

Perbandingan *Logistic Regression*, *Random Forest*, *XGBoost* untuk Prediksi *Employee Attrition* dengan *SMOTE-Pipeline* dan *SHAP*

Gaizka Wisnu Prawira^{*1}, Roni Andarsyah², Muhammad Yusril Helmi Setyawan³

^{1,2,3}Sarjana Terapan Teknik Informatika, Universitas Logistik dan Bisnis Internasional, Indonesia
Email: ¹714220011@std.ulbi.ac.id, ²roniandarsyah@ulbi.ac.id, ³yusrilhelmi@ulbi.ac.id

Abstrak

Tingkat keluar-masuk karyawan (*employee attrition*) merupakan tantangan utama dalam manajemen sumber daya manusia karena berdampak langsung pada biaya operasional dan stabilitas organisasi. Penelitian ini bertujuan membandingkan performa tiga algoritma *machine learning* yaitu *Logistic Regression*, *Random Forest*, dan *XGBoost* dalam memprediksi *attrition* karyawan berdasarkan 1.057 data ulasan karyawan di Indonesia (*employee reviews*) beserta atribut skor kepuasan kerja. Ketidakseimbangan kelas pada label target ditangani menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)* yang digabungkan dengan *StandardScaler* dalam satu *pipeline* untuk mencegah kebocoran data selama validasi silang. Optimasi *hyperparameter* dilakukan dengan *RandomizedSearchCV* dan validasi *5-fold stratified*, serta pengujian stabilitas melalui *repeated stratified k-fold* sebanyak 50 kali pengulangan. Evaluasi pada data uji dilakukan dengan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Matthews Correlation Coefficient*, dan *ROC-AUC*. Hasil menunjukkan *XGBoost* memberikan *ROC-AUC* tertinggi (0,598), diikuti *Random Forest* (0,573) dan *Logistic Regression* (0,564). Analisis *feature importance* dan *SHAP (SHapley Additive exPlanations)* secara konsisten mengidentifikasi fitur *firm*, *years_working*, *overall*, dan *management* sebagai faktor paling dominan dalam keputusan *attrition*. Nilai *SHAP* yang relatif kecil menunjukkan sinyal prediktif dari data yang tersedia masih lemah, sehingga akurasi model belum optimal meskipun pemodelan telah dilakukan secara sistematis. Temuan ini merekomendasikan penguatan program retensi pada aspek kepuasan kerja, manajemen, dan masa kerja karyawan. Secara implikatif, penelitian ini berkontribusi pada bidang pendidikan dan teknologi dengan menunjukkan penerapan praktis teknik pembelajaran mesin (*SMOTE-Pipeline* dan *SHAP*) yang dapat diadopsi sebagai bahan ajar analitik data dan sistem pendukung keputusan berbasis kecerdasan buatan di lingkungan pendidikan vokasi maupun industri.

Kata kunci: *Pengunduran Diri Karyawan, Feature Importance, Logistic Regression, Random Forest, SHAP, XGBoost*

A Comparison of Logistic Regression, Random Forest, and XGBoost for Predicting Employee Turnover Rates Using the SMOTE Pipeline and SHAP

Abstract

Employee attrition is a major challenge in human resource management because it directly affects operational costs and organizational stability. This study aims to compare the performance of three machine learning algorithms, namely Logistic Regression, Random Forest, and XGBoost, in predicting employee attrition based on 1,057 employee review records in Indonesia along with job satisfaction score attributes. Class imbalance in the target label was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) combined with StandardScaler within a single pipeline to prevent data leakage during cross-validation. Hyperparameter optimization was performed using RandomizedSearchCV with 5-fold stratified validation, and stability testing was conducted through repeated stratified k-fold cross-validation with 50 repetitions. Evaluation on the test data was carried out using Accuracy, Precision, Recall, F1-Score, Matthews Correlation Coefficient, and ROC-AUC metrics. The results show that XGBoost achieved the highest ROC-AUC (0.598), followed by Random Forest (0.573) and Logistic Regression (0.564). Feature importance and SHAP (SHapley Additive exPlanations) analyses consistently identified firm, years_working, overall, and management as the most dominant factors in attrition decisions. The relatively small SHAP values indicate that the predictive signal from the available data remains weak, so model accuracy has not yet reached an optimal level despite systematic modeling efforts. These findings recommend strengthening retention programs focused on job satisfaction, management quality, and employee tenure. These findings also have implications for education and technology, as the applied SMOTE-Pipeline and SHAP techniques can serve as practical teaching material for data analytics and as a foundation for developing AI-based decision support systems in vocational education and industry.

Keywords: *Employee Attrition, Feature Importance, Logistic Regression, Random Forest, SHAP, XGBoost*

1. PENDAHULUAN

Employee attrition atau keluarnya karyawan dari suatu organisasi menjadi salah satu masalah krusial dalam manajemen sumber daya manusia (HR) modern. Keluarnya karyawan, baik secara sukarela maupun tidak, menimbulkan biaya rekrutmen dan pelatihan ulang, menurunkan produktivitas tim, serta berpotensi mengganggu kesinambungan budaya organisasi [1], [2]. Oleh karena itu, kemampuan memprediksi risiko *attrition* secara dini menjadi kebutuhan strategis bagi perusahaan agar dapat mengambil langkah intervensi yang tepat sasaran sebelum karyawan benar-benar memutuskan untuk keluar [3].

Urgensi mitigasi *attrition* semakin tinggi apabila ditinjau dari kondisi pasar tenaga kerja di Indonesia, di mana ketersediaan lowongan kerja masih jauh tertinggal dibanding jumlah pencari kerja. Data Badan Pusat Statistik (BPS) mencatat bahwa pada tahun 2024 terdapat sekitar 909 ribu pencari kerja terdaftar, sementara lowongan kerja yang tersedia hanya sebanyak 630 ribu, sehingga terdapat kesenjangan sekitar 279 ribu pencari kerja yang belum terserap [4]. Meski jumlah lowongan kerja cenderung meningkat dari tahun ke tahun, laju peningkatannya tidak sebanding dengan pertumbuhan jumlah pencari kerja baru [4]. Kondisi ketimpangan pasar kerja ini menegaskan bahwa kehilangan karyawan berpengalaman melalui *attrition* menjadi persoalan yang lebih kritis bagi organisasi, karena posisi yang ditinggalkan tidak selalu dapat digantikan secara cepat maupun dengan kualitas sumber daya manusia yang setara.

