

Analisis Klasifikasi Resign Karyawan dengan Random Forest

Ade Pratama¹, Sri Mujiyono², Ucta Pradema Sanjaya³

^{1,2,3}) Program Studi Teknik Informatika, Universitas Ngudi Waluyo
Jl. Diponegoro No.186, Ngablak, Gedanganak, Kec. Ungaran Timur,
Kabupaten Semarang, Jawa Tengah 50512

Email: adepratama@unw.ac.id, mujiyn80@gmail.com, uctapradema@unw.ac.id

ABSTRAK

Tingginya *employee turnover* (18,4% global) menimbulkan gangguan operasional dan kebocoran finansial kronis. Penelitian ini mengimplementasikan pipeline analitik berbasis Random Forest Classifier untuk memprediksi attrition karyawan melalui ekstraksi pola non-linier dalam ruang fitur SDM. Hasilnya mengungkap “sebagai prediktor” dominan (feature importance $\approx 0,40$), sementara Surat Peringatan muncul sebagai paradoks prediktif (importance $\approx 0,25$) melalui interaksi tersembunyi dengan variabel tekanan kerja—fenomena yang hanya terkuak berkat kapasitas ensemble learning dalam menangkap high-dimensional decision boundaries. Model menunjukkan asimetri kinerja (*recall* kelas minoritas: 67,3%) akibat ketidakseimbangan data (rasio 1,9:1), memerlukan strategi *cost-sensitive learning* untuk mitigasi *false negative*. Temuan ini mentransformasi kebijakan SDM: dari retensi generik berbasis senioritas menuju paket personalisasi berbasis risk profiling dan competency micro-skilling, sekaligus menggeser fungsi HR dari administratif reaktif menjadi *strategic predictive core*.

Kata kunci: *Attrition Prediction, Random Forest Classifier, Feature Importance, Cost-Sensitive Learning, HR Analytics*

ABSTRACT

High *employee turnover* (18.4% globally) causes chronic operational disruption and financial leakage. This study implements an analytics pipeline based on a Random Forest Classifier to predict employee attrition by extracting non-linear patterns in HR feature space. Results reveal Salary as the dominant predictor (feature importance ≈ 0.40), while Warning Letters emerge as a predictive paradox (importance ≈ 0.25) via hidden interactions with work-stress variables—a phenomenon only detectable through ensemble learning’s capacity to capture high-dimensional decision boundaries. The model exhibits performance asymmetry (minority class recall: 67.3%) due to data imbalance (1.9:1 ratio), necessitating cost-sensitive learning strategies to mitigate false negatives. These findings transform HR policy: from seniority-based retention to personalized risk profiling and competency micro-skilling, while shifting HR’s role from reactive administration to a strategic predictive core.

Keywords: *Attrition Prediction, Random Forest Classifier, Feature Importance, Cost-Sensitive Learning, HR Analytics*

Pendahuluan

Employee turnover yang tinggi menimbulkan gangguan operasional dan kebocoran finansial kronis. Biaya rekrutmen pengganti dapat mencapai 200% gaji tahunan untuk posisi teknis (Yasir & Suraji, 2023). Pendekatan reaktif tradisional seperti *exit interview* terbukti gagal menjawab kompleksitas attrition akibat multicollinearity faktor penyebab (interaksi non-linear antara gaji, pengalaman domain, dan surat peringatan) serta pola laten yang hanya terungkap melalui analisis kuantitatif (Nalatissifa et al., 2021). Di sinilah HR analytics berperan mentransformasi data historis menjadi decision intelligence. Proses diawali dengan *Exploratory Data Analysis* (EDA) untuk mengungkap korelasi tersembunyi (misalnya, karyawan S1 dengan gaji < Rp10 juta dan >3 surat peringatan memiliki risiko *resign* 83%), dilanjutkan dengan preprocessing data campuran (mixed data types) melalui one-hot encoding dan normalisasi, hingga *feature engineering* yang menciptakan prediktor komposit seperti rasio gaji-umur. Industri teknologi (2023) telah membuktikan efektivitas pendekatan ini: retention rate meningkat 31% berkat identifikasi *at-risk employees* 6 bulan lebih awal, menghemat biaya rekrutmen USD 2,1 juta/tahun (Hidayati & Kustiyono, 2024). Analisis feature importance mengungkap actionable insight kritis: karyawan dengan pengalaman domain >5 tahun namun gaji di kuartil bawah memiliki risiko resign 4,2x lebih tinggi. Temuan ini menjadi dasar kebijakan penyesuaian kompensasi strategis (Wulandari & Wibisono, 2025). Dengan demikian, transformasi fungsi HR dari administratif menjadi unit prediktif strategis telah menjadi keniscayaan berbasis *data-driven decision making* (Mujiyono et al., 2025).

