dari (Islam et al., 2018) juga menekankan bahwa faktor logistik seperti jarak tempuh dan moda transportasi dapat meningkatkan akurasi prediksi bila dimasukkan ke dalam model *Random Forest* berbasis seleksi fitur.

Sebaliknya, fitur seperti gaya hidup dan status pernikahan menunjukkan kontribusi yang lebih rendah dalam model prediktif, dengan skor *Random Forest* masing-masing sebesar 0.04 dan 0.03. Meskipun secara statistik pengaruhnya terbatas, kedua fitur ini tetap dapat dipertimbangkan sebagai faktor

pendukung dalam pendekatan manajerial yang lebih holistik, terutama jika dikaitkan dengan kesejahteraan karyawan.

Secara keseluruhan, hasil ini menegaskan bahwa kombinasi antara variabel demografis dan logistik memberikan pengaruh signifikan terhadap risiko keterlambatan pegawai. Informasi ini dapat dimanfaatkan oleh bagian sumber daya manusia untuk merancang kebijakan yang lebih adaptif, misalnya dengan intervensi berbasis usia, atau kebijakan fleksibilitas bagi pegawai dengan jarak tempuh tertentu.

4. Kesimpulan

Studi ini berhasil membuktikan bahwa algoritma *machine learning* dapat digunakan secara efektif untuk memprediksi keterlambatan pegawai berdasarkan informasi keterlambatan dan data demografis pegawai yang mencakup usia, lama bekerja, status pernikahan, jarak tempat tinggal, jenis kendaraan, dan gaya hidup. Tiga model yang dianalisis *Logistic Regression*, *Support Vector Machine* (SVM), dan *Random Forest* menunjukkan kinerja yang memadai untuk dijadikan sebagai sistem prediksi awal.

Hasil evaluasi menunjukkan bahwa *Random Forest* merupakan model dengan performa terbaik. Model ini juga menunjukkan kestabilan performa dan kemampuan generalisasi yang baik. Analisis lebih lanjut terhadap *feature importance* mengungkap bahwa usia dan masa kerja merupakan dua faktor dominan, yang disusul oleh faktor jarak tempuh ke kantor dan jenis kendaraan. Sementara itu, fitur seperti status perkawinan dan gaya hidup memiliki kontribusi yang lebih kecil, namun tetap relevan dalam konteks personalisasi kebijakan *Human Resource* (HR).

Secara keseluruhan, penelitian ini menunjukkan bahwa pemanfaatan algoritma pembelajaran mesin tidak hanya dapat meningkatkan efisiensi dalam proses SDM, tetapi juga membuka ruang baru untuk memahami perilaku pegawai secara lebih objektif dan berbasis data.

Model prediksi keterlambatan berbasis *Random Forest* yang telah dibangun dalam penelitian ini, selanjutnya dapat dimanfaatkan sebagai alat bantu dalam sistem kehadiran digital untuk mendukung pengambilan keputusan di bidang SDM. Pengembangan ke depan sebaiknya mencakup perluasan variabel, termasuk aspek psikologis dan perilaku, serta pembaruan model secara berkala untuk menjaga akurasi. Selain itu, temuan mengenai pentingnya usia dan masa kerja dapat menjadi dasar penyusunan kebijakan berbasis segmen, seperti *mentoring* karyawan baru atau pengaturan waktu kerja yang lebih fleksibel. Pendekatan ini diharapkan dapat membuat strategi manajemen kehadiran lebih responsif, personal, dan efektif.

5. Daftar Pustaka

Abdulrahman Aspita Ishiaka. (2025). Assessing The Impact Of Employee Discipline On Organizational Survival: The Human Resource Management Perspective. *Journal of Human, Social and Political Science Research*, *7*(6 SE-Articles). https://doi.org/10.70382/sjhspsr.v7i6.018

Aditya, M. R., Sutanto, T., Budiman, H., Ridha, M. R. N., Syapotro, U., & Azijah, N. (2024). Machine Learning Models for Classification of Anemia from CBC Results: Random Forest, SVM, and Logistic Regression. *Journal of Data Science*, *2024*(SE-Articles). https://iuojs.intimal.edu.my/index.php/jods/article/view/589