Kondisi ini menempatkan strategi retensi karyawan bukan sekadar upaya efisiensi biaya rekrutmen, melainkan sebagai langkah strategis di tengah pasar kerja yang kompetitif. Oleh karena itu, kemampuan organisasi dalam mengidentifikasi secara dini karyawan yang berisiko *resign* menjadi semakin relevan, sehingga intervensi retensi dapat difokuskan pada segmen karyawan berisiko tinggi sebelum kekosongan posisi tersebut berdampak pada produktivitas dan keberlangsungan operasional perusahaan.

Pendekatan konvensional dalam analisis *attrition* umumnya mengandalkan statistik deskriptif atau model regresi sederhana yang hanya mampu menangkap hubungan linear antar variabel. Pendekatan tersebut memiliki keterbatasan dalam menangkap pola non-linear dan interaksi kompleks antar fitur, misalnya kombinasi antara masa kerja, kepuasan kerja, dan kualitas manajemen terhadap keputusan resign [5]. Perkembangan teknik *machine learning*, seperti *Logistic Regression*, *Random Forest* [6], dan *XGBoost* [7], menawarkan kemampuan lebih baik dalam menangkap pola-pola tersebut, namun performa masing-masing algoritma sangat bergantung pada karakteristik data yang digunakan sehingga perbandingan sistematis antar model tetap diperlukan [1], [6].

Permasalahan lain yang sering diabaikan pada penelitian prediksi *attrition* adalah ketidakseimbangan kelas (*class imbalance*), di mana jumlah karyawan yang bertahan jauh lebih banyak dibanding yang *resign* [8]. Kondisi ini membuat model cenderung bias terhadap kelas mayoritas apabila tidak ditangani dengan tepat. Selain itu, penerapan teknik penyeimbangan data seperti *SMOTE* yang dilakukan sebelum proses validasi silang berpotensi menimbulkan kebocoran data (*data leakage*), yaitu informasi dari data uji ikut memengaruhi proses pelatihan sehingga hasil evaluasi menjadi terlalu optimis dan tidak mencerminkan performa sesungguhnya di dunia nyata [8], [9].

Tabel 1. Gap dan Novelty

Atribut	Yang Sudah Diketahui	Gap yang Tersisa	Posisi Penelitian Ini
Klasifikasi <i>attrition</i> berbasis <i>machine learning</i>	Random Forest dan XGBoost terbukti unggul menangkap pola non-linear dibanding regresi logistik biasa.	Perbandingan antar algoritma sering dilakukan tanpa validasi statistik berlapis dan sering hanya melaporkan satu model terbaik.	Membandingkan tiga arsitektur (linear, bagging, boosting) secara sistematis dengan tuning, validasi silang berulang, dan beberapa strategi threshold.
Penanganan ketidakseimbangan kelas (SMOTE)	SMOTE efektif meningkatkan recall kelas minoritas pada data <i>attrition</i> yang timpang.	SMOTE sering diterapkan di luar skema validasi silang sehingga berisiko menimbulkan kebocoran data dan performa yang terlalu optimis.	Mengintegrasikan SMOTE dan StandardScaler ke dalam satu pipeline agar resampling hanya “melihat” data latih pada setiap fold.
Interpretabilitas model (Explainable AI)	SHAP mampu menjelaskan kontribusi fitur secara global maupun lokal pada model black-box.	Interpretasi umumnya hanya dihitung untuk satu model sehingga rawan bias terhadap karakteristik algoritma tertentu.	Menghitung SHAP pada tiga model berbeda arsitektur sekaligus menguji konsistensi peringkat fitur penting antar model.

Selain akurasi prediksi, interpretabilitas model menjadi aspek yang tidak kalah penting, khususnya bagi pengambil keputusan di bidang HR [6], [10]. Model *black-box* seperti *Random Forest* dan *XGBoost* sulit dijelaskan secara langsung, sehingga diperlukan pendekatan eksplainabilitas seperti *SHAP* (*SHapley Additive exPlanations*) [6], [11] untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap prediksi, baik secara global maupun pada level individu (lokal), agar hasil model dapat diterjemahkan menjadi rekomendasi strategi retensi karyawan yang konkret [7], [12].

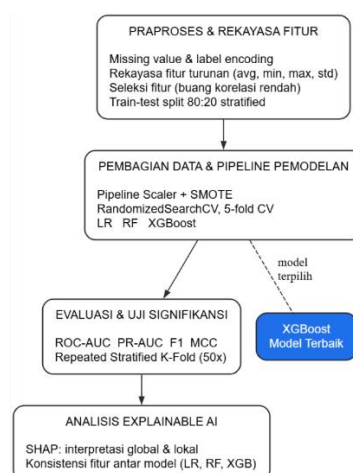
Penelitian terdahulu menunjukkan bahwa masih terdapat beberapa celah penelitian yang perlu dijawab. Pertama, penerapan pendekatan *Explainable AI* (XAI) seperti *SHAP* dan *LIME* masih sangat terbatas [3], [10], sehingga model prediksi *attrition* cenderung tetap bersifat *black-box* bagi pengambil keputusan HR. Kedua, penanganan ketidakseimbangan kelas (*class imbalance*) melalui teknik seperti *SMOTE* masih diterapkan secara tidak konsisten antar studi, tanpa jaminan metodologis yang eksplisit terhadap risiko kebocoran data (*data leakage*) akibat *resampling* yang dilakukan di luar skema validasi silang. Ketiga, sebagian besar studi hanya melaporkan satu algoritma terbaik tanpa menguji konsistensi interpretasi fitur penting antar beberapa arsitektur model yang berbeda, sehingga rekomendasi strategi retensi yang dihasilkan berpotensi bias terhadap karakteristik satu algoritma saja [12].

