ANALISIS PREDIKTIF EMPLOYEE ATTRITION MENGGUNAKAN XGBOOST, BAYESIAN TUNING, DAN SHAP

Nikko Riestian Putra Wardoyo[1], Anisah Nabilah[2], Siti Dinarti[3], Tito Dwi Saputro[4]

[1, 2, 4,] Program Studi Informatika, Universitas PGRI Jombang, [3,] Program Studi Pendidikan Matematika, Universitas PGRI Jombang

Jln. Pattimura III No.20, Sengon, Kec. Jombang, Kabupaten Jombang, Jawa Timur 61411
[1] ncool750@gmail.com, [2] anisah11nabilah@gmail.com, [3] dinarti.matem@gmail.com, [4] titosaputro59@gmail.com

Abstract

Employee turnover (attrition) is a strategic challenge that impacts a company's operational inefficiency. This study aims to develop a predictive classification model for early and transparent attrition detection using the Extreme Gradient Boosting (XGBoost) algorithm. The IBM HR Analytics dataset suffers from extreme class imbalance, which is effectively handled through the Synthetic Minority Over-sampling Technique (SMOTE). To avoid inefficiency in parameter search, optimization is implemented using Bayesian Tuning via the Optuna framework. Computational evaluation shows the optimized model achieves a Recall of 34% and an F1-Score of 40%. Although these metrics represent the empirical limitations of tabular data in mapping the complexity of human sentiments, the figures are operationally considered proportional for reducing the risk of False Negatives in minority class detection without drastically compromising precision. In addition to metric performance, this study solves the black-box problem in decision tree algorithms by integrating Explainable AI based on SHapley Additive exPlanations (SHAP). The extraction of SHAP values reveals that the absence of stock option allocation (StockOptionLevel), placement outside the research department, low job satisfaction, and high overtime frequency are the four main anomalies triggering resignation. The results of this study provide not only a predictive model but also a data-driven diagnostic system that can be directly interpreted by HR management to formulate precise human resource retention strategies.

Keywords : *employee attrition, XGBoost, SMOTE, bayesian tuning, SHAP*

Abstrak

Perputaran karyawan (employee attrition) merupakan tantangan strategis yang berdampak pada inefisiensi operasional perusahaan. Penelitian ini bertujuan mengembangkan model klasifikasi prediktif untuk mendeteksi attrition secara dini dan transparan menggunakan algoritma Extreme Gradient Boosting (XGBoost). Dataset IBM HR Analytics berisiko ketidakseimbangan kelas ekstrem, yang ditangani secara efektif melalui teknik Synthetic Minority Over-sampling Technique (SMOTE). Untuk menghindari inefisiensi pencarian parameter, optimasi diimplementasikan menggunakan Bayesian Tuning melalui kerangka kerja Optuna. Evaluasi komputasi menunjukkan model teroptimasi mencapai nilai Recall 34% dan F1-Score 40%. Meskipun capaian metrik ini merepresentasikan batasan empiris dari data tabular dalam memetakan kompleksitas sentimen manusia, angka tersebut secara operasional dinilai proporsional untuk mereduksi risiko False Negative pada deteksi kelas minoritas tanpa mengorbankan ketepatan prediksi secara drastis. Selain performa metrik, penelitian ini memecahkan masalah black-box pada algoritma pohon keputusan dengan mengintegrasikan Explainable AI berbasis SHapley Additive exPlanations (SHAP). Ekstraksi nilai SHAP mengungkapkan bahwa ketiadaan alokasi opsi saham (StockOptionLevel), penempatan di luar departemen riset, kepuasan kerja yang rendah, dan tingginya frekuensi lembur merupakan empat anomali utama pemicu pengunduran diri. Hasil penelitian ini tidak sekadar memberikan model prediktif, tetapi juga menghasilkan sistem diagnostik berbasis data yang dapat diinterpretasikan secara langsung oleh manajemen HR untuk merumuskan strategi retensi sumber daya manusia yang presisi.

Kata kunci : *employee attrition, XGBoost, SMOTE, bayesian tuning, SHAP*

1. PENDAHULUAN

Sumber daya manusia (SDM) merupakan aset paling berharga bagi kelangsungan dan daya saing sebuah perusahaan. Saat ini, fenomena perputaran karyawan (employee turnover atau attrition) menjadi tantangan manajerial yang pelik karena berdampak langsung pada penurunan performa organisasional [17]. Biaya operasional yang harus dikeluarkan perusahaan untuk proses rekrutmen dan pelatihan karyawan baru jauh lebih tinggi dibandingkan upaya mempertahankan talenta terbaik yang sudah ada. Oleh karena itu, kemampuan mengidentifikasi dan memprediksi karyawan yang berisiko keluar secara dini menjadi sebuah urgensi.

