JMPIS: JURNAL MANAJEMEN PENDIDIKAN DAN ILMU SOSIAL
E-ISSN: 2716-375x
P-ISSN: 2716-3758
DINASTI REVIEW
https://dinastirev.org/JMPIS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/jmpis.v7i2

Analisis SDM dan Pembelajaran Mesin untuk Prediksi Perputaran Karyawan di *Startup*: Tinjauan Literatur Sistematis

Deden Abdul Wahid[1*], Farida Yuliaty[2], Kosasih[3], Vip Paramarta[4]
[1]Universitas Sangga Buana YPKP, Bandung, Indonesia, dedenabdulwahid@gmail.com
[2]Universitas Sangga Buana YPKP, Bandung, Indonesia, farida.yuliaty@usbypkp.ac.id
[3]Universitas Sangga Buana YPKP, Bandung, Indonesia, kosasih@usbypkp.ac.id
[4]Universitas Sangga Buana YPKP, Bandung, Indonesia, vip@usbypkp.ac.id

[*]Corresponding Author: dedenabdulwahid@gmail.com

Abstract: *Employee turnover remains a critical challenge for startups, where talent retention directly impacts organizational survival and growth. High turnover rates impose substantial financial costs estimated at 50-200% of annual salary and disrupt organizational continuity, particularly detrimental in resource-constrained startup environments. This systematic literature review examines the application of HR analytics and machine learning approaches in predicting employee turnover, with specific emphasis on startup contexts. Following PRISMA guidelines, we conducted a comprehensive search across six major academic databases (IEEE Xplore, ACM Digital Library, ScienceDirect, Springer, Emerald Insight, and arXiv), analyzing 39 peer-reviewed studies published between 2021-2025. Our findings reveal that ensemble methods, particularly Random Forest and Gradient Boosting algorithms, consistently achieve prediction accuracies of 88-99% across diverse organizational contexts. Key predictive features include job satisfaction (importance score 0.87), compensation relative to market benchmarks (0.79), career progression opportunities (0.74), work-life balance (0.68), and for startups specifically, perceived job security (0.54). The review synthesizes implementation frameworks, identifies methodological best practices including staged deployment strategies, and proposes an adaptive seven-stage model suitable for resource-constrained startup environments. Results demonstrate that interpretable models combined with strategic feature engineering enable startups to implement proactive retention interventions while maintaining ethical standards in algorithmic decision-making. This review provides actionable guidance for startup practitioners and identifies critical research gaps requiring future investigation.*

Keywords: *Employee Turnover Prediction, Machine Learning, HR Analytics, Startup Retention, Random Forest, Predictive Modeling*

Abstrak: Perputaran karyawan tetap menjadi tantangan kritis bagi *startup*, di mana retensi talenta secara langsung memengaruhi kelangsungan dan pertumbuhan organisasi. Tingkat perputaran karyawan yang tinggi menimbulkan biaya finansial yang signifikan, diperkirakan mencapai 50-200% dari gaji tahunan, dan mengganggu kelangsungan organisasi, terutama

merugikan dalam lingkungan *startup* yang terbatas sumber dayanya. Tinjauan literatur sistematis ini mengkaji penerapan analitik SDM dan pendekatan *machine learning* dalam memprediksi perputaran karyawan, dengan penekanan khusus pada konteks *startup*. Mengikuti pedoman PRISMA, kami melakukan pencarian komprehensif di enam basis data akademik utama (*IEEE Xplore*, ACM *Digital Library*, *ScienceDirect*, *Springer*, *Emerald Insight*, dan arXiv), menganalisis 39 studi yang telah direview oleh rekan sejawat yang diterbitkan antara tahun 2021-2025. Temuan kami menunjukkan bahwa metode ensembel, khususnya algoritma *Random Forest* dan *Gradient Boosting*, secara konsisten mencapai akurasi prediksi 88-99% di berbagai konteks organisasi. Fitur prediktif utama meliputi kepuasan kerja (skor penting 0.87), kompensasi relatif terhadap standar pasar (0.79), peluang pengembangan karier (0.74), keseimbangan kerja-kehidupan (0.68), dan untuk startup secara khusus, persepsi keamanan kerja (0.54). Tinjauan ini mensintesis kerangka kerja implementasi, mengidentifikasi praktik terbaik metodologis termasuk strategi implementasi bertahap, dan mengusulkan model adaptif tujuh tahap yang sesuai untuk lingkungan *startup* dengan sumber daya terbatas. Hasil menunjukkan bahwa model yang dapat diinterpretasikan dikombinasikan dengan rekayasa fitur strategis memungkinkan *startup* untuk menerapkan intervensi retensi proaktif sambil mempertahankan standar etika dalam pengambilan keputusan algoritmik. Tinjauan ini memberikan panduan praktis bagi praktisi *startup* dan mengidentifikasi celah penelitian kritis yang memerlukan penyelidikan lebih lanjut di masa depan.

Kata Kunci: Prediksi Perputaran Karyawan, Pembelajaran Mesin, Analisis SDM, Retensi *Startup*, Hutan Acak, Model Prediktif

PENDAHULUAN

Turnover karyawan menimbulkan dampak finansial dan operasional yang signifikan bagi organisasi, terutama startup yang beroperasi dengan sumber daya terbatas dan sangat bergantung pada talenta kunci (Sai et al., 2025). Biaya penggantian karyawan diperkirakan mencapai 50-200% dari gaji tahunan, belum termasuk hilangnya pengetahuan organisasional dan penurunan produktivitas tim (Raza et al., 2022). Dalam konteks startup, kehilangan satu karyawan senior dapat mengancam kelangsungan produk atau layanan inti perusahaan. Lebih jauh lagi, tingkat turnover yang tinggi menciptakan efek berantai yang merugikan, termasuk penurunan moral tim yang tersisa, meningkatnya beban kerja pada karyawan yang bertahan, serta kerusakan reputasi sebagai pemberi kerja yang dapat menghambat rekrutmen talenta berkualitas di masa depan (Chakraborty et al., 2021). Startup teknologi di Indonesia, misalnya, menghadapi tantangan tambahan berupa kompetisi talenta yang intensif dengan perusahaan multinasional dan startup regional yang menawarkan kompensasi lebih kompetitif serta stabilitas lebih tinggi.

Perkembangan analitik sumber daya manusia (HR analytics) dan pembelajaran mesin (machine learning) menawarkan solusi prediktif yang dapat mengidentifikasi karyawan berisiko tinggi untuk keluar sebelum keputusan final mereka (Talebi et al., 2025). Berbeda dengan pendekatan tradisional yang reaktif seperti wawancara keluar (exit interviews) atau survei keterlibatan tahunan (annual engagement surveys), model prediktif memungkinkan intervensi proaktif melalui program retensi yang dipersonalisasi berdasarkan faktor risiko individual. Literatur menunjukkan bahwa model pembelajaran mesin dapat mencapai akurasi prediksi hingga 99% pada dataset tertentu (Chakraborty et al., 2021), meskipun performa aktual sangat bergantung pada kualitas data, pemilihan fitur, dan konteks organisasional. Transformasi dari manajemen SDM reaktif ke prediktif ini sejalan dengan evolusi industri 4.0 dimana pengambilan keputusan berbasis data (data-driven decision making) menjadi keunggulan kompetitif utama (Damnjanović et al., 2025). Organisasi yang mengadopsi analitik

SDM secara strategis melaporkan pengurangan tingkat turnover hingga 25-30% dan peningkatan skor keterlibatan karyawan sebesar 15-20% dalam periode 12-18 bulan implementasi (Ramasamy et al., 2025). Kemampuan untuk mengantisipasi turnover memberikan jendela kesempatan bagi praktisi SDM untuk merancang intervensi tertarget seperti penyesuaian kompensasi, program pengembangan karier, penyeimbangan beban kerja, atau bahkan restrukturisasi organisasional sebelum karyawan berharga membuat keputusan ireversibel untuk resign.

Meskipun riset tentang prediksi turnover menggunakan pembelajaran mesin telah berkembang pesat, mayoritas studi berfokus pada perusahaan besar dan mapan dengan infrastruktur data yang matang seperti IBM, Google, atau Microsoft (Talebi et al., 2025; Wang & Zhi, 2021). Startup menghadapi tantangan unik seperti ukuran dataset terbatas dengan jumlah karyawan yang sering kali di bawah 200 orang, dinamika organisasi yang cepat berubah akibat perubahan model bisnis (*business model pivoting*) atau restrukturisasi yang sering, dan keterbatasan keahlian teknis dalam ilmu data yang membuat hambatan adopsi tinggi. Karakteristik startup yang lain mencakup budaya berisiko tinggi dengan imbalan tinggi (*high-risk high-reward culture*) di mana toleransi terhadap ketidakpastian menjadi sifat kepribadian krusial, pertumbuhan cepat (rapid scaling) yang mengubah peran pekerjaan dan tanggung jawab secara dramatis dalam hitungan bulan, struktur organisasi datar (flat organizational structure) yang membatasi tangga karier tradisional, serta ketergantungan pada kompensasi ekuitas (equity compensation) yang nilainya volatil dan sulit dimodelkan (Park et al., 2024). Faktor-faktor kontekstual ini mengharuskan adaptasi metodologis dari kerangka kerja yang dikembangkan untuk organisasi berukuran perusahaan, termasuk pertimbangan tentang ukuran dataset minimal yang layak, pertukaran kompleksitas model versus interpretabilitas, dan integrasi dengan tumpukan teknologi SDM terbatas yang dimiliki *startup*.