Penelitian ini secara khusus dirancang untuk menjawab ketiga celah tersebut. Pertama, interpretabilitas model diperkuat melalui penerapan *SHAP* secara menyeluruh, baik pada level global maupun lokal, sejalan dengan rekomendasi penelitian yang menekankan integrasi *XAI* pada studi prediksi *attrition* [6], [13]. Kedua, teknik *SMOTE* digabungkan ke dalam satu pipeline bersama *StandardScaler* sehingga proses *resampling* hanya diterapkan pada data latih di setiap *fold* validasi silang, sebagai bentuk kehati-hatian metodologis terhadap risiko kebocoran data yang belum banyak ditekankan pada studi-studi sebelumnya [8], [9]. Ketiga, konsistensi fitur penting diuji lintas tiga algoritma dengan arsitektur berbeda seperti model *linear* (*Logistic Regression*), *ensemble* berbasis *bagging* (*Random Forest*), dan *ensemble* berbasis *boosting* (*XGBoost*) untuk menghasilkan dasar rekomendasi retensi karyawan yang lebih *robust* dan tidak bergantung pada satu jenis algoritma saja [7], [14].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk: (1) membandingkan performa tiga algoritma klasifikasi, yaitu *Logistic Regression*, *Random Forest*, dan *XGBoost*, dalam memprediksi *employee attrition*; (2) menerapkan *SMOTE* di dalam pipeline pemodelan untuk menangani ketidakseimbangan kelas tanpa menimbulkan kebocoran data; (3) mengevaluasi stabilitas dan signifikansi performa model melalui validasi silang berulang dan strategi ambang batas klasifikasi; serta (4) menganalisis faktor-faktor dominan penyebab *attrition* menggunakan *feature importance* dan *SHAP* sebagai dasar penyusunan rekomendasi strategi retensi karyawan. Penelitian ini juga relevan dengan ruang lingkup JPTI pada bidang Teknologi, khususnya Teknik Informatika dan Ilmu Pengetahuan dan Teknologi, karena menghasilkan alur kerja (pipeline) pembelajaran mesin yang teruji secara metodologis serta pendekatan *explainable AI* yang dapat menjadi rujukan pengembangan sistem analitik data dan sistem pendukung keputusan berbasis kecerdasan buatan.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen kuantitatif dengan alur kerja (*pipeline*) *machine learning* yang sistematis, mulai dari praproses data, pelatihan model, evaluasi statistik, hingga analisis eksplainabilitas model (*explainable AI*). Arsitektur menyeluruh dari tahapan penelitian ditampilkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

Sebagaimana ditunjukkan pada Gambar 1, penelitian ini terdiri atas lima tahap utama, yaitu praproses data, pemilihan fitur, pembagian data latih dan uji, pelatihan model dengan pipeline *SMOTE*, *tuning hyperparameter*, evaluasi statistik berlapis, dan analisis *explainable AI* (XAI) menggunakan *SHAP*.

2.1. Dataset

Data yang digunakan pada penelitian ini adalah dataset *employee_reviews.csv* yang berisi 1.057 baris data karyawan dengan 12 atribut awal, meliputi *firm*, *job_title*, *years_working*, *location*, serta lima skor penilaian (*overall*, *work_life_balance*, *culture*, *Diversity_Equity_Inclusion*, *career_opp*, *benefit*, dan *management*) dalam skala 1 sampai 5. Variabel target penelitian adalah *is_resign* yang bersifat biner, dengan distribusi 632 karyawan berstatus tidak resign (0) dan 425 karyawan berstatus resign (1). Distribusi ini menunjukkan adanya ketidakseimbangan kelas sedang (rasio $\pm 60:40$) yang perlu ditangani pada tahap pemodelan. Tabel *Dataset* dapat dilihat pada Tabel 2 di bawah. Pemilihan dataset *employee_reviews.csv* didasarkan pada ketersediaan atribut ulasan karyawan Indonesia yang mencakup skor kepuasan kerja multidimensi serta label resign eksplisit, yang relatif jarang ditemukan pada dataset publik lain untuk konteks Indonesia. Ukuran sampel sebanyak 1.057 baris data dinilai memadai untuk skema validasi silang 5-fold berulang yang digunakan pada penelitian ini, meskipun disadari lebih kecil dibandingkan dataset benchmark internasional seperti *IBM HR Analytics*.

Tabel 2. Dataset

Atribut	Detail
Sumber Data	Kaggle (Indonesia Employee Reviews Dataset from Glassdoor)
Ukuran Data	1057
Indikator Kunci	<i>firm</i> , <i>job_title</i> , <i>years_working</i> , <i>location</i> , <i>overall</i> , <i>work_life_balance</i> , <i>culture</i> , <i>Diversity_Equity_Inclusion</i> , <i>career_opp</i> , <i>benefit</i> , <i>management</i>
Variabel Target	<i>is_resign</i> (0/1)
Jumlah Fitur	12

2.2. Praproses dan Rekayasa Fitur

Tahap praproses diawali dengan pemeriksaan nilai kosong (*missing value*); hasil audit menunjukkan tidak terdapat nilai kosong pada dataset. Selanjutnya dilakukan rekayasa fitur (*feature engineering*) dengan membentuk lima fitur turunan dari skor kepuasan kerja, yaitu *avg_score* (rata-rata skor), *min_score*, *max_score*, *score_std*

(standar deviasi antar skor), dan *overall_vs_avg* (selisih skor *overall* terhadap rata-rata skor). Fitur kategorikal *job_title* dan *location* dikodekan menggunakan *Label Encoding*. Sebelum pemodelan, dilakukan audit label target untuk memastikan *is_resign* bukan hasil transformasi langsung dari salah satu skor rating; korelasi tertinggi antara *is_resign* dengan *overall* dan *avg_score* hanya sebesar -0,139 dan -0,154, sehingga label dinyatakan valid dan tidak bocor dari fitur skor. Seleksi fitur dilakukan dengan membuang fitur berkorelasi sangat rendah terhadap target (*overall_vs_avg* dan *job_title*), sehingga tersisa 14 fitur yang digunakan pada tahap pemodelan.