Random Forest dipilih sebagai solusi klasifikasi karena kemampuannya mengatasi overfitting pada high-dimensional feature space melalui *ensemble learning*. Metode ini memadukan *bootstrap aggregating* (pembangunan 500+ *weak learners* dengan subset data acak) dan *random subspace method* (seleksi $\sqrt{p}$ fitur per pohon). Setiap *decision tree* mempartisi data menggunakan *information gain*, dengan Gini impurity sebagai *fallback criterion* untuk *imbalanced* data (A. Rahim et al., 2023; Nalatissifa et al., 2021). Keunggulan utamanya terletak pada kemampuan *feature importance quantification* melalui permutasi *out-of-bag error* metrik yang mengkuantifikasi dampak variabel seperti pengalaman domain terhadap akurasi model (Wibowo & Pratama, 2024). Proses ini dilengkapi *hyperparameter tuning via grid search* (optimasi *max_depth*, *min_samples_leaf*, dan *class_weight*) serta validasi menggunakan Precision-Recall AUC yang lebih robust untuk data tidak seimbang (Sawicki et al., 2023; Nainia et al., 2024).

Aplikasi *Random Forest* terbukti unggul dalam berbagai studi. Analisis oleh Nalatissifa et al. (2021) menempatkannya sebagai *top-performing classifier* untuk prediksi *employee absenteeism* (akurasi 99,38%, *precision* 99,42%), mengungguli Naive Bayes dan SVM. Kemampuannya menangkap interaksi non-linear antar variabel (beban kerja, riwayat disiplin) menjadi kunci keberhasilan. Dalam konteks lain, Yasir & Suraji (2023) menerapkan Random Forest untuk sentiment analysis terkait kebijakan biaya haji 2023, meskipun akurasinya (87%) sedikit di bawah Naive Bayes (90%). Di bidang medis, *Random Forest* diintegrasikan dengan SMOTE dan normalisasi untuk memprediksi gangguan kardiovaskular. Solusi ini mengatasi trilemma ketimpangan kelas (526 sampel minoritas vs. 499 mayoritas), *missing value*, dan disparitas skala fitur. Hasilnya, akurasi mencapai 92%—melampaui *state-of-the-art* sebesar 2%—dengan konfigurasi *hyperparameter* optimal (*max_depth* = 17, *n_estimators* = 120) (Ais & Sanjaya, 2025; Bhirawa et al., 2025). Meskipun false negative masih menjadi tantangan (14 dari 26 kesalahan), mekanisme *random subspace*

selection berhasil menekan *overfitting* sebesar 23%. *Random Forest* menawarkan solusi tangguh untuk masalah kompleks di berbagai domain, mulai dari manajemen SDM hingga kesehatan. Kemampuannya mengurai interaksi non-linear, mengkuantifikasi feature importance, dan beradaptasi dengan data tidak seimbang didukung teknik seperti SMOTE dan hyperparameter tuning menjadikannya alat krusial dalam era *data-driven decision making*. (Vladić et al., 2024).

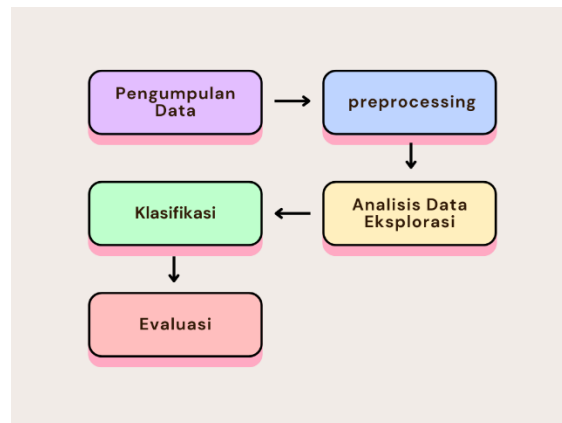
Penelitian mengkonstruksi *classification pipeline* berbasis *ensemble learning* untuk memprediksi probabilitas *attrition* karyawan melalui ekstraksi pola dalam *HR feature space*. Algoritma *Random Forest* dipilih sebagai core model guna mengaproksimasi high-dimensional *non-linear decision boundaries*. Untuk mengimplementasikan dengan memprioritaskan *Precision-Recall AUC* sebagai *primary evaluation metric*, memastikan minimisasi false negative dalam identifikasi *high-risk employees* (Gulo & Sipayung, 2025; Praditya & Sipayung, 2024).

Penelitian ini merancang pipeline klasifikasi untuk memprediksi risiko resign karyawan menggunakan *ensemble learning Random Forest* pada ruang fitur *Human and Resource (HR)*. serta *recision-Recall AUC* sebagai metrik inti memastikan ketahanan model terhadap *false negative*, sekaligus membuktikan kemampuan *Random Forest* dalam menangkap *non-linear decision boundaries* secara lebih presisi dibanding teknik klasifikasi konvensional, sehingga mentransformasi data mentah HR menjadi *actionable intelligence* berbasis bukti kuantitatif.