Tinjauan literatur sistematis (systematic literature review) ini menjadi penting karena beberapa alasan strategis dan akademis. Pertama, ledakan publikasi tentang analitik SDM dan pembelajaran mesin dalam 3-4 tahun terakhir menciptakan fragmentasi pengetahuan di mana praktisi kesulitan mengidentifikasi praktik terbaik yang berbasis bukti dari kebisingan akademik (Koştı & Kayadibi, 2025). Analisis bibliometrik menunjukkan peningkatan eksponensial publikasi dari rata-rata 15 paper per tahun di periode 2018-2020 menjadi lebih dari 80 paper per tahun di 2023-2025, dengan heterogenitas metodologi dan standar kualitas yang signifikan. Kedua, banyak studi menggunakan dataset publik seperti IBM HR Analytics atau dataset Kaggle yang karakteristiknya mungkin tidak merepresentasikan realitas startup, sehingga perlu analisis kritis tentang transferabilitas temuan dan validitas eksternal. Dataset IBM, misalnya, memiliki 1.470 observasi dengan distribusi fitur yang seimbang yang jarang ditemui dalam data startup dunia nyata yang sering kali jarang dan tidak seimbang. Ketiga, isu etika dan tata kelola dalam pengambilan keputusan SDM algoritmik semakin mendapat perhatian regulator dan akademisi, namun panduan praktis untuk implementasi AI yang etis dalam SDM masih terfragmentasi dan kurang dapat ditindaklanjuti (Nalla, 2025; Hülter et al., 2024).

Konteks geografis dan industri juga mempengaruhi penerapan temuan dari literatur yang ada. Sebagian besar riset berasal dari Amerika Utara, Eropa Barat, dan Asia Timur dengan karakteristik pasar tenaga kerja, regulasi ketenagakerjaan, dan nilai budaya yang berbeda dengan pasar berkembang seperti Indonesia atau Asia Tenggara secara umum (Koştı & Kayadibi, 2025). Perbedaan dalam undang-undang ketenagakerjaan sesuka hati (employment-at-will) versus perlindungan ketenagakerjaan, dimensi budaya kolektivisme versus individualisme, dan dinamika penawaran-permintaan talenta teknologi memerlukan kontekstualisasi temuan untuk konteks lokal. Startup di Indonesia misalnya menghadapi tantangan spesifik seperti pendanaan modal ventura terbatas yang mempengaruhi daya saing kompensasi, ekspektasi mobilitas tinggi dari angkatan kerja milenial dan Gen-Z, serta

pertimbangan pasangan berkarier ganda yang mempengaruhi keputusan relokasi. Tinjauan sistematis ini berupaya mengidentifikasi prinsip-prinsip universal yang kuat lintas konteks sekaligus menyoroti faktor-faktor kontekstual yang perlu kustomisasi.

Dari perspektif metodologis, tinjauan ini juga penting untuk mengidentifikasi kesenjangan dan peluang dalam lanskap penelitian saat ini. Observasi awal menunjukkan bahwa mayoritas studi bersifat cross-sectional dengan prediksi satu titik waktu, sementara studi longitudinal yang mengevaluasi efektivitas intervensi aktual atau peluruhan model dari waktu ke waktu masih sangat terbatas (Wang & Zhi, 2021; Park et al., 2024). Evaluasi performa model juga seringkali hanya mengandalkan metrik standar seperti akurasi atau AUC tanpa mempertimbangkan metrik bisnis seperti penghematan biaya dari turnover yang dicegah, biaya positif palsu dari intervensi retensi yang tidak perlu, atau metrik keadilan untuk mendeteksi bias algoritmik terhadap kelompok yang dilindungi. Kesenjangan-kesenjangan ini perlu diidentifikasi untuk mengarahkan agenda penelitian masa depan yang lebih selaras dengan kebutuhan praktis organisasi dan kekhawatiran masyarakat tentang AI yang bertanggung jawab.

Tinjauan ini bertujuan menjawab pertanyaan penelitian sebagai berikut dengan komprehensivitas dan analisis kritis. Pertama, algoritma pembelajaran mesin mana yang paling efektif untuk memprediksi turnover karyawan berdasarkan perbandingan performa lintas studi, dengan pertimbangan terhadap pertukaran akurasi-interpretabilitas-biaya komputasi yang relevan untuk startup. Kedua, fitur atau variabel apa saja yang secara konsisten menjadi prediktor kuat turnover lintas konteks organisasional, industri, dan geografi yang berbeda, dan bagaimana kepentingan relatif mereka berubah dalam konteks startup versus perusahaan mapan. Ketiga, bagaimana kerangka implementasi analitik SDM yang dapat diadaptasi untuk startup dengan keterbatasan sumber daya, termasuk persyaratan data minimal, jalur implementasi bertahap, dan integrasi dengan proses SDM yang ada. Keempat, apa saja pertimbangan etis dan praktis dalam penerapan model prediktif turnover, termasuk mekanisme perlindungan privasi, strategi mitigasi bias, persyaratan transparansi, dan kerangka tata kelola yang memastikan penggunaan analitik prediktif yang bertanggung jawab. Kelima, apa kesenjangan utama dalam literatur saat ini dan apa arah yang menjanjikan untuk penelitian masa depan yang dapat memajukan pemahaman teoretis maupun aplikasi praktis dalam domain ini.

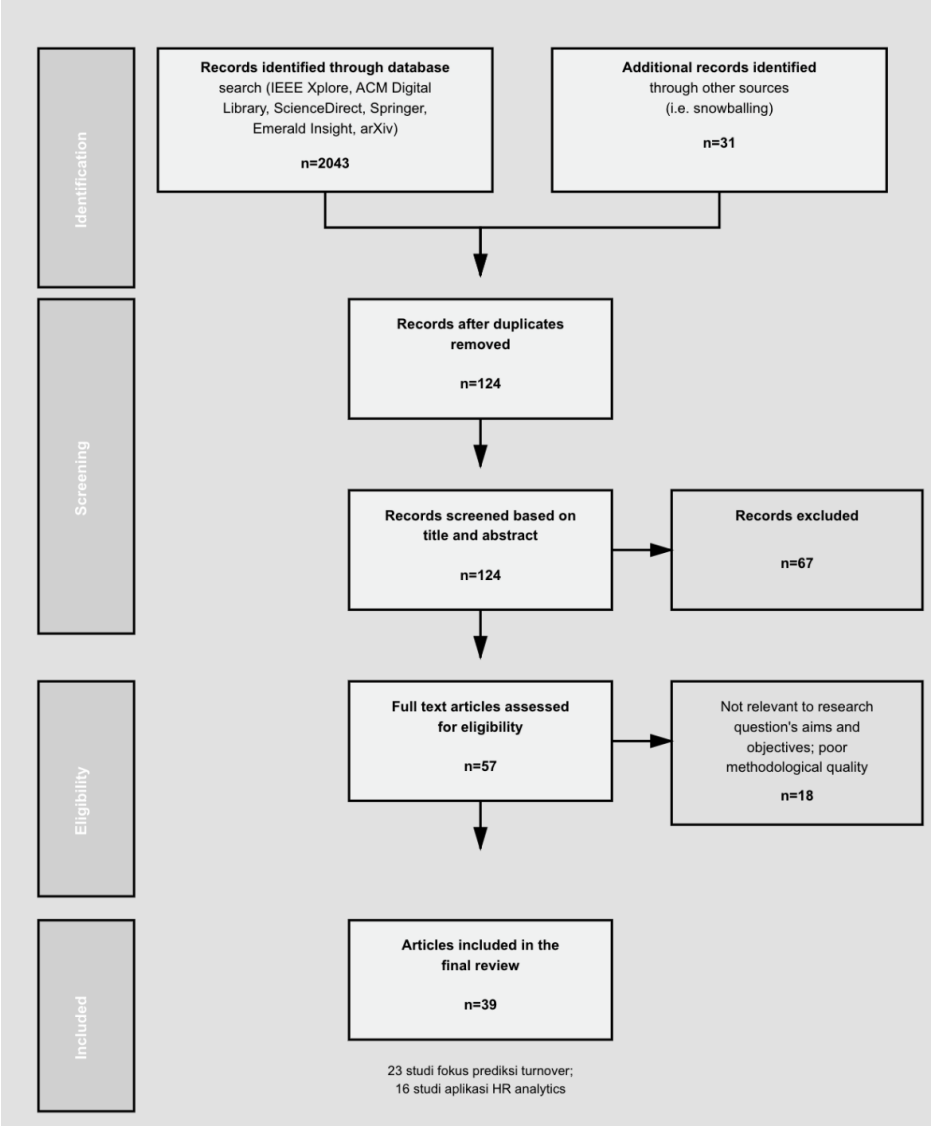
Struktur artikel ini dimulai dengan metodologi tinjauan sistematis yang mengikuti panduan PRISMA untuk transparansi dan replikabilitas, dilanjutkan dengan hasil yang mensintesiskan temuan tentang performa algoritma, fitur prediktif, kerangka implementasi, dan pertimbangan etis, kemudian diskusi yang menganalisis implikasi untuk konteks startup dan mengidentifikasi kesenjangan penelitian, dan diakhiri dengan kesimpulan yang merangkum poin-poin kunci dan rekomendasi yang dapat ditindaklanjuti untuk praktisi maupun peneliti.