2.3. Pembagian Data dan Pipeline Pemodelan

Data dibagi menjadi data latih dan data uji dengan proporsi 80:20 secara *stratified* terhadap target, menghasilkan 845 data latih (505 tidak *resign*, 340 *resign*) dan 212 data uji (127 tidak *resign*, 85 *resign*). Untuk mengatasi ketidakseimbangan kelas sekaligus mencegah kebocoran data, *StandardScaler* dan *SMOTE* digabungkan ke dalam satu *Pipeline* (*imblearn.Pipeline*) bersama algoritma klasifikasi. Dengan konfigurasi tersebut, proses penskalaan dan penyeimbangan data hanya diterapkan pada data latih di setiap fold validasi silang, sedangkan pada fold validasi maupun data uji, scaler dan model yang telah dilatih sebelumnya hanya digunakan untuk mentransformasi dan memprediksi data tersebut tanpa ikut membentuk *parameter scaling* maupun menghasilkan sampel sintesis SMOTE baru, sehingga estimasi performa model tetap mencerminkan kondisi data yang belum pernah digunakan pada proses pelatihan.

2.4. Algoritma dan Tuning Hyperparameter

Tiga algoritma klasifikasi dibandingkan pada penelitian ini, yaitu *Logistic Regression*, *Random Forest*, dan *XGBoost*. Optimasi *hyperparameter* dilakukan menggunakan *RandomizedSearchCV* dengan 30 iterasi pencarian, validasi silang 5-fold stratified (*StratifiedKfold*), dan metrik skoring *F1*. Ruang pencarian parameter meliputi

regularisasi C dan $class_weight$ untuk *Logistic Regression*; jumlah pohon, kedalaman maksimum, dan pengaturan sampel daun untuk *Random Forest*; serta $learning_rate$, kedalaman pohon, subsample, $colsample_bytree$, $gamma$, dan $parameter$ regularisasi (reg_alpha , reg_lambda) untuk *XGBoost*. Model terbaik hasil tuning kemudian diuji stabilitasnya menggunakan *Repeated Stratified K-Fold* (5-fold, 10 pengulangan, total 50 kali pengujian) pada metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Ruang pencarian *parameter tuning* ditetapkan berdasarkan rentang umum yang direkomendasikan pada literatur dan dokumentasi resmi masing-masing pustaka (*scikit-learn* dan *XGBoost*), dengan mempertimbangkan ukuran *dataset* yang relatif kecil sehingga rentang kedalaman pohon dan jumlah *estimator* dibatasi untuk menghindari *overfitting*, sedangkan pemilihan metrik skoring F1 didasarkan pada karakteristik ketidakseimbangan kelas pada data target.

2.5. Analisis Explainable AI

Untuk mengatasi keterbatasan interpretasi *feature importances* bawaan (yang hanya mencerminkan kekuatan pengaruh, bukan arah, dan tidak dapat dibandingkan langsung antar jenis algoritma), penelitian ini menerapkan *SHAP* (*SHapley Additive exPlanations*). *TreeExplainer* digunakan untuk model berbasis pohon (*Random Forest* dan *XGBoost*), sedangkan *LinearExplainer* digunakan untuk *Logistic Regression*, dengan sampel sebanyak 100 baris data uji yang diproses melalui *scaler* yang sama dengan yang dilihat oleh masing-masing *classifier*. Analisis *SHAP* menghasilkan (1) interpretasi global berupa kekuatan dan arah pengaruh setiap fitur, (2) interpretasi lokal yang menjelaskan kontribusi fitur terhadap satu prediksi individu, dan (3) analisis konsistensi peringkat fitur penting antar ketiga model.

2.6. Formulasi Matematis Model dan Metode

Bagian ini menguraikan dasar perhitungan matematis dari masing-masing algoritma, teknik penyeimbangan data, dan metode eksplainabilitas yang digunakan pada penelitian ini.

Logistic Regression memodelkan probabilitas kelas positif (*resign*) menggunakan fungsi *sigmoid* terhadap kombinasi linear fitur, sebagaimana dirumuskan pada Persamaan (1).

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (1)$$

dengan w_i adalah vektor bobot fitur, x adalah vektor input, dan b adalah bias. *Parameter* model dioptimasi dengan meminimalkan fungsi *loss log-loss* (*binary cross-entropy*) pada Persamaan (2).

$$L = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2)$$

Random Forest membangun T buah *decision tree* secara independen pada subset data dan fitur yang berbeda (*bootstrap aggregating*). Setiap pemisahan (*split*) *node* pada pohon dipilih dengan meminimalkan impuritas *Gini*, dirumuskan pada Persamaan (3).

$$Gini = 1 - \sum_{i=1}^k p_i^2 \quad (3)$$

dengan p_i adalah proporsi sampel kelas i pada suatu *node*. Prediksi akhir *Random Forest* diperoleh melalui agregasi (*majority voting* untuk klasifikasi) dari seluruh pohon, seperti pada Persamaan (4).

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\} \quad (4)$$

XGBoost membangun model secara bertahap (*boosting*) dengan meminimalkan fungsi objektif yang terdiri dari fungsi *loss* dan *term* regularisasi, dirumuskan pada Persamaan (5).

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (5)$$

dengan *term* regularisasi $\Omega(f)$ untuk mengontrol kompleksitas pohon dirumuskan pada Persamaan (6).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (6)$$

Pemilihan titik *split* terbaik pada *XGBoost* didasarkan pada *gain*, yang dihitung dari gradien pertama (G) dan gradien kedua (H) sampel pada *node* kiri (L) dan kanan (R), dirumuskan pada Persamaan (7).

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (7)$$

Penyeimbangan kelas dengan *SMOTE* menghasilkan sampel sintetis kelas minoritas melalui interpolasi linear antara sampel x_i dan salah satu tetangga terdekatnya x_{zi} , sebagaimana Persamaan (8).

$$x_{new} = x_i + \lambda \cdot (x_{zi} - x_i), \lambda \in [0,1] \quad (8)$$

Interpretasi model dilakukan dengan menghitung nilai *Shapley* (*SHAP value*) berbasis teori permainan kooperatif, yang mengukur kontribusi rata-rata suatu fitur i terhadap prediksi di seluruh kemungkinan kombinasi *subset* fitur F , dirumuskan pada Persamaan (9).

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (10)$$

Signifikansi performa antar model turut dievaluasi menggunakan *Matthews Correlation Coefficient* (*MCC*) yang memperhitungkan keempat sel *confusion matrix* secara seimbang, dirumuskan pada Persamaan (10).

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (11)$$

dengan *TP*, *TN*, *FP*, dan *FN* berturut-turut adalah jumlah *true positive*, *true negative*, *false positive*, dan *false negative* pada hasil klasifikasi data uji.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dapat diuraikan mengenai hasil dari penelitian beserta pengujian yang telah dilakukan. Selain itu, disampaikan juga mengenai pembahasan dari penelitian maupun pengujian yang telah dilakukan.