Angulakshmi, M., Madhumithaa, N., Dokku, S., Pachar, S., Sneha, K., & Lenin, S. (2024). Predictive HR Analytics: Using Machine Learning to Forecast Workforce Trends and Needs. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, *7*, 1399–1405. https://doi.org/10.1109/IC3I61595.2024.10829013

Beauty, C., & Thomas, K. T. (2021). *Disciplinary Procedures, Employee Punctuality and Employee Performance at Ndola City Council (Zambia)*. https://api.semanticscholar.org/CorpusID:235503456

Borzikowsky, C., Raimer, S., & Kowalski, J. (2023). New work in modern times: predicting employees' choice to work from home. *Facilities*, *41*(13/14), 855–867. https://doi.org/10.1108/F-04-2023-0037

De Clercq, D., Aboramadan, M., & Kundi, Y. M. (2024). How employee pandemic fears may escalate into a lateness attitude, and how a safe organizational climate can mitigate this challenge. *Personnel Review*, *53*(4), 1039–1058. https://doi.org/10.1108/PR-11-2022-0764

Douglas, S., & Roberts, R. (2020). Employee age and the impact on work engagement. *Strategic HR Review*, *19*(5), 209–213. https://doi.org/10.1108/SHR-05-2020-0049

Flynn, J. J. (2017). *How commuting affects employees' Wellbeing and Work-Life Balance: The perspective of full-time employees commuting within the greater Dublin area*. https://api.semanticscholar.org/CorpusID:134524321

Gandhi, P. N., Mhaske, G. S., Jangale, A., & Kadlag, A. (2020). Employee Attrition Prediction using Machine Learning. *Journal of Emerging Technologies and Innovative Research*. https://api.semanticscholar.org/CorpusID:262131418

Gao, X., Wen, J., & Zhang, C. (2019). An Improved Random Forest Algorithm for Predicting Employee Turnover. *Mathematical Problems in Engineering*, *2019*(1), 4140707. https://doi.org/https://doi.org/10.1155/2019/4140707

Islam, M. K., Alam, M. M., Islam, M. B., Mohiuddin, K., Das, A. K., & Kaonain, M. S. (2018). *An Adaptive Feature Dimensionality Reduction*

Technique Based on Random Forest on Employee Turnover Prediction Model BT - Advances in Computing and Data Sciences (M. Singh, P. K. Gupta, V. Tyagi, J. Flusser, & T. Ören (eds.); pp. 269–278). Springer Singapore.

Kushwaha, P. K., Rana, A., Srivastava, S., Saifi, A., Tavish, A., & Chaturvedi, P. (2023). Employee Absenteeism Prediction Using Machine Learning. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, *10*, 116–121. https://doi.org/10.1109/UPCON59197.2023.10434342

Le, T. P. L., & Trinh, T. A. (2016). Encouraging Public Transport Use to Reduce Traffic Congestion and Air Pollutant: A Case Study of Ho Chi Minh City, Vietnam. *Procedia Engineering*, *142*, 236–243. https://doi.org/https://doi.org/10.1016/j.proeng.2016.02.037

Luksyte, A., Waite, E., Avery, D. R., & Roy, R. (2013). Held to a different standard: Racial differences in the impact of lateness on advancement opportunity. *Journal of Occupational and Organizational Psychology*, *86*(2), 142–165. https://doi.org/https://doi.org/10.1111/joop.12010

Maharjan, R. (2022). *Employee Churn Prediction using Logistic Regression and Support Vector Machine*. https://api.semanticscholar.org/CorpusID:261723658

Mendis, M. V. S., & Weerakkody, W. A. S. (2014). *The Relationship between Work Life Balance and Employee Performance: With Reference to Telecommunication Industry of Sri Lanka*. https://api.semanticscholar.org/CorpusID:113705959

Mercara, J. L. D. (2020). Prediction of Employees' Lateness Determinants using Machine Learning Algorithms. *International Journal of Advanced Trends in Computer Science and Engineering*. https://api.semanticscholar.org/CorpusID:215941567

Mohbey, K. K. (2020). *Employee's Attrition Prediction Using Machine Learning Approaches*. https://api.semanticscholar.org/CorpusID:226580682

Murzaeva, A., Illik, V., & Koc, K. Y. (2024). Employee Turnover Prediction on Synthetic and Real Datasets. *2024 9th International Conference on Computer Science and Engineering (UBMK)*, 1–5. https://doi.org/10.1109/UBMK63289.2024.10773575