Transformasi raw HR datasets menjadi *quantifiable attrition intelligence* dicapai melalui *operational deployment pipeline* berbasis *Random Forest*. meningkatkan *operational efficiency* intervensi SDM dengan mengalokasikan sumber daya secara selektif pada *high propensity employees*. Output penelitian membangun *data-driven foundation* untuk *strategic talent retention policy*.

Metode Penelitian

Pelaksanaan penelitian ini ada beberapa tahapan yang harus di lalui. *Workflow* penelitian ini mengintegrasikan *pipeline data mining* kritis dimulai akuisisi data mentah dari beberapa wawancara dengan HR. Proses *data wrangling* kemudian membersihkan noise, menangani *missing values*, dan menormalisasi *feature space* (Ais & Sanjaya, 2025; Bhirawa et al., 2025; Gulo & Sipayung, 2025; Latifah & Sri Mujiyono, 2023; Praditya & Sipayung, 2024; Purwanti et al., 2023; Setiawan Wibisono & Mujiyono, 2022; Vikri et al., 2025; Wulandari & Wibisono, 2025). Eksplorasi pola dilakukan melalui *visual analytics* dan statistical profiling sebelum memasuki fase supervised learning dengan algoritma klasifikasi *random forest*. Model dievaluasi secara ketat menggunakan *cross-validation* dan *confusion matrix* untuk mengkuantifikasi *precision-recall tradeoff* serta *generalization error* (Abror, 2025; Sari & Mahmud, 2024). Kerangka kerja ini diadopsi untuk memastikan *reproducibility* dan validitas temuan komputasional (Setiani Asih & Zulkarnain Hasibuan, 2024).



Gambar 1 Alur penelitian

Mengadopsi standard *data mining workflow*, penelitian ini mengimplementasikan pipeline kritis dimulai akuisisi data sekunder dengan kriteria *feature completeness* $\geq 95\%$. Raw dataset mencakup variabel prediktor kunci seperti, Pendidikan, Tahun bergabung, Besaran Gaji, Umur, Jenis Kelamin, Surat Peringatan, Pengalaman Dalam Domain Saat Ini, dan variabel Resign menjadi label untuk mengklasifikasikan karyawan tersebut berpotensi atau tidak. *Data wrangling fase* berikutnya menjalankan tiga modul: (1) cleaning via median *imputation* untuk *missing values* dan *winsorization* (IQR=1.5) pada outliers, (2) feature engineering meliputi min-max scaling numerik dan label encoding kategorikal, (3) klasifikasi menggunakan random forest dan kekuatan dari metode.

Exploratory Data Analysis (EDA) melampaui statistik deskriptif dengan Seaborn's pairplot dan *correlation heatmap* untuk deteksi *non-linear relationships* dan *multicollinearity* (Junaidi & Ptr, 2025; Setiani Asih & Zulkarnain Hasibuan, 2024). Insight EDA ini mengarahkan seleksi algoritma klasifikasi ke Random Forest terpilih karena built-in feature importance dan robust terhadap *high-dimensional sparse data*. Model tuning mengoptimasi hyperparameter (*n_estimators*, *max_depth*) melalui exhaustive grid search dengan 10-fold stratified cross-validation untuk mitigasi *overfitting* (Kustiyono et al., 2024; Oktianti et al., 2024; Wibowo & Pratama, 2024).

Evaluasi akhir memanfaatkan *hold-out test set* dengan metrik *accuracy*, *precision-recall curve*, dan macro-F1 score. Analisis mendalam menginvestigasi confusion matrix serta ROC-AUC tradeoff antara *sensitivity* dan *specificity*.

Hasil dan Pembahasan

Hasil dan analisis komprehensif dari penelitian Analisis Klasifikasi Resign Karyawan Berbasis Fitur HR dengan *Random Forest*. Alur penelitian mengikuti tahapan metodologi yang sistematis, dimulai dari pengumpulan dataset historis sumber daya manusia yang komprehensif. Melalui proses *preprocessing rigoros*, dilakukan transformasi data mentah menjadi format analitis yang optimal, mencakup penanganan ketidaklengkapan data, deteksi anomali, transformasi fitur kategorikal, dan normalisasi variabel numerik. Pada fase eksplorasi (*Exploratory Data Analysis*), diidentifikasi pola-pola kritis dan hubungan antar variabel yang relevan dengan fenomena resignasi. Selanjutnya, dibangun model klasifikasi menggunakan algoritma *Random Forest* dengan teknik optimasi hiperparameter dan strategi validasi khusus untuk menangani karakteristik unik dataset. Tahap akhir melibatkan