3.1. Hasil Tuning Hyperparameter

Hasil *RandomizedSearchCV* menunjukkan kombinasi hyperparameter terbaik untuk masing-masing model beserta skor *F1* validasi silang, sebagaimana disajikan pada Tabel 3.

Model	Parameter Terbaik	Best CV F1
Logistic Regression	C = 0,1; class_weight = None	0,505
Random Forest	n_estimators = 400; max_depth = 5; min_samples_leaf = 4; class_weight = balanced	0,494
XGBoost	n_estimators = 400; max_depth = 4; learning_rate = 0,1; subsample = 0,9; gamma = 0,2	0,494

Nilai *Best CV F1* pada ketiga model relatif setara di kisaran 0,49–0,50, mengindikasikan bahwa perbedaan performa antar algoritma pada tahap *tuning* tidak terlalu signifikan.

3.2. Uji Stabilitas Repeated Stratified K-Fold

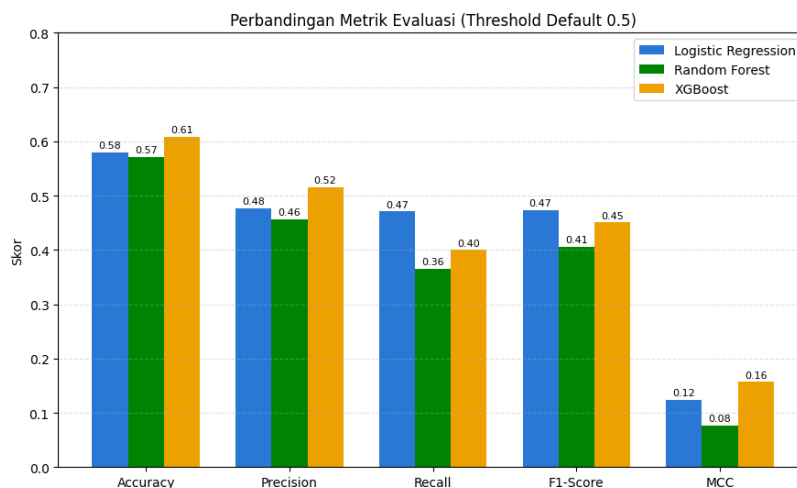
Untuk memastikan performa model tidak dipengaruhi oleh pembagian data tertentu, dilakukan pengujian stabilitas dengan *Repeated Stratified K-Fold* (5-fold x 10 pengulangan = 50 kali pengujian) tanpa kebocoran data. Hasil rata-rata dan simpangan baku tiap metrik ditampilkan pada Tabel 4.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0,566 ± 0,035	0,465 ± 0,040	0,508 ± 0,058	0,484 ± 0,043	0,583 ± 0,035
Random Forest	0,583 ± 0,030	0,483 ± 0,039	0,478 ± 0,066	0,478 ± 0,045	0,599 ± 0,030
XGBoost	0,593 ± 0,031	0,493 ± 0,043	0,453 ± 0,064	0,471 ± 0,051	0,587 ± 0,035

Tabel 4 menunjukkan bahwa *Random Forest* memiliki rata-rata *ROC-AUC* validasi silang tertinggi (0,599), sedikit di atas *XGBoost* (0,587) dan *Logistic Regression* (0,583). Nilai simpangan baku yang relatif kecil (0,03–0,06) pada seluruh metrik menandakan performa ketiga model cukup stabil dan tidak terlalu sensitif terhadap pembagian *fold* data.

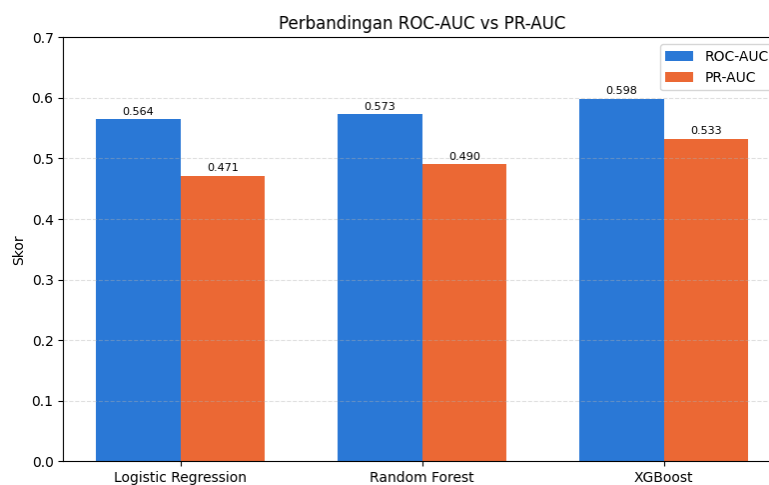
3.3. Evaluasi pada Data Uji

Evaluasi akhir pada data uji (*held-out test set*) dilakukan dengan membandingkan tiga strategi ambang batas klasifikasi. Ringkasan hasil evaluasi pada ambang, akurasi, presisi, *recall*, dan *F1-Score* ditampilkan pada Tabel 4.



Gambar 2. Evaluasi Metrik

Berdasarkan Gambar 2, *XGBoost* mencapai akurasi tertinggi (0,610), Presisi Tertinggi (0,520) dan MCC tertinggi (0,160) pada *Threshold Default*, namun dengan *recall* yang rendah (0,400) sehingga banyak karyawan yang benar-benar resign tidak terdeteksi. Sebaliknya, *Logistic Regression* menghasilkan keseimbangan terbaik antara precision (0,480) dan recall (0,470), tercermin dari *F1-Score* tertinggi (0,473) di antara ketiga model pada strategi ambang batas ini. *Random Forest* berada pada posisi menengah dengan performa yang relatif stabil. Perbandingan *ROC-AUC* dan *PR-AUC* yang tidak bergantung pada pemilihan *threshold* ditampilkan pada Tabel 5.