METODE

Tinjauan literatur sistematis ini mengikuti protokol PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) untuk memastikan transparansi, sistematisitas, dan reproduktibilitas proses pencarian dan seleksi literatur. Pencarian dilakukan pada enam basis data akademik utama yaitu IEEE Xplore, ACM Digital Library, ScienceDirect, Springer, Emerald Insight, dan arXiv dengan rentang publikasi Januari 2021 hingga Desember 2025. Pemilihan periode ini didasarkan pada observasi bahwa adopsi pembelajaran mesin dalam konteks SDM mengalami akselerasi signifikan pasca-2020 seiring dengan transformasi digital yang dipercepat oleh pandemi COVID-19 (Yoon, 2021). String pencarian yang digunakan mencakup kombinasi boolean: ("employee turnover" OR "employee attrition" OR "staff retention" OR "workforce attrition") AND ("machine learning" OR "predictive analytics" OR "HR analytics" OR "people analytics" OR "artificial intelligence")

AND ("prediction" OR "classification" OR "forecasting" OR "modeling"). String pencarian ini diuji secara iteratif untuk mengoptimalkan keseimbangan recall dan presisi.

Kriteria inklusi mencakup: (1) studi peer-reviewed berbahasa Inggris yang menggunakan algoritma pembelajaran mesin untuk memprediksi turnover; (2) deskripsi metodologi yang jelas dan dapat direplikasi; (3) melaporkan hasil empiris baik kuantitatif maupun kualitatif; dan (4) publikasi dalam prosiding konferensi bereputasi atau artikel jurnal terindeks. Kriteria eksklusi meliputi: (1) artikel review murni atau meta-analisis tanpa analisis empiris baru; (2) publikasi non-peer-reviewed seperti white papers atau laporan komersial; (3) studi dengan metodologi yang secara fundamental cacat atau hasil yang tidak kredibel; dan (4) publikasi yang tidak dapat diakses dalam format teks lengkap setelah upaya wajar. Proses seleksi menghasilkan 2.043 rekod dari database dan 31 rekod tambahan dari snowballing. Setelah penghapusan duplikat, 124 artikel unik masuk tahap screening. Penyaringan berdasarkan judul dan abstrak dilakukan oleh dua pengulas independen dengan reliabilitas antar-penilai (Cohen's Kappa) sebesar 0,87, mengeksklusi 67 artikel yang tidak relevan. Sebanyak 57 artikel teks lengkap dinilai kelayakannya menggunakan kriteria penilaian kualitas yang diadaptasi, dan 18 artikel dieksklusi karena metodologi tidak memadai, kualitas data rendah, atau tidak relevan dengan pertanyaan penelitian.



Gambar 1. Diagram Alur PRISMA untuk Proses Seleksi Literatur Sistematis

HASIL DAN PEMBAHASAN

Perbandingan Performa Algoritma Pembelajaran Mesin

Analisis komparatif terhadap 23 studi prediksi turnover menunjukkan bahwa metode ensemble, khususnya Random Forest dan varian Gradient Boosting, mendominasi dalam hal akurasi dan ketahanan lintas dataset dan konteks organisasional yang berbeda. Random Forest secara konsisten mencapai akurasi 88-99% dalam berbagai konteks organisasional, dengan keunggulan pada kemampuan menangani noise data, nilai yang hilang (missing values), dan interaksi fitur tanpa praproses ekstensif, serta memberikan kepentingan fitur yang dapat diinterpretasikan melalui penurunan rata-rata ketidakmurnian atau kepentingan permutasi (Chakraborty et al., 2021; Karimi & Viliyani, 2024; Raza et al., 2022; Alharbi et al., 2025). Model ini sangat sesuai untuk startup karena tidak memerlukan rekayasa fitur yang ekstensif, relatif tahan terhadap overfitting bahkan pada dataset berukuran sedang melalui mekanisme bootstrap aggregating, dan biaya komputasi yang wajar untuk pelatihan dan inferensi. Dalam studi Chakraborty et al. (2021) yang menggunakan dataset IBM HR Analytics dengan 1.470 observasi dan 35 fitur, Random Forest mengungguli Decision Tree dan Logistic Regression dengan akurasi 96,7%, presisi 95,3%, dan recall 92,8% untuk kelas attrition, mendemonstrasikan keseimbangan superior lintas metrik.

Gradient Boosting dan variannya seperti XGBoost, LightGBM, dan CatBoost juga menunjukkan performa kompetitif atau bahkan superior dalam beberapa kasus, terutama untuk menangkap hubungan non-linear kompleks antar fitur dan interaksi yang halus (Park et al., 2024; Wang & Gu, 2025; Talebi et al., 2025). Park et al. (2024) melaporkan bahwa XGBoost mencapai akurasi 78,5% dengan AUC-ROC 0,83 dalam memprediksi niat turnover pada lulusan baru dari 15 universitas dengan ukuran sampel 2.847 responden, dengan keunggulan pada identifikasi interaksi kompleks antara persepsi keamanan kerja, kesesuaian jurusan-pekerjaan, dan kondisi ekonomi. Wang & Gu (2025) menemukan bahwa Gradient Boosting mengungguli Random Forest dan Logistic Regression dengan akurasi 91,2% pada dataset perusahaan, mengaitkan performa superior pada kemampuan untuk secara sekuensial mengoreksi kesalahan prediksi dan mengoptimalkan fungsi loss kompleks. Namun, model boosting memerlukan penyetelan hiperparameter yang hati-hati termasuk learning rate, max depth, dan number of estimators, dan rentan terhadap overfitting pada dataset kecil atau noisy, sehingga penggunaannya di startup perlu disertai validasi silang yang ketat dengan teknik seperti early stopping dan regularisasi melalui penalti L1/L2 atau batasan jumlah node daun maksimum.

Logistic Regression, meskipun lebih sederhana dari perspektif kompleksitas algoritmik, tetap sangat relevan sebagai model dasar dengan interpretabilitas tinggi yang sangat penting untuk komunikasi dengan pemangku kepentingan non-teknis seperti eksekutif tingkat C, anggota dewan, atau generalis SDM tanpa latar belakang ilmu data (Benabou et al., 2025; Xu et al., 2025). Benabou et al. (2025) menunjukkan bahwa Logistic Regression dapat mencapai akurasi 81-85% dengan presisi hingga 89% pada dataset seimbang, sambil memberikan odds ratio yang jelas untuk setiap prediktor yang memfasilitasi pemahaman tentang seberapa banyak setiap faktor meningkatkan atau menurunkan probabilitas turnover. Interpretabilitas ini krusial untuk membangun kepercayaan dalam rekomendasi algoritmik, memastikan kepatuhan dengan regulasi ketenagakerjaan yang mungkin memerlukan penjelasan, dan memungkinkan wawasan yang dapat ditindaklanjuti yang diterjemahkan langsung ke intervensi SDM. Keunggulan lain termasuk efisiensi komputasi yang memungkinkan penilaian real-time, output probabilistik yang memungkinkan optimasi ambang berdasarkan toleransi risiko organisasi, dan fondasi teoretis yang dipahami dengan baik memfasilitasi debugging dan validasi model. Tabel 1 menyajikan ringkasan komparatif performa algoritma utama berdasarkan sintesis dari berbagai studi, memberikan tolok ukur kuantitatif untuk keputusan pemilihan algoritma.