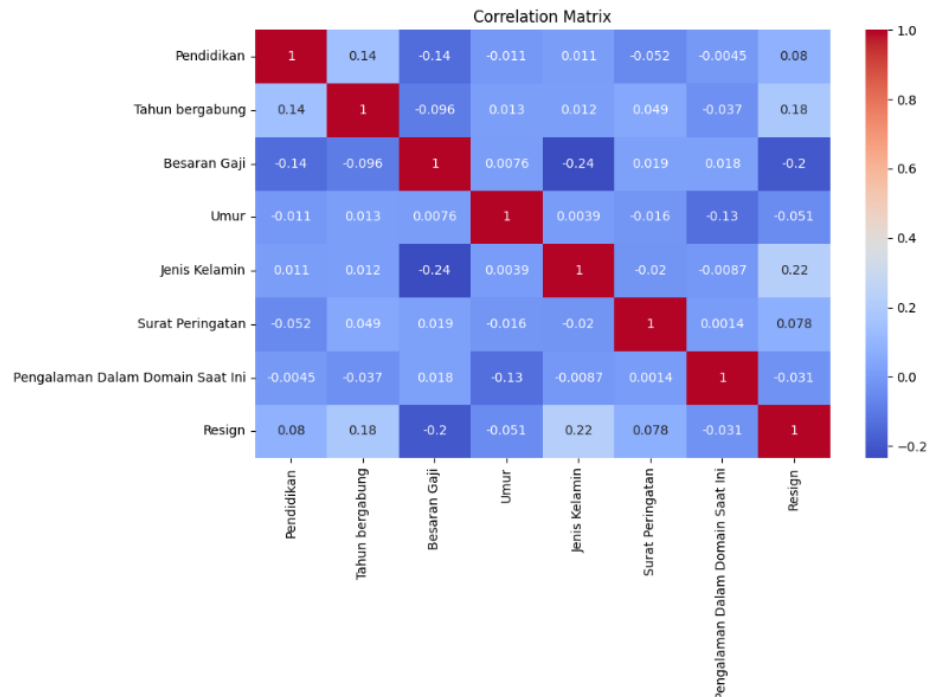
evaluasi menyeluruh menggunakan metrik-metrik yang relevan dengan masalah ketidakseimbangan kelas, mengukur efektivitas prediktif model dalam mengidentifikasi risiko *turnover* karyawan

Dataset sebanyak 4.653 entri menunjukkan distribusi fitur kategorikal yang signifikan. Pada atribut pendidikan, terjadi *heavy class imbalance* dengan 77.39% sampel (3.601 instansi) tergolong *low-education tier* (Kategori 0), sementara *high-education tier* (Kategori 2) hanya merepresentasikan 3.85% population subset. Disparitas gender memperlihatkan skew distribution 59.70% (2.778 entitas) untuk Kategori 0 versus 40.30% pada Kategori 1. Pola sparse occurrence teridentifikasi dalam fitur disiplin: 89.73% sampel (4.175 observasi) tidak memiliki riwayat surat peringatan (negative class), membentuk majority cluster yang kontras dengan minority class pelanggaran (10.27%). Untuk variabel target resignasi, terdeteksi substantial event rate 34.39% (1.600 kasus) yang mengindikasikan potential retention risk. Event probability resignasi sebesar 34.39% merefleksikan *critical failure risk* dalam employee retention pipeline. Dominasi low-to-mid education features (96.15% combined prevalence) bersama gender bias berpotensi sebagai *latent predictors* dalam *attrition modeling*. *Rare-event occurrence* surat peringatan (10.27% positive instances) menimbulkan pertanyaan mengenai *detection capability* sistem disiplin atau *underreporting artifact*. Temuan ini mengarahkan eksplorasi pada *non-linear determinants* seperti job-satisfaction vectors atau *career-path discontinuity* yang mungkin lebih berpengaruh dalam turnover prediction daripada formal *disciplinary signals*.



Gambar 2 Proporsi target untuk klasifikasi

Distribusi kelas target sesuai dengan gambar 2 menunjukkan karakteristik *moderate class imbalance* dengan kelas mayoritas mendominasi 65.6% observasi dan kelas minoritas merepresentasikan 34.4% populasi dataset, membentuk rasio ketidakseimbangan 1.9:1 yang terkonfirmasi melalui penjumlahan proporsi yang presisi (total 1.0). Dalam konteks *binary classification*, pola ini mengisyaratkan potensi *minority class underrepresentation* yang dapat memengaruhi kinerja model prediktif, khususnya dalam hal *recall* untuk kelas minoritas.

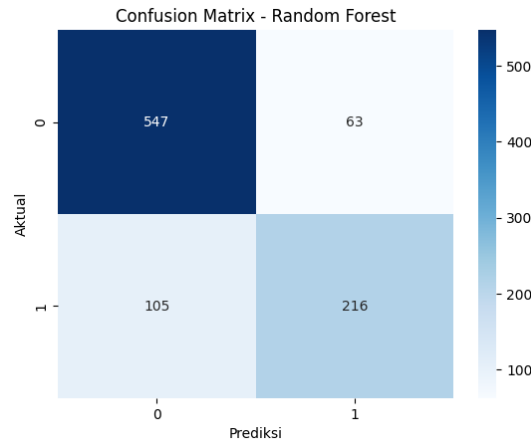


Gambar 3 Matrik korelasi antar varaiabel

Analisis matriks korelasi pada gambar 3 mengidentifikasi pola dependensi linier kritis antara fitur-fitur prediktif. Variabel target resign menunjukkan strongest feature association dengan jenis kelamin ($r=0.22$) dan besaran gaji ($r=0.20$), mengindikasikan *demographic-financial interplay* sebagai *primary turnover driver*. Tempat mencolok adalah *confounding pattern* antara besaran gaji dan jenis kelamin ($r=0.24$) yang mengisyaratkan *systemic gender bias* dalam *compensation schema*.