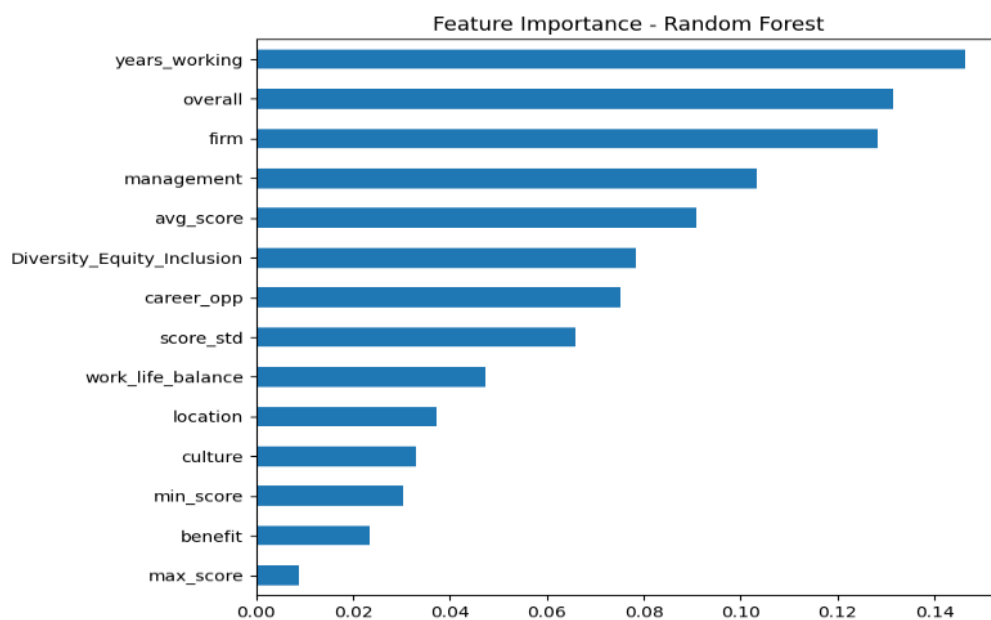


Gambar 3. ROC-AUC dan PR-AUC pada Data Uji

Gambar 3 menunjukkan bahwa *XGBoost* memberikan *ROC-AUC* tertinggi pada data uji sebesar 0,598, diikuti *Random Forest* (0,573) dan *Logistic Regression* (0,564). Nilai *ROC-AUC* pada rentang 0,56–0,60 tergolong moderat, mengindikasikan bahwa ketiga model memiliki kemampuan diskriminasi yang lebih baik

dibanding tebakan acak (AUC 0,5), namun belum kuat, sehingga secara keseluruhan *XGBoost* dipilih sebagai model dengan performa terbaik pada penelitian ini ditinjau dari $ROC-AUC$, sedangkan *Logistic Regression* lebih unggul apabila prioritas evaluasi adalah keseimbangan *precision* dan *recall* ($F1-Score$).

3.4. Feature Importance dan Analisis SHAP



Gambar 4. Feature Importance

Berdasarkan Gambar 4, Analisis *feature importance* berbasis *Random Forest* menunjukkan bahwa fitur *years_working* (0,147), *overall* (0,132), *firm* (0,128), dan *management* (0,103) merupakan empat fitur dengan kontribusi tertinggi terhadap prediksi *attrition*, sedangkan fitur *max_score* (0,009) dan *benefit* (0,024) memberikan kontribusi paling kecil.

Untuk memperoleh interpretasi yang *comparable* antar ketiga algoritma sekaligus arah pengaruhnya, dilakukan analisis *SHAP* terhadap 100 sampel data uji. Ringkasan lima fitur teratas menurut nilai rata-rata $|SHAP|$ pada masing-masing model ditampilkan pada Tabel 5.

Tabel 5. Lima Fitur Teratas Berdasarkan Rata-Rata $|SHAP\ Value|$ per Model

Peringkat	Logistic Regression	Random Forest	XGBoost
1	firm (Negatif)	overall (Negatif)	firm (Negatif)
2	years_working (Negatif)	firm (Negatif)	years_working (Negatif)
3	overall (Negatif)	management (Negatif)	avg_score (Negatif)
4	benefit (Positif)	years_working (Negatif)	score_std (Negatif)
5	management (Negatif)	career_opp (Negatif)	overall (Negatif)

Tabel 5 memperlihatkan bahwa fitur *firm*, *years_working*, *overall*, dan *management* secara konsisten muncul sebagai fitur paling berpengaruh pada ketiga model, meskipun urutannya sedikit berbeda. Arah pengaruh yang dominan negatif pada fitur-fitur tersebut mengindikasikan bahwa semakin rendah nilai fitur (misalnya masa kerja yang pendek, skor *overall* yang rendah, atau penilaian manajemen yang buruk), semakin tinggi kontribusi *SHAP* terhadap prediksi *resign*. Sebaliknya, fitur *benefit* menunjukkan arah pengaruh positif pada *Logistic Regression* dan *XGBoost*, yang secara intuitif konsisten dengan logika bisnis yaitu benefit yang lebih baik cenderung menahan karyawan untuk tidak *resign*.

Analisis konsistensi lintas model (Tabel 6) menghitung jumlah kesamaan fitur pada lima fitur teratas antar pasangan model.

Tabel 6. Konsistensi Interpretasi Fitur Penting Antar Model

Pasangan Model	Jumlah Fitur Sama (Top 5)	Fitur yang Sama
Logistic Regression vs Random Forest	4	firm, management, overall, years_working
Logistic Regression vs XGBoost	3	firm, overall, years_working
Random Forest vs XGBoost	3	firm, overall, years_working

Tingginya kesamaan fitur teratas antar model (3–4 dari 5 fitur) menunjukkan bahwa hasil interpretasi bersifat *robust* dan tidak bergantung pada satu algoritma tertentu, sehingga fitur *firm*, *years_working*, dan *overall* dapat dijadikan dasar yang kuat untuk rekomendasi strategi retensi karyawan.