Seiring dengan masifnya adopsi teknologi Data Science, tahapan Knowledge Discovery in Databases (KDD) banyak dimanfaatkan untuk menggali wawasan dari basis data yang masif [9]. Pendekatan Machine Learning (ML) mulai banyak dieksplorasi sebagai solusi inovatif dalam ranah analitik sumber daya manusia [1], [11]. Berbagai studi literatur membuktikan bahwa metode ensemble learning mampu memprediksi turnover karyawan dengan tingkat presisi yang jauh lebih baik dibandingkan metode statistik tradisional [8]. Secara lebih spesifik, algoritma berbasis Gradient Boosting terbukti menjadi salah satu pendekatan analitik terbaik untuk membedah dan mengklasifikasikan kasus attrition yang memiliki struktur data tabular kompleks [14]. Keandalan XGBoost dalam mengekstraksi dan menyeleksi fitur yang krusial juga telah dibuktikan pada berbagai komputasi prediktif terdahulu [15].

Meskipun model pembelajaran mesin konvensional seperti Support Vector Machine (SVM) yang dioptimasi menggunakan algoritma metaheuristik Particle Swarm Optimization (PSO) telah terbukti sangat andal dalam memecahkan masalah pengenalan pola yang rumit [18], [19], penerapannya secara langsung di ranah analitik manajemen seringkali menemui hambatan praktis. Model-model berakurasi tinggi ini umumnya bersifat black-box (kotak hitam), sehingga penggunaannya sangat disarankan untuk dihindari pada proses pengambilan keputusan berisiko tinggi (high-stakes decisions) karena alur prediksinya tidak transparan [16]. Dalam konteks strategis retensi SDM, pihak manajemen tidak sekadar membutuhkan prediksi "siapa" yang akan keluar, melainkan membutuhkan penjelasan transparan mengenai "mengapa" karyawan tersebut memutuskan untuk keluar agar intervensi yang tepat sasaran dapat dirumuskan.

Untuk memecahkan masalah tersebut, penelitian ini mengusulkan pengembangan model prediktif menggunakan algoritma Extreme Gradient Boosting (XGBoost) [7]. Namun, implementasi pada data SDM umumnya berhadapan dengan masalah ketidakseimbangan kelas (imbalanced data). Untuk itu, teknik prapemrosesan Synthetic Minority Over-sampling Technique (SMOTE) diaplikasikan guna menyeimbangkan distribusi target agar model tidak menghasilkan prediksi yang bias [6]. Selanjutnya, karena performa XGBoost sangat dipengaruhi oleh konfigurasi hyperparameter, pencarian parameter konvensional yang memakan waktu komputasi tinggi digantikan oleh metode probabilistik yang cerdas [5]. Penerapan hyperparameter tuning yang presisi sangat krusial karena terbukti secara empiris mampu meningkatkan akurasi prediksi secara signifikan [12]. Metode Bayesian Optimization [10], [20] melalui kerangka kerja Optuna [2] diimplementasikan untuk mencari nilai parameter global yang paling optimal secara otomatis dan efisien.

Guna mengatasi masalah black-box pada pemodelan, penelitian ini mengintegrasikan pendekatan Explainable Artificial Intelligence (XAI) [4]. Metode SHapley Additive exPlanations (SHAP) digunakan untuk menghitung kontribusi masing-masing fitur terhadap hasil prediksi akhir [13]. Penerapan SHAP pada analitik HR telah terbukti secara efektif mampu memberikan transparansi dengan memetakan faktor-faktor dominan pemicu turnover secara individual maupun global [3].

Berdasarkan latar belakang tersebut, tujuan utama penelitian ini adalah merancang, mengoptimasi, dan mengevaluasi model prediktif employee attrition berbasis XGBoost teroptimasi Bayesian Tuning, serta menginterpretasikan keputusannya menggunakan metode SHAP. Hasil capaian yang ditargetkan dari kegiatan penelitian ini bukan hanya berupa peningkatan performa komputasi algoritma, melainkan terbentuknya rekomendasi operasional berupa strategi retensi SDM berbasis data (data-driven) yang objektif, transparan, dan aplikatif bagi manajemen perusahaan.