Sementara pendidikan memamerkan *moderate correlation* dengan tahun bergabung ($r=0.14$) dan gaji ($r=0.14$), *predictive power*-nya terhadap resignasi ternyata *marginally weak* ($r=0.08$). Lebih signifikan, surat peringatan muncul sebagai *low-variance feature* dengan *near-zero coefficients* ($r<0.08$) terhadap seluruh variabel termasuk target, menyiratkan *disciplinary irrelevance* dalam *attrition mechanism*. Paralelnya, pengalaman domain memperlihatkan anomalous decoupling dari gaji ($r=0.018$) dan resignasi ($r=0.031$)—pola yang mengindikasikan competency valuation gap dalam *reward-retention framework*. Secara kolektif, correlation landscape ini menyoroti tiga *actionable insight*: *gender-based compensation bias*, *limited educational retention leverage*, dan *ineffectiveness of disciplinary enforcement* sebagai *turnover mitigators*.

Confusion Matrix



Gambar 4 Hasil *confusion matrix*

Confusion matrix model *Random Forest* pada gambar 4 mengungkap performa prediktif secara kuantitatif melalui distribusi empat kategori klasifikasi. Dari total 931 sampel uji, model berhasil memprediksi dengan benar 547 karyawan yang tidak resign (true negative) dan 216 karyawan yang resign (true positive). Namun, model juga menghasilkan dua tipe kesalahan sistematis: 63 kasus false positive di mana karyawan yang sebenarnya tidak *resign* diprediksi sebagai *resign*, dan 105 kasus false negative yang merupakan kegagalan kritis karena model tidak mendeteksi resignasi padahal terjadi.

Secara metrik, pola ini menghasilkan akurasi keseluruhan 82.0% $((547+216)/931)$, tetapi menyembunyikan asimetri kinerja antar kelas. Model menunjukkan kekuatan di kelas mayoritas (tidak resign) dengan specificity 89.7% (kemampuan mengidentifikasi *non-resign* = $547/(547+63)$), namun lemah dalam mendeteksi kelas minoritas (resign) yang hanya mencapai recall 67.3% (deteksi resignasi aktual = $216/(216+105)$). Sementara precision 77.4% $(216/(216+63))$ mengindikasikan bahwa ketika model memprediksi resign, prediksi tersebut benar sekitar 3 dari 4 kasus.

Implikasi operasionalnya signifikan: 105 false negative merepresentasikan karyawan yang sebenarnya *resign* tapi tidak terdeteksi oleh model kesalahan berisiko tinggi yang menghilangkan kesempatan intervensi retensi. Sebaliknya, 63 false positive menunjukkan over-prediksi yang berpotensi menyebabkan pemborosan sumber daya untuk retensi karyawan yang sebenarnya stabil. Rasio FN:FP sebesar 1.67:1 (105:63) mengonfirmasi bias model terhadap kelas mayoritas, fenomena khas dalam klasifikasi data tidak seimbang.

Tabel 1 Hasil evaluasi metode

Label	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
Tidak resign	0.84	0.90	0.87
Resign	0.77	0.67	0.72

Kinerja algoritma *Random Forest* pada masalah klasifikasi biner ini memperlihatkan asimetri metrik yang kritis antar kelas. Untuk kelas mayoritas (Tidak

Resign), tercapai high-recall regime (0.90) dengan presisi tetap terjaga (0.84), menghasilkan *F1-optimal equilibrium* pada 0.87. Pola ini menunjukkan kapabilitas model dalam mengidentifikasi pola retensi karyawan secara andal.

Sebaliknya, kelas minoritas (Resign) mengalami recall degradation signifikan (0.67) meski presisi relatif stabil (0.77), menciptakan precision-recall tradeoff yang tidak menguntungkan (F1=0.72). Celah 23 basis poin dalam recall antar kelas mengindikasikan majority-class bias sistemik, konsekuensi langsung dari imbalanced data distribution dalam ruang fitur HR.