Mutiara, S., & Candra, A. (2024). Pengaruh Motivasi Kerja Dan Disiplin Kerja Terhadap Kinerja Karyawan Di Bpjs Kesehatan Cabang Bandar Lampung. *Journal of Management : Small and Medium Enterprises (SMEs)*, *17*(3 SE-Articles). https://doi.org/10.35508/jom.v17i3.19035

Nayak, S., & Palai, P. (2023). Employee Attrition System Prediction using Random Forest Classifier. *International Journal of Computer and Communication Technology*. https://api.semanticscholar.org/CorpusID:261021246

Odionu, C. S., Bristol-Alagbariya, B., & Okon, R. (2024). Data-driven decision making in human resources to optimize talent acquisition and retention. *International Journal of Scholarly Research and Reviews*. https://api.semanticscholar.org/CorpusID:274988834

Padmanabhan, L., & Magesh, D. R. (2016). Difference between Employees Marital Status and Performance Level in IT Industry. *Imperial Journal of Interdisciplinary Research*, *2*. https://api.semanticscholar.org/CorpusID:54757538

Patil, J., & Jadhav, P. (2023). Predicting HR Churn with Python and Machine Learning. *Journal of Advanced Zoology*. https://doi.org/10.53555/jaz.v44is8.3526

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Puji Lestari, U., & Sinambela, E. A. (2021). The Role Of Tenure And Incentives On Employee Behavior. *Current Advanced Research On Sharia Finance And Economic Worldwide*, *1*(1 SE-Articles), 12–17. https://doi.org/10.55047/cashflow.v1i1.17

Reznychenko, T., Uglickich, E., & Nagy, I. (2024). Accuracy Comparison of Logistic Regression, Random Forest, and Neural Networks Applied to Real MaaS Data. *2024 Smart City Symposium Prague (SCSP)*, 1–5. https://doi.org/10.1109/SCSP61506.2024.10552715

Rosett, C. M., & Hagerty, A. (2021). *Analytics About Employees BT - Introducing HR Analytics with Machine Learning: Empowering Practitioners, Psychologists, and Organizations* (C. M. Rosett & A. Hagerty (eds.); pp. 7–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-67626-1_2

Salazar-Serna, K., Cadavid, L., Franco, C. J., & Carley, K. M. (2023). *Simulating Transport Mode Choices in Developing Countries BT - Social, Cultural, and Behavioral Modeling* (R. Thomson, S. Al-khateeb, A. Burger, P. Park, & A. A. Pyke (eds.); pp. 209–218). Springer Nature Switzerland.

Sari, S. F., & Lhaksmana, K. M. (2022). Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification. *Journal of Computer System*

and Informatics (JoSYC). https://api.semanticscholar.org/CorpusID:255684494

Shahikhaneh, A., Azari, K. A., & Aghayan, I. (2020). Modeling the Transport Mode Choice Behavior of Motorcyclists. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *44*(1), 175–184. https://doi.org/10.1007/s40996-019-00236-4

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13–22.

Sujaini, H. (2020). Image Classification of Tourist Attractions with K-Nearest Neighbor, Logistic Regression, Random Forest, and Support Vector Machine. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(6 SE-Articles), 2207–2212. https://doi.org/10.18517/ijaseit.10.6.9098

Sujatha, P., & Dhivya, R. S. (2021). *Qualitative Assessment of Machine Learning Classifiers for Employee Performance Prediction BT - Intelligent Computing and Innovation on Data Science* (S.-L. Peng, S.-Y. Hsieh, S. Gopalakrishnan, & B. Duraisamy (eds.); pp. 339–349). Springer Nature Singapore.

Valery, M. B., Santati, P., & Hadjri, M. I. (2023). The Influence of Work-Life Balance on Employee Performance: (Empirical Study at Telkomsel Regional Sumbagsel Office). *JPIM (Jurnal Penelitian Ilmu Manajemen)*, *8*(2 SE-), 208–217. https://doi.org/10.30736/jpim.v8i2.1601

Zupančič, P., & Panov, P. (2024). Predicting Employee Absence from Historical Absence Profiles with Machine Learning. In *Applied Sciences* (Vol. 14, Issue 16). https://doi.org/10.3390/app14167037