Meskipun berbagai algoritma Machine Learning konvensional seperti Support Vector Machine (SVM) atau Random Forest telah terbukti andal dalam memprediksi status karyawan, sebagian besar penelitian terdahulu masih menyisakan kelemahan fundamental (research gap). Pertama, pemodelan sebelumnya kerap mengabaikan fenomena majority class illusion

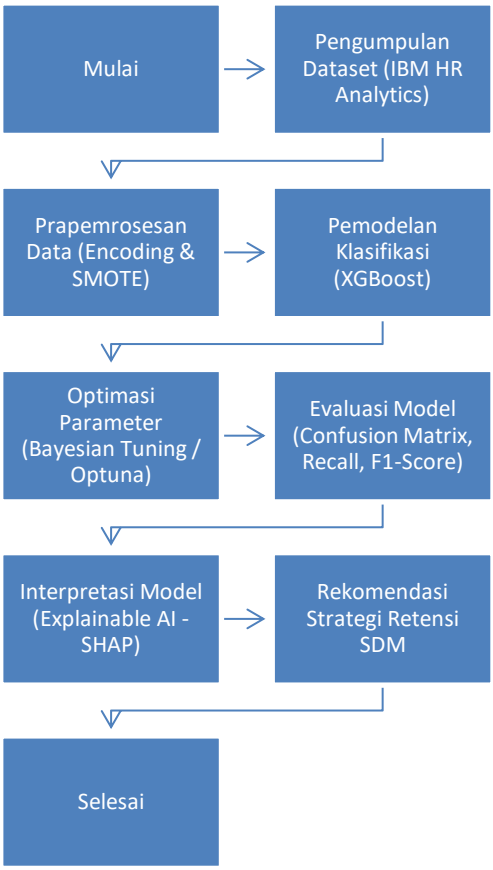
akibat ketidakseimbangan kelas (imbalanced data) yang ekstrem pada basis data SDM. Kedua, model ensemble atau deep learning yang berkinerja tinggi umumnya bersifat black-box, sehingga pihak manajemen hanya mendapatkan prediksi biner tanpa wawasan (insight) kausalitas mengenai akar penyebab pengunduran diri tersebut.

Oleh karena itu, kebaruan (novelty) dari penelitian ini terletak pada pengusulan arsitektur integratif tiga pilar: (1) Synthetic Minority Over-sampling Technique (SMOTE) untuk penanganan data timpang pada level prapemrosesan, (2) optimasi probabilitas cerdas Bayesian Tuning (Optuna) untuk efisiensi komputasi algoritma Extreme Gradient Boosting (XGBoost), dan (3) modul Explainable AI berbasis game theory (SHAP). Integrasi ini secara spesifik dirancang tidak sekadar untuk mengejar metrik akurasi (mengingat tingginya risiko pengambilan keputusan/ high-stakes decisions di ranah SDM), melainkan untuk menghasilkan sistem diagnostik yang transparan, agar manajemen mampu merumuskan strategi retensi talenta secara presisi dan tepat sasaran.

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Bagian ini menguraikan langkah-langkah sistematis yang dilakukan dalam menyelesaikan masalah klasifikasi employee attrition. Penelitian ini diimplementasikan menggunakan kerangka kerja Knowledge Discovery in Databases (KDD) yang berfokus pada ekstraksi pola informasi berharga dari sekumpulan data berskala besar. Skema alur penelitian dirancang ke dalam lima tahapan utama yang berjalan secara sekuensial, mulai dari pengumpulan dataset, prapemrosesan data, pemodelan klasifikasi prediktif, optimasi parameter, hingga interpretasi model menggunakan pendekatan Explainable AI. Rincian kerangka kerja penelitian ini divisualisasikan pada Gambar 1.



Gambar 1. Skema Alur Penelitian.

2.2. Konfigurasi Eksperimen dan Validasi Model

Untuk menjamin reliabilitas dan kemampuan replikasi (reproducibility) dari eksperimen yang dilakukan, dataset dibagi menjadi data latih (training set) dan data uji (testing set) dengan rasio 80:20 menggunakan teknik stratified random sampling. Penggunaan stratifikasi ini sangat krusial untuk memastikan proporsi kelas minoritas (karyawan yang resign) tetap konsisten terwakili baik di dalam fase pelatihan maupun pengujian.

Proses validasi internal selama fase pelatihan model dikonfigurasi menggunakan skema Stratified 5-Fold Cross-Validation untuk mencegah terjadinya overfitting. Pada tahapan optimasi algoritma, pencarian ruang parameter (hyperparameter space) dilakukan melalui 30 iterasi percobaan (trials) menggunakan algoritma Tree-structured Parzen Estimator (TPE) yang tertanam di dalam kerangka kerja Optuna. Ruang pencarian (search space) difokuskan pada parameter kunci XGBoost yang secara langsung

memengaruhi kompleksitas pohon dan regularisasi, meliputi: learning_rate, max_depth, n_estimators, subsample, dan colsample_bytree. Seluruh skenario komputasi dalam eksperimen ini dieksekusi menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka integratif Scikit-learn, XGBoost, dan Optuna.