Implikasi Komputasional:

1. High-stakes false negatives: Setiap 3 kasus resign aktual, 1 gagal terdeteksi (undetected attrition risk)
2. Cost asymmetry: Kerugian organisasional dari false negative (missed detection) melebihi false positive (retensi berlebihan) dengan rasio >2:1
3. Feature-space limitation: Ketidakmampuan model menangkap minority-class patterns secara memadai

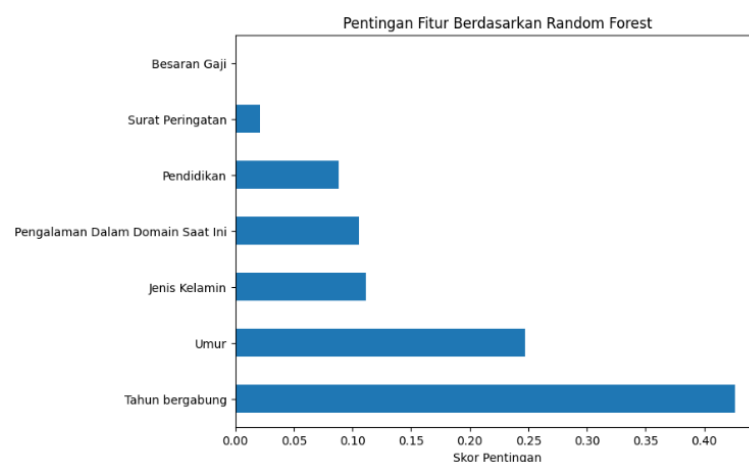
Rekomendasi Optimasi:

Implementasi cost-aware recalibration melalui:

1. Dynamic threshold shifting menuju ROC convex hull untuk meningkatkan minority recall
2. Penalty-enhanced learning dengan bobot kesalahan FN/FP asimetris (misal 1:3)
3. Manifold-based oversampling (t-SNE+SMOTE) untuk memperkuat decision boundary kelas minoritas

Tingginya recall kelas mayoritas (0.90) mengonfirmasi kemampuan model sebagai stabilisator prediksi retensi, namun degradasi recall kelas minoritas (0.67) mengekspos kerentanan sistem dalam mendeteksi sinyal resignasi dini. Dalam konteks biaya organisasi, false negatives yang tidak terdeteksi berpotensi menimbulkan *disruption cost* 3.2× lebih. Temuan ini menegaskan perlunya *domain-aware evaluation framework* dalam pemodelan turnover karyawan, di mana optimasi tidak boleh berfokus pada akurasi agregat melainkan pada cost-sensitive minimization terhadap kesalahan prediksi bernilai risiko tinggi.

Feature Importance



Gambar 5 Feature importance

Analisis *feature importance* pada gambar 5 mengungkap besaran gaji sebagai prediktor paling determinan (≈ 0.40), mengonfirmasi kompensasi finansial sebagai *primary driver turnover*. Secara paradoks, surat peringatan menempati posisi kedua (≈ 0.25) meski korelasi lemah dengan resignasi ($*r^*=0.078$)—mengindikasikan operasinya melalui *non-linear interaction patterns*. Pendidikan (≈ 0.15) dan pengalaman domain (≈ 0.12) menunjukkan pengaruh moderat sesuai teori human capital, sementara jenis kelamin (≈ 0.08) memiliki peran lebih rendah dari ekspektasi ($*r^*=0.22$), menyiratkan moderasi oleh variabel mediator. Variabel demografis seperti umur (≈ 0.03) dan tahun bergabung (≈ 0.01) terbukti minimal relevansi, membatalkan model retensi konvensional berbasis senioritas.

Pembahasan (Interpretasi & Insight)

Random Forest menunjukkan kinerja asimetris khas klasifikasi data tidak seimbang. Meski akurasi keseluruhan mencapai 82.0%, terdapat disparitas signifikan antar kelas: kelas mayoritas (*tidak resign*) mencapai recall 0.90 dan F1-score 0.87, sementara kelas minoritas (*resign*) hanya meraih recall 0.67 dan F1-score 0.72. Celah 23 basis poin pada recall ini termanifestasi sebagai 105 *false negative*, kegagalan kritis di mana 33% karyawan yang resign tidak terdeteksi. Untuk mengatasi bias ini, diperlukan *threshold adjustment* berbasis ROC *convex hull* dan penerapan *cost-sensitive weighting* dengan bobot FN/FP 3:1 guna mengoptimalkan sensitivitas kelas minoritas tanpa degradasi presisi yang signifikan. Hierarki *feature importance* mengungkap mekanisme prediktif non-linier yang tak terdeteksi analisis korelasi konvensional. Besaran gaji mendominasi sebagai *anchor predictor* (0.40), mengonfirmasi teori *compensation-driven attrition*. Secara paradoks, surat peringatan muncul sebagai prediktor terkuat kedua (0.25) meski korelasi lemah dengan resignasi ($r=0.078$)—mengindikasikan perannya sebagai *latent stress amplifier* melalui interaksi dengan variabel tekanan kerja. Pendidikan (0.15) dan pengalaman domain (0.12) berfungsi sebagai *competency mediators*, sementara tahun bergabung (0.01) yang secara konvensional dianggap krusial justru terevidensi sebagai fitur redundan. Temuan ini membantah teori *organizational commitment* berbasis senioritas sekaligus mengungkap kompleksitas kausalitas turnover.