Analisis *SHAP* lokal pada salah satu sampel data uji, dijelaskan menggunakan model *XGBoost* sebagai model dengan *ROC-AUC* tertinggi, menunjukkan bahwa fitur *years_working* memberikan kontribusi negatif terbesar (*SHAP* = -1,190) terhadap probabilitas *resign* karyawan tersebut, diikuti oleh *Diversity Equity Inclusion* (*SHAP* = -0,278) dan *career opp* (*SHAP* = -0,274), sementara *firm* memberikan kontribusi positif terbesar (*SHAP* = 0,480). Contoh ini menegaskan bahwa selain pola

global, *SHAP* mampu menjelaskan logika keputusan model untuk setiap individu karyawan, yang bermanfaat bagi HR dalam melakukan audit kasus per kasus. Secara ringkas, temuan utama pada bagian hasil menegaskan tiga hal: (1) *XGBoost* unggul pada *ROC-AUC* namun performa ketiga model tergolong moderat; (2) fitur *firm*, *years_working*, *overall*, dan *management* konsisten menjadi faktor paling berpengaruh; dan (3) interpretasi *SHAP* terbukti stabil lintas arsitektur model yang berbeda.

3.5. Pembahasan

Secara keseluruhan, nilai *ROC-AUC* yang diperoleh ketiga model berada pada kisaran 0,56–0,60, yang tergolong moderat dan menunjukkan bahwa sinyal prediktif dari fitur-fitur yang tersedia pada *dataset Kaggle (Indonesia Employee Reviews Dataset from Glassdoor)* relatif lemah terhadap label *is_resign*. Hal ini didukung oleh korelasi *Pearson* antar fitur individual dengan target yang seluruhnya berada di bawah 0,16 (fitur *avg_score* memiliki korelasi tertinggi sebesar 0,154). Kondisi ini konsisten dengan nilai *SHAP* yang kecil dan tersebar pada banyak fitur, bukan terkonsentrasi pada satu atau dua fitur dominan, sehingga permasalahan utama lebih terletak pada kekuatan sinyal data dibandingkan pada pemilihan algoritma atau proses *tuning*.

Penerapan *pipeline SMOTE* dan *StandardScaler* yang digabungkan ke dalam satu objek *pipeline* terbukti penting untuk menghasilkan estimasi performa yang jujur. Sebagai perbandingan, penerapan *SMOTE* di luar *loop* validasi silang pada eksperimen awal menghasilkan skor *CV-AUC* yang jauh lebih tinggi (berkisar 0,71–0,74) dibandingkan performa sesungguhnya pada data uji (0,56–0,60), yang membuktikan adanya kebocoran data apabila resampling dilakukan sebelum pembagian *fold*. Temuan ini menegaskan pentingnya kehati-hatian metodologis dalam menangani ketidakseimbangan kelas pada studi prediksi *attrition*.

Dibandingkan dengan penelitian acuan yang menggabungkan *LightGBM* dan *XGBoost* dengan skema penjumlahan berbobot dan mencapai *ROC-AUC* 0,9504 pada *dataset IBM HR Analytics* [1], performa model pada penelitian ini jauh lebih rendah. Perbedaan tersebut kemungkinan besar disebabkan oleh karakteristik *dataset* yang berbeda: *dataset Kaggle (Indonesia Employee Reviews Dataset from Glassdoor)* pada penelitian ini bersumber dari ulasan karyawan dengan jumlah sampel yang lebih kecil (1.057 baris) dan fitur yang didominasi oleh skor kepuasan kategorikal, sedangkan *dataset IBM HR Analytics* memiliki 35 fitur yang lebih kaya, termasuk data kompensasi dan demografi terperinci. Perbandingan ini menggarisbawahi bahwa kualitas dan kelengkapan fitur *input* menjadi faktor penentu utama performa model prediksi *attrition*, sejalan dengan simpulan bahwa model *machine learning* secanggih apapun tetap dibatasi oleh kekuatan sinyal pada data yang tersedia.

Hal ini selaras dengan temuan Manafi et al. [15] dan Marin Diaz et al. [16] yang menyoroti peran kualitas manajemen dan lingkungan organisasi dalam keputusan *resign*, serta konsisten dengan penelitian Konar [6] yang menekankan bahwa faktor psikologis (kepuasan kerja dan hubungan atasan-bawahan) seringkali memiliki bobot lebih tinggi dibandingkan faktor kompensasi non-moneter. Rendahnya kontribusi fitur benefit pada penelitian ini juga menguatkan temuan Talebi et al. [2] bahwa faktor kompensasi *non-moneter* seringkali memiliki bobot lebih rendah dibandingkan faktor psikologis dan hubungan kerja dalam konteks dataset berbasis ulasan karyawan.

Ditinjau dari sisi interpretabilitas, hasil *SHAP* pada penelitian ini konsisten dengan temuan pada literatur terkait yang menempatkan kepuasan kerja, kompensasi/benefit, dan masa kerja sebagai faktor dominan penyebab *turnover*. Konsistensi fitur *firm*, *years_working*, *overall*, dan *management* pada tiga model yang berbeda arsitektur (*linear*, *ensemble tree* berbasis *bagging*, dan *gradient boosting*) memperkuat validitas temuan ini

sebagai dasar rekomendasi praktis. Perusahaan disarankan untuk memperkuat program retensi pada karyawan dengan masa kerja pendek, skor kepuasan kerja keseluruhan yang rendah, dan penilaian terhadap kualitas manajemen yang kurang baik, misalnya melalui program mentoring pada tahun-tahun awal bekerja, evaluasi berkala terhadap gaya kepemimpinan lini manajemen, serta peninjauan paket benefit bagi karyawan pada segmen berisiko tinggi.

Penelitian ini memiliki beberapa keterbatasan yang perlu diakui. Pertama, ukuran sampel yang relatif kecil (1.057 baris) dan cakupan fitur yang terbatas pada skor kepuasan kerja menyebabkan sinyal prediktif yang tertangkap masih moderat sehingga generalisasi hasil ke populasi karyawan yang lebih luas perlu dilakukan dengan hati-hati. Kedua, dataset bersifat *cross-sectional* sehingga tidak dapat menangkap dinamika perubahan kepuasan kerja dan risiko *resign* dari waktu ke waktu. Untuk penelitian selanjutnya, disarankan memperkaya fitur dengan data kompensasi, riwayat promosi, dan interaksi atasan-bawahan, serta mengeksplorasi pendekatan *longitudinal* atau deep learning berbasis data tekstual ulasan karyawan untuk meningkatkan kekuatan sinyal prediktif model.