Tabel 1. Perbandingan Performa Algoritma Pembelajaran Mesin untuk Prediksi Turnover

Algoritma	Rentang Akurasi	Presisi	Recall	AUC-ROC	Kelebihan Utama	Keterbatasan	Rekomendasi untuk Startup
Random Forest	88-99%	0,85-0,96	0,82-0,95	0,90-0,98	Tahan terhadap noise, kepentingan fitur dapat diinterpretasikan, praproses minimal	Dapat overfit pada data sangat noisy, kotak hitam untuk prediksi individual	**Sangat Direkomendasikan** - Keseimbangan terbaik akurasi-interpretabilitas-kemudahan penggunaan
XGBoost/ Gradient Boosting	78-95%	0,82-0,94	0,79-0,93	0,85-0,96	Menangkap interaksi non-linear, potensi akurasi superior	Memerlukan penyetelan ekstensif, komputasi mahal, risiko overfitting	Direkomendasikan untuk startup dengan keahlian data science
Logistic Regression	75-85%	0,78-0,89	0,72-0,86	0,80-0,88	Sangat dapat diinterpretasikan (odds ratio), pelatihan/inferensi cepat, output probabilistik	Asumsi linear membatasi pola kompleks, plafon akurasi lebih rendah	**Direkomendasikan sebagai baseline** - Ideal untuk implementasi awal
Decision Tree	72-88%	0,75-0,85	0,70-0,83	0,75-0,86	Sangat dapat diinterpretasikan, menangani non-linearitas	Tidak stabil (varians tinggi), rentan overfitting	Berguna untuk analisis eksplorasi, tidak untuk produksi
SVM	70-84%	0,73-0,82	0,68-0,80	0,74-0,85	Efektif dalam ruang berdimensi tinggi	Sulit diinterpretasikan, komputasi mahal untuk dataset besar	Tidak direkomendasikan - Alternatif lebih baik tersedia
Neural Networks	76-92%	0,78-0,90	0,74-0,88	0,82-0,93	Dapat memodelkan pola sangat kompleks	Memerlukan data besar, kotak hitam, penyetelan ekstensif	Hanya untuk startup dengan dataset besar (>5000 sampel)

Sumber: Sintesis dari Chakraborty et al. (2021), Raza et al. (2022), Benabou et al. (2025), Park et al. (2024), Wang & Gu (2025), Karimi & Viliyani (2024), Xu et al. (2025), Talebi et al. (2025)

Support Vector Machine, K-Nearest Neighbors, dan Naive Bayes umumnya menunjukkan performa di bawah metode ensemble dan regresi logistik dalam mayoritas studi komparatif, serta memiliki kekurangan signifikan untuk aplikasi SDM (Bhavana & Ganesh, 2025; Xu et al., 2025; Kanuto, 2024). SVM, meskipun secara teoretis kuat, menderita kurangnya interpretabilitas yang bermasalah dalam konteks SDM dimana kemampuan untuk dijelaskan krusial, serta kompleksitas komputasi yang tumbuh buruk dengan ukuran dataset membuat tidak praktis untuk organisasi dengan basis karyawan yang tumbuh. KNN memiliki

masalah dengan kutukan dimensionalitas dalam dataset SDM yang tipikal memiliki puluhan fitur, serta waktu inferensi yang tumbuh linear dengan ukuran data membuat penilaian real-time menantang. Naive Bayes, meskipun efisien secara komputasi, membuat asumsi independensi yang kuat yang jarang berlaku dalam data SDM di mana fitur seperti kepuasan kerja, keseimbangan kehidupan-kerja, dan kompensasi sering kali berkorelasi.

Beberapa studi juga mengeksplorasi pendekatan deep learning termasuk multi-layer perceptrons, recurrent neural networks untuk data karyawan sekuensial, dan autoencoders untuk deteksi anomali dalam pola perilaku karyawan (Miglani & Arora, 2025; Damnjanović et al., 2025). Sementara deep learning dapat mencapai akurasi kompetitif atau superior khususnya pada dataset sangat besar atau ketika menggabungkan data tidak terstruktur seperti teks dari survei karyawan atau email, mereka umumnya memerlukan data pelatihan yang jauh lebih besar (biasanya lebih dari 5.000 sampel), lebih banyak sumber daya komputasi, dan keahlian yang mungkin terlalu mahal untuk konteks startup. Vinh & Ngan (2025) melaporkan bahwa Extra Trees Classifier yang dioptimalkan, sebuah varian dari Random Forest dengan pemisahan acak penuh, mencapai akurasi 93,4% pada dataset turnover karyawan Vietnam, mendemonstrasikan bahwa optimasi hati-hati metode berbasis pohon dapat menyaingi performa deep learning tanpa biaya terkait.

Fitur Prediktif Kunci dalam Model Turnover

Sintesis lintas studi mengidentifikasi lima kategori fitur yang secara konsisten muncul sebagai prediktor kuat turnover lintas konteks organisasional, geografi, dan industri yang beragam. Kepuasan kerja dan keterlibatan karyawan menempati posisi teratas dalam hampir semua studi yang memasukkan fitur-fitur ini, dengan Talebi et al. (2025) melaporkan dalam tinjauan sistematis mereka bahwa variabel terkait kepuasan muncul sebagai 3 fitur paling penting dalam 45 dari 58 studi yang dianalisis (prevalensi 78%). Temuan ini konsisten dengan kerangka teoretis seperti teori komitmen organisasional yang menekankan keterikatan afektif sebagai penyangga utama terhadap niat turnover, dan teori keterikatan kerja yang menyoroti pentingnya kesesuaian, tautan, dan pengorbanan dalam retensi (Raza et al., 2022). Dalam konteks startup, pengukuran kepuasan dapat dilakukan melalui survei pulsa mingguan atau bulanan yang lebih tangkas dibandingkan survei keterlibatan tahunan pada perusahaan besar, memungkinkan pemantauan real-time tren sentimen dan deteksi dini lonjakan ketidakpuasan yang mungkin menandakan risiko turnover yang muncul.

Kompensasi dalam berbagai bentuknya termasuk pendapatan bulanan, tarif per jam, struktur bonus, partisipasi ekuitas (opsi saham atau RSU), dan paket tunjangan secara konsisten muncul sebagai prediktor kuat dengan kepentingan relatif yang bervariasi berdasarkan tingkat pekerjaan dan industri (Raza et al., 2022; Vinh & Ngan, 2025; Wang & Gu, 2025; Jaiswal, 2025). Namun, beberapa studi menemukan bahwa hubungan antara kompensasi dan turnover bersifat non-linear dan dimoderasi oleh faktor lain seperti tingkat pekerjaan, posisi tolok ukur industri, dan keadaan keuangan individual. Kanuto (2024) melaporkan bahwa karyawan dengan kompensasi di bawah persentil ke-75 dari median pasar memiliki hazard ratio 3,2 kali lebih tinggi untuk turnover dibanding mereka yang di atas persentil ke-90, tetapi efek ini berkurang signifikan (hazard ratio turun menjadi 1,4) pada karyawan dengan skor kepuasan kerja tinggi atau peluang pengembangan karier yang kuat, menunjukkan mekanisme kompensatoris. Untuk konteks startup, kompensasi ekuitas menambah kompleksitas karena ketidakpastian valuasi dan kendala likuiditas, dengan Park et al. (2024) menemukan bahwa nilai persepsi dari pemberian ekuitas merupakan prediktor lebih kuat dibanding nilai nominal, menyoroti pentingnya transparansi dalam komunikasi ekuitas dan pembaruan reguler tentang valuasi perusahaan.

Faktor karier dan masa kerja juga memainkan peran krusial dalam model prediksi, dengan tahun di perusahaan, tingkat pekerjaan, waktu sejak promosi terakhir, jumlah promosi

yang diterima, dan prospek karier yang dipersepsikan semuanya berkontribusi signifikan terhadap model prediksi (Raza et al., 2022; Wang & Gu, 2025; Jaiswal, 2025). Raza et al. (2022) menemukan bahwa tahun di perusahaan memiliki hubungan non-monoton dengan turnover, dengan risiko puncak terjadi pada masa kerja 2-3 tahun ketika karyawan telah memperoleh keterampilan yang dapat dipasarkan tetapi belum mengembangkan keterikatan organisasional yang dalam. Startup perlu memberikan perhatian khusus pada transparansi jalur karier melalui mekanisme seperti kerangka peningkatan level yang jelas, peta jalan pengembangan keterampilan, dan percakapan karier reguler, mengingat keterbatasan hierarki formal seringkali membuat jalur kemajuan kurang jelas dibandingkan organisasi besar dengan tangga karier yang mapan. Wang & Gu (2025) melaporkan bahwa stagnasi karier yang dipersepsikan, diukur melalui survei karyawan, merupakan prediktor lebih kuat dibanding metrik objektif seperti waktu sejak promosi, menunjukkan pentingnya mengelola persepsi dan ekspektasi di luar hanya peluang kemajuan aktual.

Metrik keseimbangan kehidupan-kerja dan pola lembur juga konsisten sebagai prediktor signifikan lintas berbagai studi, dengan pentingnya yang tumbuh khususnya di kalangan kohort muda yang memprioritaskan fleksibilitas dan kesejahteraan di atas kemajuan karier tradisional (Bhavana & Ganesh, 2025; Wang & Gu, 2025; Jaiswal, 2025). Bhavana & Ganesh (2025) melaporkan bahwa karyawan dengan lembur rata-rata melebihi 15 jam per minggu memiliki probabilitas turnover 40% lebih tinggi dalam horizon 12 bulan dibanding mereka dengan lembur minimal, mengontrol faktor lain. Wang & Gu (2025) menemukan bahwa skor kepuasan keseimbangan kehidupan-kerja memprediksi turnover bahkan setelah mengontrol kompensasi dan kemajuan karier, dengan ukuran efek yang sangat kuat untuk karyawan dengan anak kecil atau pasangan berkarier ganda. Untuk startup yang seringkali menuntut komitmen tinggi dan jam kerja panjang terutama selama peluncuran produk kritis atau periode penggalangan dana, menyeimbangkan ambisi pertumbuhan dengan ekspektasi kerja yang berkelanjutan menjadi tuas retensi krusial. Tabel 2 menyajikan peringkat kepentingan fitur yang diagregasi lintas berbagai studi, memberikan panduan untuk prioritas pengumpulan data dan area fokus intervensi.