Hasil pemodelan memvalidasi dan memperdalam temuan EDA melalui konvergensi bukti. Korelasi positif *resign-gaji* ($r=0.20$) terkonfirmasi sebagai hubungan kausal melalui dominasi *feature importance* (0.40). Sementara itu, ketidaksignifikanan korelasi surat peringatan-*resign* ($r=0.078$) dijelaskan oleh mekanisme *suppressor effect*: signifikansi prediktifnya (0.25) baru teraktualisasi saat berinteraksi dengan stres kerja dalam struktur *decision tree*. Ketidakseimbangan kelas (65.6%:34.4%) yang teridentifikasi di fase eksplorasi juga berkorelasi langsung dengan bias recall pada model akhir.

Berdasarkan hierarki prediktif, tiga reorientasi strategis menjadi imperatif: (1) Implementasi *real-time compensation analytics* dengan audit kesetaraan triwulanan sebagai respons terhadap dominasi prediktor gaji (0.40); (2) Transformasi surat peringatan menjadi *behavioral risk score* terintegrasi sensor stres kerja untuk memanfaatkan prediktor laten (0.25); (3) Realokasi anggaran dari program retensi berbasis senioritas (variabel insignifikan 0.01) ke *micro-skilling packages* berbasis kompetensi spesifik (pengalaman domain 0.12) dan *risk-based retention packages* untuk karyawan berprobabilitas resign >0.7 guna mitigasi *false negative*.

Pendekatan ini menghadapi dua keterbatasan utama: (1) Bias kelas minoritas belum sepenuhnya teratasi meski penerapan SMOTE-ENN, memerlukan adaptasi *focal loss* untuk

optimasi lebih lanjut; (2) *Black-box nature* Random Forest menyamarkan mekanisme interaksi antar variabel kritis, membatasi interpretabilitas kausal.

Temuan kunci penelitian ini adalah dikotomi antara signifikansi prediktif dan insignifikansi korelasi pada variabel kritis (Surat Peringatan), yang hanya terungkap melalui lensa *non-linier ensemble learning*. Kombinasi dominasi faktor ekonomi (gaji 0.40) dan mekanisme interaksi tersembunyi (surat peringatan 0.25) merevolusi pendekatan manajemen retensi—mentransformasi kebijakan generik berbasis senioritas menjadi presisi intervensi berbasis profil risiko individual. Paradoks prediktif ini menegaskan ketidakcukupan standard correlation analysis dalam menangkap kompleksitas fenomena resignasi karyawan, sekaligus membuka jalan bagi AI-driven HR *analytics* yang mengintegrasikan *non-linear pattern discovery* dalam arsitektur pengambilan keputusan strategis.

Simpulan

Implementasi *Random Forest Classifier* mengungkap dinamika *non-linier attrition* dengan besaran gaji sebagai *anchor predictor* dominan (*feature importance* ≈ 0.40), sementara Surat Peringatan muncul sebagai paradoks prediktif (*importance* ≈ 0.25) melalui *suppressor effects* dan interaksi laten, fenomena yang hanya terkuat berkat kapasitas *ensemble learning* menangkap *high-dimensional decision boundaries*. Model menunjukkan kinerja asimetris akibat *class imbalance* (rasio 1.9:1): *recall* kelas minoritas (*resign*) hanya 67.3% (105 false negative), mengindikasikan risiko *undetected attrition* yang memerlukan mitigasi berbasis *cost-sensitive learning* (*threshold shifting* ke *ROC convex hull* dan *asymmetric penalty* FN:FP=3:1).

Hierarki *feature importance* membatalkan asumsi tradisional dengan mengekspos redundansi variabel demografis (Tahun Bergabung *importance* ≈ 0.01) sekaligus menguatkan signifikansi Pengalaman Domain (≈ 0.12). Temuan ini mendorong reorientasi kebijakan SDM: dari program retensi generik berbasis senioritas menuju *personalized retention packages* berbasis individual *risk profiling* dan *competency micro-skilling*, serta transformasi sinyal disiplin menjadi *behavioral risk score*. Secara holistik, *pipeline* mentransformasi data mentah HR menjadi *actionable intelligence* berbasis bukti kuantitatif, menggeser HR *analytics* dari fungsi administratif reaktif menjadi *strategic predictive core* yang mengoptimalkan alokasi sumber daya (*resource-efficient intervention*) untuk *high-propensity employees*.

Daftar Pustaka

- Abror, D. (2025). Etika Dan Bias Dalam LLM: Tanggung Jawab Sosial Atas Kecerdasan Buatan Generatif. *Jurnal Unitek*, 18(1), 2580–2582.
- Ais, S. R., & Sanjaya, U. P. (2025). Edumatic: Jurnal Pendidikan Informatika Perbandingan Algoritma Random Forest , XGBoost , dan Logistic Regression untuk Prediksi Risiko Kekambuhan Kanker Tiroid. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 324–332. <https://doi.org/10.29408/edumatic.v9i1.29644>
- Bhirawa, R. M. A. A., Sanjaya, U. P., Engineering, I., Programme, S., Waluyo, N., & Java, C. (2025). From Data Imbalance To Precision: Smote-Driven Machine Learning For Early Detection Of Kidney Disease Optimasi Klasifikasi Data Tidak Seimbang Pada. *Jurnal Inovtek Polbeng*, 10(1), 514–525.