4. KESIMPULAN

Penelitian ini membandingkan *Logistic Regression*, *Random Forest*, dan *XGBoost* untuk memprediksi *employee attrition* pada dataset ulasan karyawan dengan penanganan ketidakseimbangan kelas menggunakan *SMOTE* yang diintegrasikan ke dalam *pipeline* agar terhindar dari kebocoran data. Hasil evaluasi pada data uji menunjukkan *XGBoost* memberikan *ROC-AUC* tertinggi sebesar 0,598 dan akurasi tertinggi sebesar 0,608 pada ambang batas *Default* (0,5), sementara *Logistic Regression* memberikan keseimbangan *precision-recall* terbaik dengan *F1-Score* tertinggi sebesar 0,473. Analisis *feature importance* dan SHAP secara konsisten mengidentifikasi *firm*, *years_working*, *overall*, dan *management* sebagai faktor paling dominan yang memengaruhi keputusan *resign* karyawan, dengan tingkat kesamaan fitur teratas antar model mencapai 60–80%. Nilai *ROC-AUC* yang tergolong moderat (0,56–0,60) mengindikasikan bahwa kekuatan sinyal prediktif pada fitur yang tersedia masih terbatas, sehingga peningkatan performa model ke depan lebih bergantung pada penambahan fitur yang lebih kaya (misalnya data kompensasi, riwayat promosi, dan interaksi dengan atasan) dibandingkan hanya mengandalkan penyesuaian algoritma. Secara praktis, hasil penelitian ini dapat dijadikan dasar bagi manajemen SDM untuk menyusun strategi retensi yang berfokus pada peningkatan kepuasan kerja keseluruhan, kualitas manajemen lini, serta program khusus bagi karyawan dengan masa kerja pendek.

Secara ilmiah, kontribusi utama penelitian ini terletak pada penerapan *pipeline SMOTE-StandardScaler* yang bebas kebocoran data serta pengujian konsistensi interpretasi SHAP lintas tiga arsitektur model berbeda, yang belum banyak dilakukan secara bersamaan pada studi *attrition* sebelumnya. Keterbatasan penelitian meliputi ukuran sampel dan cakupan fitur yang masih terbatas, sehingga penelitian lanjutan disarankan memperkaya fitur (misalnya data kompensasi dan riwayat promosi) serta menguji generalisasi model pada dataset dari sektor industri yang berbeda.

DAFTAR PUSTAKA

- [1] R. Qin, X. Qi, Y. Yuan, and B. Alatas, "Employee Turnover Prediction Research of Human Resource Management on Machine Learning Algorithms and Big Data Analysis," *Journal of Organizational and End User Computing*, vol. 38, no. 1, Jan. 2026, doi: 10.4018/JOEUC.399146.
- [2] H. Talebi, A. K. Bardsiri, and V. K. Bardsiri, "Developing a hybrid machine learning model for employee turnover prediction: Integrating LightGBM and genetic algorithms," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 2, Jun. 2025, doi: 10.1016/j.joitmc.2025.100557.
- [3] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," *Applied Sciences (Switzerland)*, vol. 12, no. 13, Jul. 2022, doi: 10.3390/app12136424.
- [4] Badan Pusat Statistik, "Pencari Kerja Terdaftar, Lowongan Kerja Terdaftar, dan Penempatan/Pemenuhan Tenaga Kerja Menurut Provinsi dan Jenis Kelamin, 2024," BPS-Statistics Indonesia. Accessed: Jul. 29, 2026. [Online]. Available: <https://www.bps.go.id/en/statistics-table/3/VEU5VVVERIVWM0JwYTNvdkIISkpWR3R1VUhaVmR6MDkjMw==/pencari-kerja-terdaftar--lowongan-kerja-terdaftar--dan-penempatan-pemenuhan-tenaga-kerja-menurut-provinsi-dan-jenis-kelamin--2024.html?year=2024>
- [5] C. Zhang and W. Han, "Ensembles of decision trees and gradientbased learning for employee turnover rate prediction," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2387.

-
- [6] K. Konar, S. Das, S. Das, and S. Misra, "Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI," *SN Comput. Sci.*, vol. 6, no. 6, Aug. 2025, doi: 10.1007/s42979-025-04204-w.
 - [7] S. Barman et al., "An Explainable Machine Learning Framework for Prediction of Employee Attrition," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2025, pp. 615–627. doi: 10.1007/978-981-96-2721-9_40.
 - [8] Q. Ke and Y. Xu, "Integrating Textual Features with Survival Analysis for Predicting Employee Turnover," *Behavioral Sciences*, vol. 16, no. 2, Feb. 2026, doi: 10.3390/bs16020174.
 - [9] W. Wu and S. Fukui, "Using Human Resources Data to Predict Turnover of Community Mental Health Employees: Prediction and Interpretation of Machine Learning Methods," *Int. J. Ment. Health Nurs.*, vol. 33, no. 6, pp. 2180–2192, Dec. 2024, doi: 10.1111/inm.13387.
 - [10] M. I. Alyousef, H. W. Khan, and M. U. Sattar, "A Hybrid Predictive Model for Employee Turnover: Integrating Ensemble Learning and Feature-Driven Insights from IBM HR Analytics," *Information (Switzerland)*, vol. 17, no. 2, Feb. 2026, doi: 10.3390/info17020208.
 - [11] Y. Li, Y. Yang, P. Song, L. Duan, and R. Ren, "An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-09506-w.
 - [12] J. Park, Y. Feng, and S. P. Jeong, "Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-023-50593-4.
 - [13] W. Li, "A transformer-based deep learning framework to predict employee attrition," *PeerJ Comput. Sci.*, vol. 9, 2023, doi: 10.7717/peerj-cs.1570.
 - [14] İ. T. Baydili and B. Tasci, "Predicting Employee Attrition: XAI-Powered Models for Managerial Decision-Making," *Systems*, vol. 13, no. 7, Jul. 2025, doi: 10.3390/systems13070583.
 - [15] S. Manafi Varkiani, F. Pattarin, T. Fabbri, and G. Fantoni, "Predicting employee attrition and explaining its determinants," *Expert Syst. Appl.*, vol. 272, May 2025, doi: 10.1016/j.eswa.2025.126575.
 - [16] G. Marín Díaz, J. J. Galán Hernández, and J. L. Galdón Salvador, "Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making," *Mathematics*, vol. 11, no. 22, Nov. 2023, doi: 10.3390/math11224677.