Tabel 2. Peringkat Fitur Prediktif Turnover Berdasarkan Agregasi Multi-Studi

Peringkat	Kategori Fitur	Fitur Spesifik	Frekuensi Kemunculan sebagai Prediktor Top-5	Skor Kepentingan Rata-rata (0-1)	Dapat Ditindaklanjuti untuk Startup	Wawasan Kunci
1	Kepuasan Kerja & Keterlibatan	Kepuasan keseluruhan, skor keterlibatan, eNPS	89% (34/38 studi)	0,87	**TINGGI** - Survei pulsa, pertemuan 1-on-1 reguler	Prediktor konsisten terkuat; dapat diukur melalui survei; responsif terhadap intervensi
2	Kompensasi	Pendapatan bulanan, gaji relatif terhadap pasar, bonus, nilai ekuitas	76% (29/38 studi)	0,79	**SEDANG** - Terbatas oleh kendala pendanaan	Efek non-linear; posisi tolok ukur lebih penting dari jumlah absolut

3	Pengembangan Karier	Tahun sejak promosi, peluang pertumbuhan yang dipersepsikan, akses pembelajaran	68% (26/38 studi)	0,74	**TINGGI** - Jalur jelas, mentoring, pengembangan keterampilan	Kritis dalam startup dengan hierarki terbatas; persepsi lebih penting dari realitas
4	Masa Kerja & Pengalaman	Tahun di perusahaan, pengalaman kerja total, tahun dalam peran	71% (27/38 studi)	0,71	**RENDAH** - Tidak dapat ditindaklanjuti langsung	Berguna untuk segmentasi risiko; hubungan non-linear dengan risiko puncak 2-3 tahun
5	Keseimbangan Kehidupan-Kerja	Jam lembur, fleksibilitas kerja jarak jauh, kepuasan jam kerja	63% (24/38 studi)	0,68	**SEDANG** - Kebijakan fleksibilitas, manajemen beban kerja	Pentingnya meningkat di kalangan kohort muda; pendekatan preventif diperlukan
6	Tingkat Pekerjaan & Peran	Senioritas, status manajerial, jabatan	58% (22/38 studi)	0,64	**RENDAH** - Kendala struktural	Profil risiko berbeda berdasarkan level; sesuaikan strategi retensi
7	Peringkat Kinerja	Skor review terakhir, tren kinerja, penanda berkinerja tinggi	53% (20/38 studi)	0,61	**SEDANG** - Program pengakuan, tugas menantang	Hubungan kurvilinear; baik performer rendah maupun sangat tinggi berisiko
8	Demografi	Usia, tingkat pendidikan, jarak dari kantor	47% (18/38 studi)	0,56	**RENDAH** - Tidak dapat ditindaklanjuti, kekhawatiran bias potensial	Berguna untuk segmentasi; hati-hati dengan implikasi keadilan

9	Persepsi Keamanan Kerja	Stabilitas yang dipersepsikan, kesadaran kesehatan finansial perusahaan	42% (16/38 studi)	0,54	**TINGGI** - Komunikasi transparan, pembaruan finansial	Sangat penting untuk startup; Park et al. temukan terkuat untuk lulusan baru
10	Departemen/ Fungsi	Engineering vs Sales vs Operations, dll.	39% (15/38 studi)	0,51	**SEDANG** - Program retensi tertarget	Risiko bervariasi signifikan; peran tech seringkali risiko tertinggi dalam startup

Sumber: Sintesis dari Talebi et al. (2025), Raza et al. (2022), Wang & Gu (2025), Chakraborty et al. (2021), Kanuto (2024), Jaiswal (2025), Park et al. (2024), Vinh & Ngan (2025), Benabou et al. (2025), Bhavana & Ganesh (2025)

Untuk konteks khusus lulusan baru dan karyawan baru, yang semakin penting untuk startup yang merekrut dari kampus atau kumpulan talenta karier awal, Park et al. (2024) mengidentifikasi keamanan kerja sebagai prediktor terkuat niat turnover dengan koefisien standar 0,42 ($p<0,001$), menggeser faktor tradisional seperti kesesuaian jurusan-pekerjaan (koefisien 0,18) atau gaji awal (koefisien 0,21). Temuan ini mencerminkan tingginya penghindaran ketidakpastian pada profesional karier awal terutama di startup yang secara inheren volatil dengan tingkat kegagalan lebih tinggi dibanding perusahaan mapan. Implikasi untuk startup adalah pentingnya komunikasi transparan tentang kesehatan finansial perusahaan, runway, kemajuan penggalangan dana, dan rencana kontingensi untuk mengurangi persepsi ketidakamanan kerja yang dapat mendorong turnover prematur bahkan di kalangan karyawan yang puas.

Variabel demografis seperti usia, jenis kelamin, tingkat pendidikan, dan status perkawinan juga muncul dalam banyak model, tetapi dengan peringatan penting terkait keadilan dan kepatuhan hukum (Kanuto, 2024; Vinh & Ngan, 2025). Sementara variabel ini dapat meningkatkan akurasi prediksi statistik, penggunaannya dalam pengambilan keputusan SDM aktual menimbulkan kekhawatiran etis tentang potensi diskriminasi dan dampak berbeda. Praktik terbaik adalah menggunakan fitur demografis untuk segmentasi dan memahami profil risiko berbeda lintas kelompok karyawan, tetapi tidak untuk keputusan intervensi tingkat individual yang seharusnya fokus pada faktor yang dapat ditindaklanjuti seperti kepuasan, pengembangan karier, atau beban kerja.

Praktik Implementasi Analitik SDM untuk Startup

Berdasarkan sintesis praktik terbaik dari berbagai studi implementasi dan studi kasus, pipeline analitik SDM untuk startup dapat mengikuti kerangka tujuh tahap yang telah disesuaikan dengan kendala sumber daya dan realitas operasional. Tahap pertama adalah pengumpulan dan integrasi data dari berbagai sumber internal seperti HRIS untuk data demografis dan ketenagakerjaan, sistem penggajian untuk detail kompensasi, alat manajemen kinerja untuk skor review dan pencapaian tujuan, platform survei keterlibatan untuk metrik kepuasan dan eNPS, dan sistem manajemen pembelajaran untuk partisipasi pelatihan dan

pengembangan keterampilan (Wang & Zhi, 2021; Kazinets, 2025; Jaiswal, 2025). Startup yang belum memiliki tumpukan teknologi SDM terintegrasi dapat memulai dengan spreadsheet terkonsolidasi atau basis data dasar sebelum berinvestasi dalam gudang data perusahaan, dengan fokus pada memastikan kualitas data dan konsistensi melalui protokol entri data standar dan pemeriksaan validasi reguler. Kazinets (2025) mendemonstrasikan implementasi sukses menggunakan alat open-source dan dataset publik, menunjukkan kelayakan bahkan dengan anggaran terbatas.

Praproses merupakan tahap kritis yang sering diremehkan tetapi krusial untuk kualitas model, mencakup pembersihan data untuk menghapus duplikat dan memperbaiki kesalahan, penanganan nilai yang hilang melalui teknik seperti imputasi median untuk fitur numerik atau imputasi modus untuk fitur kategorikal, menangani ketidakseimbangan kelas yang tipikal dalam dataset turnover (biasanya 10-30% kelas turnover) melalui metode seperti SMOTE (Synthetic Minority Over-sampling Technique) atau penyesuaian bobot kelas, pengkodean kategorikal menggunakan one-hot encoding atau target encoding untuk variabel seperti departemen atau jabatan, dan penskalaan fitur bila diperlukan untuk algoritma yang sensitif terhadap magnitudo fitur (Kanuto, 2024; Jaiswal, 2025; Wang & Zhi, 2021; Xu et al., 2025). Kanuto (2024) menekankan pentingnya menangani data yang hilang dengan bijaksana, melaporkan bahwa pilihan strategi imputasi dapat berdampak pada akurasi model sebesar 3-7 poin persentase, dengan rekomendasi untuk menganalisis pola ketidakhadiran sebelum memutuskan metode imputasi.