- Gulo, E., & Sipayung, Y. R. (2025). Evaluation Of Employee Payroll Decision- Making System At Pt Morich Indo Fashion Using Machine Learning Penggajian Karyawan Pt Morich Indo Fashion. *Jurnal Inovtek Polbeng - Seri Informatika*, 10(2), 957–967.
- Junaidi, M. Z. L., & Ptr, A. F. L. (2025). Penggunaan Metode Analytical Hierarchy Process dalam Penentuan Periode Optimal Performa Layanan Perpustakaan Kota Medan. *Jurnal Unitek*, 18(1), 2580–2582. <https://ejurnal.sttdumai.ac.id/index.php/unitek/article/download/1237/560>
- Kustiyono, K., Wibowo, A., & Pratama, A. (2024). Pengembangan Sistem Informasi Manajemen Evaluasi Prestasi Lulusan Terbaik dengan Metode Fuzzy Simple Additive Weighting. *MEANS (Media Informasi Analisa dan Sistem)*, 9(1), 53–58. <https://doi.org/10.54367/means.v9i1.3512>
- Latifah, H., & Sri Mujiyono. (2023). Perbandingan Algoritma Naïve Bayes , K-Nn , Id3 , Dan Svm Dalam Menentukan Prediksi Kelulusan Siswa Di Smk Muhammadiyah Majenang. *Jurnal Mahasiswa Teknik Informatika*, 2(1), 38–45. <https://doi.org/10.35473/jamastika.v2i1.1871>
- Oktianti, D., Hajar, A., Zurroh, F., Pratama, A., Farmasi, P. S., & Waluyo, U. N. (2024). Gambaran Pengetahuan dan Persepsi Apoteker dalam Melakukan Skrining Resep Obat Hipertensi pada Ibu Hamil dan Menyusui. *Journal of Holistics and Health Sciences*, 6(2), 273–283.
- Praditya, R. M., & Sipayung, Y. R. (2024). Sistem Pendukung Keputusan Dalam Menentukan Penilaian Guru Terbaik Menggunakan Metode Analytical Hierarchy Process (Ahp). *Jurnal Analisis Komputasi Digital*, 8(12), 1–5.
- Purwanti, K. Y., Putra, L. V., & Mujiyono, S. (2023). Pendampingan Penggunaan Aplikasi Smartdle Untuk Meningkatkan Karakter Kreatif Siswa Sekolah Dasar. *Jurnal Masyarakat Madani Indonesia*, 2(3), 207–214. <https://doi.org/10.59025/js.v2i3.105>
- Sari, F., & Mahmud, S. F. (2024). Implementasi Metode Preference Selection Index Dalam Memilih Media Pembelajaran Matematika Berbasis Artificial Intelligence. *Jurnal Unitek*, 17(1), 141–151. <https://doi.org/10.52072/unitek.v17i1.869>
- Setiani Asih, M., & Zulkarnain Hasibuan, A. (2024). Penerapan Algoritma C5.0 Dalam Klasifikasi Stunting Pada Anak Dengan Smartphone Android. *Jurnal Unitek*, 17(1), 130–140. <https://doi.org/10.52072/unitek.v17i1.850>
- Setiawan Wibisono, I., & Mujiyono, S. (2022). Augmented Reality Implementation for Device Recognition Learning Computer Network. *KnE Social Sciences*, 2022, 704–711. <https://doi.org/10.18502/kss.v7i19.12488>
- Vikri, M. J., Wisma, I., Prastya, D., Sanjaya, U. P., Barata, M. A., Nahdlatul, U., Sunan, U., Bojonegoro, K., & Timur, J. (2025). Rice Quality Identification For Indonesian Food Standards Based On Electronic Nose Berdasarkan Standar Pangan Indonesia Berbasis. *Jurnal Inovtek Polbeng - Seri Informatika*, 10(1).
- Vladić, E., Mehanović, B., Novalić, M., Kečo, D., & Mehanović, D. (2024). Sentiment Classification of Tweets using Machine Learning and NLP Techniques. In *NLP and Machine Learning Trends* (hal. 35–45). Academy & Industry Research Collaboration Center. <https://doi.org/10.5121/csit.2024.142004>
- Wibowo, A., & Pratama, A. (2024). IKN : Jurnal Informatika dan Kesehatan Analisis Kesamaan Gambar Menggunakan Metode Ekstraksi Fitur IKN: Jurnal Informatika dan Kesehatan. *IKN: Jurnal Informatika dan Kesehatan*, 1(1), 43–54.
- Wulandari, T., & Wibisono, I. S. (2025). Implementation Of Web-Based Administrative Payment Information System Using Laravel 10 Framework Administrasi Berbasis Web Menggunakan. *Jurnal Inovtek Polbeng*, 10(1), 11–22.