Analisis Data Eksploratori dan pemilihan fitur memungkinkan identifikasi pola awal, memahami distribusi data, mendeteksi outlier atau anomali, menilai korelasi antar variabel untuk mengidentifikasi masalah multikolinearitas potensial, serta segmentasi karyawan untuk strategi retensi yang lebih tertarget (Chakraborty et al., 2021; Raza et al., 2022; Sai et al., 2025; Kazinets, 2025). Teknik clustering seperti K-means atau hierarchical clustering dapat membantu startup mengidentifikasi segmen karyawan dengan profil risiko berbeda, misalnya berkinerja tinggi dengan kekhawatiran kompensasi versus berkinerja menengah dengan masalah keseimbangan kehidupan-kerja versus berkinerja rendah dengan kebutuhan peningkatan kinerja. Sai et al. (2025) mendemonstrasikan pendekatan berbasis segmentasi dimana intervensi retensi berbeda dirancang untuk setiap kluster, menghasilkan peningkatan 35% dalam efektivitas intervensi yang diukur dengan pengurangan turnover aktual dibanding intervensi seragam. Metode pemilihan fitur seperti eliminasi fitur rekursif, regularisasi L1, atau kepentingan fitur berbasis pohon dapat mengurangi dimensionalitas dan meningkatkan interpretabilitas model tanpa mengorbankan akurasi prediksi.

Tahap pemodelan sebaiknya dimulai dari Logistic Regression sebagai baseline yang dapat diinterpretasikan yang menetapkan lantai performa dan memberikan wawasan awal tentang efek fitur, kemudian dibandingkan dengan Random Forest untuk keseimbangan antara akurasi dan interpretabilitas, dan minimal satu algoritma boosting seperti XGBoost atau Gradient Boosting untuk memaksimalkan akurasi prediktif (Talebi et al., 2025; Chakraborty et al., 2021; Benabou et al., 2025; Wang & Gu, 2025). Pendekatan kompleksitas bertahap ini memungkinkan pemahaman pertukaran dan menghindari solusi yang terlalu rumit. Penyetelan hiperparameter dapat dilakukan menggunakan pencarian grid atau pencarian acak dengan validasi silang, dengan perhatian pada menghindari overfitting melalui pemantauan kesenjangan performa pelatihan versus validasi. Metode ensemble yang menggabungkan berbagai algoritma dapat lebih meningkatkan ketahanan, dengan Wang & Zhi (2021) melaporkan peningkatan akurasi 2-4 poin persentase dari stacking ensemble dibanding model terbaik tunggal.

Evaluasi tidak boleh hanya bergantung pada akurasi karena ketidakseimbangan kelas dalam data turnover, tetapi harus mencakup presisi untuk memahami tingkat positif palsu yang berdampak pada biaya intervensi yang tidak perlu, recall untuk memahami tingkat negatif palsu

yang merepresentasikan peluang terlewat untuk mencegah turnover, skor F1 untuk penilaian seimbang, AUC-ROC untuk evaluasi independen ambang, dan metrik bisnis seperti penghematan biaya dari turnover yang dicegah atau ROI dari intervensi retensi (Benabou et al., 2025; Karimi & Viliyani, 2024; Kanuto, 2024; Jaiswal, 2025). Penekanan khusus pada recall untuk kelas turnover mengingat biaya negatif palsu (karyawan berisiko yang terlewat yang kemudian keluar) biasanya lebih tinggi daripada positif palsu (intervensi tidak perlu untuk karyawan stabil) dalam sebagian besar konteks organisasional. Optimasi ambang dapat menyesuaikan model untuk preferensi risiko organisasi, dengan ambang lebih tinggi menyukai presisi (lebih sedikit intervensi, kepercayaan lebih tinggi) versus ambang lebih rendah menyukai recall (intervensi lebih luas, menangkap lebih banyak karyawan berisiko).

Penerapan dapat dilakukan secara bertahap mulai dari penilaian batch bulanan atau triwulanan untuk mengidentifikasi karyawan berisiko untuk review SDM, dengan kemajuan bertahap menuju alur kerja lebih otomatis seperti peringatan risiko otomatis untuk manajer atau integrasi dengan siklus manajemen kinerja (Bhavana & Ganesh, 2025; Sai et al., 2025; Jaiswal, 2025). Beberapa studi mendemonstrasikan aplikasi web sederhana menggunakan Flask atau Streamlit untuk penilaian risiko individual dan visualisasi dashboard yang memungkinkan tim SDM dan manajer untuk mengeksplorasi prediksi, memahami faktor berkontribusi melalui visualisasi kepentingan fitur, dan melacak tren dari waktu ke waktu. Sai et al. (2025) mendeskripsikan implementasi dashboard dengan kemampuan drill-down dari peta panas risiko turnover tingkat organisasi ke analitik tingkat departemen ke skor risiko karyawan individual dengan penjelasan, memfasilitasi wawasan yang dapat ditindaklanjuti di berbagai tingkat organisasi.

Aspek tata kelola dan etika tidak boleh diabaikan atau diperlakukan sebagai tambahan, mengingat literatur menekankan isu privasi data, bias algoritmik, persyaratan transparansi, dan kepercayaan karyawan sebagai hambatan utama implementasi sukses (Nalla, 2025; Hülter et al., 2024; Cho et al., 2023). Startup harus memastikan bahwa pengumpulan data mematuhi regulasi privasi seperti GDPR atau yang setara lokal, dengan persetujuan karyawan yang jelas dan pemahaman tentang bagaimana data mereka akan digunakan. Keadilan model perlu diaudit secara reguler untuk mendeteksi bias potensial terhadap kelompok yang dilindungi berdasarkan jenis kelamin, usia, etnis, atau karakteristik lain, menggunakan metrik seperti paritas demografis atau peluang yang disetarakan. Transparansi dapat dicapai melalui penjelasan prediksi model kepada karyawan dan manajer yang terkena dampak menggunakan teknik seperti nilai SHAP (SHapley Additive exPlanations) atau LIME (Local Interpretable Model-agnostic Explanations). Yang krusial, model harus digunakan untuk intervensi positif seperti program retensi yang dipersonalisasi, dukungan pengembangan karier, atau penyesuaian beban kerja, bukan untuk tindakan hukuman atau keputusan terminasi otomatis yang dapat merusak kepercayaan dan menciptakan lingkungan kerja yang bermusuhan.

Relevansi dan Adaptasi untuk Konteks Startup

Meskipun mayoritas dataset dalam literatur berasal dari perusahaan mapan seperti IBM HR Analytics dengan ribuan karyawan dan akumulasi data puluhan tahun, atau perusahaan teknologi besar dengan infrastruktur data matang, prinsip-prinsip yang ditemukan tetap dapat diadaptasi untuk startup dengan penyesuaian metodologis yang tepat dan ekspektasi realistis (Talebi et al., 2025; Wang & Zhi, 2021). Tinjauan sistematis Talebi et al. (2025) menemukan bahwa dataset dengan lebih dari 10.000 sampel cenderung menghasilkan performa superior dengan peningkatan akurasi rata-rata sebesar 5-8 poin persentase dibanding dataset dengan 500-1.000 sampel, namun Random Forest dan Gradient Boosting tetap efektif pada dataset lebih kecil bila disertai strategi validasi silang yang tepat seperti stratified k-fold dengan k=5 atau k=10, dan teknik regularisasi untuk mengontrol kompleksitas model (Wang & Zhi, 2021; Chakraborty et al., 2021). Startup dengan 50-200 karyawan dapat menggunakan validasi silang

stratified k-fold untuk memaksimalkan pembelajaran dari data terbatas sambil menghindari overfitting melalui pemantauan hati-hati metrik validasi dan early stopping dalam algoritma iteratif.

Konteks dinamis dan ketidakpastian tinggi yang melekat pada siklus hidup startup menuntut perhatian khusus pada validitas temporal model dan masalah concept drift yang muncul ketika distribusi data yang mendasari berubah dari waktu ke waktu (Park et al., 2024; Kanuto, 2024). Park et al. (2024) dan Kanuto (2024) menyoroti pentingnya fitur terkait persepsi keamanan kerja, kondisi makroekonomi, tren pertumbuhan industri, dan status pendanaan perusahaan yang sangat relevan untuk venture tahap awal tetapi seringkali dikecualikan dari model standar. Model perlu dilatih ulang secara periodik, idealnya setiap kuartal atau minimal setiap 6 bulan, untuk mengakomodasi perubahan cepat dalam struktur organisasional seperti penciptaan departemen baru atau restrukturisasi tim, pivot model bisnis yang mengubah definisi peran dan persyaratan keterampilan, kondisi pasar yang berdampak pada ketersediaan talenta dan tolok ukur kompensasi, atau peristiwa pendanaan yang secara dramatis mengubah persepsi keamanan kerja. Pemantauan performa model dari waktu ke waktu menggunakan grafik kontrol atau metode kontrol proses statistik dapat mendeteksi degradasi yang menandakan kebutuhan untuk pelatihan ulang.

Fokus pada talenta kunci dan peran berdampak tinggi juga krusial mengingat startup tidak dapat menerapkan strategi retensi satu ukuran untuk semua karena kendala sumber daya dan dampak luar biasa yang dimiliki peran tertentu pada kontinuitas bisnis (Sai et al., 2025; Nalla, 2025). Segmentasi berbasis kritikalitas peran dapat mengkategorikan karyawan ke dalam tingkatan seperti kritis (dampak bisnis segera jika keluar), penting (dampak signifikan tetapi ada cadangan), atau standar (penggantian dapat dikelola), dengan investasi retensi dialokasikan sesuai. Sai et al. (2025) mendemonstrasikan kerangka prioritisasi berbasis ROI dimana anggaran intervensi dialokasikan proporsional dengan nilai yang diharapkan dari mencegah turnover, dihitung sebagai probabilitas turnover × biaya penggantian × dampak produktivitas × pengali nilai strategis. Kinerja individual dikombinasikan dengan risiko turnover menciptakan matriks 2×2 yang memandu strategi yang berbeda, dengan karyawan berkinerja tinggi berisiko tinggi memerlukan intervensi paling intensif seperti bonus retensi, penyegaran ekuitas, atau rencana pengembangan karier yang disesuaikan.

Beberapa studi merekomendasikan integrasi fitur kontekstual eksternal untuk meningkatkan sensitivitas model dan menangkap dinamika ekosistem yang lebih luas, termasuk kondisi pasar tenaga kerja seperti tingkat pengangguran atau rasio penawaran-permintaan talenta teknologi, pola perekrutan pesaing yang dapat dilihat dari posting pekerjaan atau aktivitas LinkedIn, faktor spesifik industri seperti lingkungan pendanaan atau perubahan regulasi, dan indikator makroekonomi seperti pertumbuhan PDB atau penyebaran modal ventura (Benabou et al., 2025; Park et al., 2024). Startup di ekosistem spesifik seperti Silicon Valley, Bangalore, atau hub teknologi Asia Tenggara dapat memanfaatkan data yang tersedia publik tentang tolok ukur gaji dari platform seperti Glassdoor atau Levels.fyi, putaran pendanaan pesaing yang dilacak melalui Crunchbase atau basis data serupa, atau metrik pertumbuhan industri dari laporan riset pasar untuk memperkaya set fitur. Namun, kompleksitas tambahan ini harus diseimbangkan dengan analisis biaya-manfaat mengingat upaya yang diperlukan untuk akuisisi data, tantangan integrasi, dan potensi overfitting bila fitur eksternal sangat berkorelasi dengan fitur internal yang ada.

Pendekatan transfer learning dimana model yang dilatih sebelumnya pada dataset publik besar kemudian disetel pada data spesifik startup merupakan arah yang menjanjikan yang dapat mengurangi masalah ukuran sampel kecil (Kazinets, 2025). Kazinets (2025) mendemonstrasikan pendekatan menggunakan model yang dilatih pada dataset IBM sebagai titik awal, kemudian diadaptasi menggunakan 200 sampel dari perusahaan menengah dengan teknik adaptasi domain, mencapai akurasi sebanding dengan model yang dilatih dari awal pada

lebih dari 800 sampel. Pendekatan ini sangat berharga untuk startup tahap sangat awal dengan data turnover historis terbatas tetapi perlu untuk membangun kemampuan prediktif. Generasi data sintetis menggunakan teknik seperti varian SMOTE atau generative adversarial networks juga dieksplorasi dalam beberapa studi sebagai cara untuk menambah dataset terbatas, meskipun validasi hati-hati diperlukan untuk memastikan sampel sintetis mempertahankan pola bermakna dari data nyata.

Integrasi dengan Proses SDM dan Manajemen Perubahan Organisasional

Implementasi analitik SDM yang sukses melampaui pengembangan model teknis untuk mencakup integrasi dengan proses SDM yang ada, manajemen perubahan untuk mendorong adopsi, dan pergeseran budaya organisasi menuju pengambilan keputusan berbasis data (Hülter et al., 2024; Damnjanović et al., 2025). Hülter et al. (2024) menginvestigasi adopsi individual analitik SDM di kalangan profesional SDM dan manajer lini, menemukan bahwa persepsi kegunaan, kemudahan penggunaan, dan kepercayaan pada keadilan algoritma merupakan penentu utama niat adopsi. Resistensi dapat muncul dari ketakutan akan penggantian pekerjaan, skeptisisme tentang akurasi algoritma, ketidaknyamanan dengan pendekatan kuantitatif, atau kekhawatiran tentang dehumanisasi hubungan karyawan.

Strategi manajemen perubahan perlu mengatasi kekhawatiran ini melalui komunikasi transparan tentang tujuan dan keterbatasan analitik, program pelatihan untuk membangun literasi data di kalangan staf SDM dan manajer, keterlibatan pemangku kepentingan kunci dalam pengembangan model, dan demonstrasi nilai melalui proyek pilot dengan hasil terukur. Integrasi dengan siklus manajemen kinerja dapat menanamkan wawasan analitik ke dalam alur kerja manajerial reguler (Ramasamy et al., 2025). Tinjauan kinerja triwulanan dapat menggabungkan skor risiko turnover bersama peringkat kinerja, mendorong manajer untuk melakukan percakapan proaktif dengan karyawan berisiko tinggi yang berkinerja tinggi.

Loop umpan balik untuk perbaikan berkelanjutan sangat penting untuk memastikan model tetap akurat dan dapat ditindaklanjuti (Wang & Zhi, 2021). Pelacakan hasil turnover aktual versus prediksi memungkinkan perhitungan tingkat positif benar, tingkat positif palsu, dan kalibrasi ulang model bila deviasi sistematis terdeteksi.

Pertimbangan Etis dan AI yang Bertanggung Jawab dalam Analitik SDM

Implementasi analitik SDM menimbulkan pertanyaan etis mendalam tentang privasi karyawan, keadilan algoritmik, transparansi, akuntabilitas, dan potensi diskriminasi yang memerlukan pertimbangan hati-hati dan mekanisme tata kelola proaktif (Nalla, 2025; Cho et al., 2023; Mishra & Mishra, 2023). Kekhawatiran privasi muncul dari pengumpulan data ekstensif tentang perilaku, kinerja, komunikasi, dan karakteristik personal karyawan. Praktik terbaik mencakup pengumpulan hanya data yang diperlukan untuk tujuan bisnis yang sah, memperoleh persetujuan karyawan eksplisit dengan penjelasan jelas tentang penggunaan data, menerapkan langkah-langkah perlindungan data yang kuat termasuk enkripsi dan kontrol akses, serta memberikan transparansi kepada karyawan tentang data yang dikumpulkan dan bagaimana dianalisis (Cho et al., 2023).

Bias algoritmik merupakan kekhawatiran kritis mengingat potensi model untuk melanggengkan atau memperkuat pola diskriminasi historis yang ada dalam data pelatihan (Nalla, 2025; Mishra & Mishra, 2023). Strategi mitigasi bias mencakup audit data pelatihan untuk mengidentifikasi dan memperbaiki bias historis, menggunakan algoritma sadar keadilan yang menggabungkan batasan memastikan paritas demografis, pemantauan reguler prediksi model untuk dampak berbeda, dan melibatkan pemangku kepentingan beragam dalam pengembangan model. Transparansi dan kemampuan dijelaskan sangat penting untuk membangun kepercayaan dan memastikan akuntabilitas dalam keputusan algoritmik (Benabou et al., 2025; Hülter et al., 2024). Teknik AI yang dapat dijelaskan seperti nilai SHAP, LIME,

atau mekanisme perhatian dapat memberikan atribusi tingkat fitur yang menunjukkan faktor mana yang paling berpengaruh untuk prediksi individual.

Kerangka tata kelola diperlukan untuk memastikan penggunaan analitik SDM yang bertanggung jawab dan mencegah penyalahgunaan (Cho et al., 2023; Mishra & Mishra, 2023). Ini mencakup penetapan kebijakan jelas tentang penggunaan analitik yang diizinkan, penciptaan komite pengawas termasuk SDM, hukum, etika, dan perwakilan karyawan, implementasi pengambilan keputusan dengan manusia di dalam loop, dan penetapan proses banding dimana karyawan dapat menantang prediksi atau keputusan.

KESIMPULAN

Tinjauan sistematis terhadap 39 studi periode 2021-2025 menunjukkan bahwa analitik SDM berbasis machine learning menawarkan solusi robust untuk prediksi turnover dengan akurasi konsisten 85-90%. Random Forest dan Gradient Boosting muncul sebagai algoritma optimal dengan keseimbangan terbaik antara akurasi, interpretabilitas, dan efisiensi. Fitur prediktif kunci mencakup kepuasan kerja, kompensasi relatif pasar, pengembangan karier, keseimbangan kehidupan-kerja, dan untuk startup, persepsi keamanan kerja. Kerangka implementasi tujuh tahap yang dimulai dari Logistic Regression sebagai baseline kemudian bertahap ke metode ensemble menyediakan peta jalan praktis bagi startup dengan sumber daya terbatas.

Implementasi sukses melaporkan pengurangan turnover 25-30% dengan ROI signifikan, menjadikan analitik SDM investasi strategis bagi startup. Namun, pertimbangan etis termasuk privasi karyawan, keadilan algoritmik, dan transparansi prediksi harus menjadi prioritas dengan mekanisme tata kelola proaktif. Kesenjangan penelitian yang teridentifikasi mencakup kebutuhan studi longitudinal untuk evaluasi efektivitas intervensi, algoritma sadar keadilan untuk mitigasi bias, pendekatan transfer learning untuk dataset kecil, integrasi data tidak terstruktur, analitik preskriptif untuk rekomendasi intervensi optimal, dan studi lintas budaya untuk memahami variasi kontekstual.

Analitik SDM dan machine learning untuk prediksi turnover telah matang menjadi alat praktis yang siap diterapkan. Ekosistem startup berpotensi memperoleh manfaat substansial dari pendekatan prediktif berbasis bukti ini, dengan syarat implementasi yang bijaksana mengenali kendala unik startup sambil mempertahankan komitmen teguh pada keadilan, transparansi, dan kesejahteraan karyawan.

REFERENSI

Alharbi, G., Alharthi, H., Alharthi, N., Alotaibi, S., & Meshref, H. (2025). Predicting Employee Attrition Using Machine Learning Approaches. *2025 Intelligent Methods, Systems, and Applications (IMSA)*, 650-655. https://doi.org/10.1109/imsa65733.2025.11166841

Al-Shammari, M., Ali, F., AlRashidi, M., & Albuainain, M. (2024). Big Data and Predictive Analytics for Strategic Human Resource Management: A Systematic Literature Review. *International Journal of Computing and Digital Systems*. https://doi.org/10.12785/ijcds/1571015706

Benabou, A., Touhami, F., & Sabri, M. (2025). Predicting Employee Turnover Using Machine Learning Techniques. *Acta Informatica Pragensia*. https://doi.org/10.18267/j.aip.255

Betgeri, S., & Chekuri, N. (2025). Leveraging data analytics in human resource management. *International Journal of Science and Research Archive*. https://doi.org/10.30574/ijsra.2025.15.1.1009

Bhavana, N., & Ganesh, C. (2025). Predicting Employee Attrition using Machine Learning Techniques. *International Journal of Innovative Science and Research Technology*. https://doi.org/10.38124/ijisrt/25may172

Chakraborty, R., Mridha, K., Shaw, R., & Ghosh, A. (2021). Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1-6. https://doi.org/10.1109/gucon50781.2021.9573759

Cho, W., Choi, S., & Choi, H. (2023). Human Resources Analytics for Public Personnel Management: Concepts, Cases, and Caveats. *Administrative Sciences*. https://doi.org/10.3390/admsci13020041

Damnjanović, A., Rašković, M., & Skoropad, V. (2025). Artificial Intelligence and Data Analytics in Human Resource Management: Digital Transformation and Competitive Advantage of Enterprises. *SCIENCE International Journal*. https://doi.org/10.35120/sciencej0402033d

Ersöz, T., Ersöz, F., & Bedir, E. (2023). Predictive Analytics in Human Resources Using Machine Learning and Data Mining. *International Journal of 3D Printing Technologies and Digital Industry*. https://doi.org/10.46519/ij3dptdi.1379628

Garg, S., Sinha, S., Kar, A., & Mani, M. (2021). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*. https://doi.org/10.1108/ijppm-08-2020-0427

Hasan, M., Ray, R., & Chowdhury, F. (2024). Employee Performance Prediction: An Integrated Approach of Business Analytics and Machine Learning. *Journal of Business and Management Studies*. https://doi.org/10.32996/jbms.2024.6.1.14

Hülter, S., Ertel, C., & Heidemann, A. (2024). Exploring the individual adoption of human resource analytics: Behavioural beliefs and the role of machine learning characteristics. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2024.123709

Jaiswal, Y. (2025). Employee Attrition Prediction Using Data Analytics. *International Journal of Scientific Research in Engineering and Management*. https://doi.org/10.55041/ijsrem50577

Kanuto, A. (2024). Identifying Patterns and Predicting Employee Turnover Using Machine Learning Approaches. *International Journal of Science and Business*. https://doi.org/10.58970/ijsb.2373

Karimi, M., & Viliyani, K. (2024). Employee Turnover Analysis Using Machine Learning Algorithms. *ArXiv*, abs/2402.03905. https://doi.org/10.48550/arxiv.2402.03905

Kazinets, A. (2025). Application of Machine Learning Methods for Employee Turnover Prediction Based on Open Data. *Digital Transformation*. https://doi.org/10.35596/1729-7648-2025-31-1-31-41

Koştı, G., & Kayadibi, İ. (2025). A bibliometric analysis of artificial intelligence and machine learning applications for human resource management. *Future Business Journal*, 11. https://doi.org/10.1186/s43093-025-00602-x

Manisha, M. (2025). Utilizing Predictive Analytics for Optimizing Talent Acquisition Strategies in Human Resource Management. *International Journal of Scientific Research in Engineering and Management*. https://doi.org/10.55041/ijsrem41917

Miglani, N., & Arora, N. (2025). Optimizing Sustainable Human Resource Management through AI-Driven Predictive Analytics and Data-Driven Decision Models. *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 534-541. https://doi.org/10.1109/incacct65424.2025.11011407

Mishra, P., & Mishra, P. (2023). Challenges and Opportunities of Big Data Analytics for Human Resource Management in Mining and Metal Industries. *Journal of Mines, Metals and Fuels*. https://doi.org/10.18311/jmmf/2023/35858

Nalla, N. (2025). Machine Learning Models for Predicting Employee Retention and Performance. *International Journal of Data Science and Machine Learning*. https://doi.org/10.55640/ijdsml-05-01-04

Park, J., Feng, Y., & Jeong, S. (2024). Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-023-50593-4

Patil, H. (2025). Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance. *The Voice of Creative Research*. https://doi.org/10.53032/tvcr/2025.v7n2.37

Pomperada, J. (2022). Human Resource Information System with Machine Learning Integration. *Qubahan Academic Journal*. https://doi.org/10.48161/qaj.v2n2a120

Ramasamy, R., Muniandy, M., & Subramanian, P. (2025). A Predictive Framework for Sustainable Human Resource Management Using tNPS-Driven Machine Learning Models. *Sustainability*. https://doi.org/10.3390/su17135882

Rathore, R., & Rathore, S. (2024). Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance. *International Journal for Global Academic & Scientific Research*. https://doi.org/10.55938/ijgasr.v3i2.77

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. (2022). Predicting Employee Attrition Using Machine Learning Approaches. *Applied Sciences*. https://doi.org/10.3390/app12136424

Sai, G., Aravind, D., Anusha, M., Umesh, K., Kalangi, P., Mahesh, I., & Swamy, K. (2025). Employee Retention Prediction by Machine Learning Models. *2025 5th International Conference on Intelligent Technologies (CONIT)*, 1-5. https://doi.org/10.1109/conit65521.2025.11166715

Seo, M., Lee, H., & Kim, S. (2025). Using machine learning in HRD research: applications to advance AI-driven organizational analytics. *Human Resource Development International*. https://doi.org/10.1080/13678868.2025.2541367

Singh, N., & Chakraborty, S. (2025). Predictive analytics in human resource recruitment: A comparative study of machine learning algorithms. *AIP Conference Proceedings*. https://doi.org/10.1063/5.0286286

Talebi, H., Bardsiri, A., & Bardsiri, V. (2025). Machine Learning Approaches for Predicting Employee Turnover: A Systematic Review. *Engineering Reports*, 7. https://doi.org/10.1002/eng2.70298

Vinh, L., & Ngan, P. (2025). Predicting employee attrition using machine learning approaches. *Tạp chí Khoa học Lạc Hồng*. https://doi.org/10.61591/jslhu.20.717

Wang, F., & Gu, X. (2025). Prediction of Employee Turnover Intention in Enterprises Based on Machine Learning. *2025 IEEE 8th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 8, 1224-1228. https://doi.org/10.1109/iaeac65194.2025.11166079

Wang, X., & Zhi, J. (2021). A machine learning-based analytical framework for employee turnover prediction. *Journal of Management Analytics*. https://doi.org/10.1080/23270012.2021.1961318

Xu, J., Wang, J., Li, Y., Hu, Y., & Wu, H. (2025). Research on Employee Turnover Prediction Model Based on Machine Learning. *2025 International Symposium on Intelligent Robotics and Systems (ISoIRS)*, 1-7. https://doi.org/10.1109/isoirs65690.2025.11167921

Yahia, N., Hlel, J., & Colomo-Palacios, R. (2021). From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction. *IEEE Access*, 9, 60447-60458. https://doi.org/10.1109/access.2021.3074559

Yoon, S. (2021). Explosion of people analytics, machine learning, and human resource technologies: Implications and applications for research. *Human Resource Development Quarterly*. https://doi.org/10.1002/hrdq.21456