2.3. Pengumpulan Data

Penelitian ini menggunakan data sekunder berbasis teks (tabular) yang berfokus pada analitik sumber daya manusia. Dataset yang digunakan adalah "IBM HR Analytics Employee Attrition & Performance" yang bersifat publik. Kumpulan data ini merepresentasikan profil historis karyawan yang terdiri dari 1470 baris data (observasi) dan mencakup variabel demografis, kepuasan kerja, metrik kinerja, serta kompensasi finansial. Objek yang dianalisis dalam penelitian ini adalah probabilitas variabel dependen (target) yaitu "Attrition" (keputusan *resign* atau bertahan) berdasarkan pola variabel independen (fitur) yang melekat pada masing-masing karyawan.

2.4. Analisa Data

Tahap analisa awal dilakukan untuk mengidentifikasi karakteristik data. Karena mayoritas algoritma *machine learning* beroperasi menggunakan komputasi matematis, seluruh variabel kategorikal berbasis teks (seperti departemen atau status pernikahan) ditransformasikan menjadi representasi numerik menggunakan teknik *One-Hot Encoding*.

Analisis eksploratori menunjukkan adanya rasio ketidakseimbangan kelas (*imbalanced data*) yang ekstrem pada target "Attrition", di mana jumlah kelas mayoritas (karyawan yang bertahan) mendominasi secara signifikan dibandingkan kelas minoritas (karyawan yang *resign*). Untuk mencegah model menghasilkan prediksi yang bias terhadap kelas mayoritas, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan. SMOTE bekerja dengan cara mensintesis sampel buatan baru untuk kelas minoritas dengan mengekstraksi jarak fitur antar tetangga terdekat (*k-nearest neighbors*), sehingga distribusi kelas pada data latih (*training set*) menjadi seimbang dan siap untuk diumpankan ke dalam algoritma klasifikasi.

2.5. Tahapan Pemodelan dan Optimasi Parameter

Tahap ini merupakan inti dari penyelesaian masalah, di mana model prediktif dibangun menggunakan algoritma *Extreme Gradient Boosting* (XGBoost). Algoritma ini beroperasi dengan membangun sekumpulan pohon keputusan (*decision trees*) secara berurutan, di mana setiap pohon baru dirancang untuk memperbaiki sisa kesalahan (*residual error*) dari pohon sebelumnya.

Untuk mencapai performa maksimal, konfigurasi *hyperparameter* pada XGBoost harus diatur secara presisi. Penelitian ini meninggalkan pendekatan konvensional (*Grid Search*) dan menerapkan *Bayesian Optimization*. Proses optimasi ini dieksekusi menggunakan pustaka Optuna, yang memanfaatkan probabilitas dari evaluasi riwayat sebelumnya untuk memilih titik parameter terbaik secara otomatis dan lebih efisien. Parameter yang dioptimasi meliputi *n_estimators, max_depth, learning_rate, subsample*, dan *colsample_bytree*, dengan target fungsi objektif untuk memaksimalkan nilai evaluasi *F1-Score* melalui mekanisme *Cross-Validation*.

2.6. Tahapan Evaluasi dan Interpretasi (*Explainable AI*)

Model klasifikasi yang telah dioptimasi dievaluasi performanya menggunakan matriks kebingungan (*Confusion Matrix*) dengan metrik pengukuran berupa *Precision*, *Recall*, dan *F1-Score*. Karena tujuan utama bisnis adalah mencegah kehilangan karyawan, metrik *Recall* pada kelas *attrition* menjadi indikator paling krusial untuk menekan angka *False Negatives* (karyawan berisiko *resign* yang gagal dideteksi).

Guna mengatasi sifat *black-box* pada XGBoost, pendekatan *Explainable AI* diaplikasikan menggunakan pustaka SHAP (*SHapley Additive exPlanations*). Metode ini menghitung nilai Shapley dari teori permainan kooperatif untuk mendistribusikan kontribusi setiap fitur terhadap prediksi akhir. Persamaan dasar dari pendekatan SHAP dapat didefinisikan sebagai penambahan atribut, seperti pada Persamaan (1).

$$g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j z'_j \quad (1)$$

di mana g adalah model penjelasan, z' bernilai 1 atau 0 yang menunjukkan ada atau tidaknya fitur terkait, M adalah jumlah fitur penyederhanaan, dan $\emptyset_j$ merupakan nilai atribusi fitur (*SHAP value*).

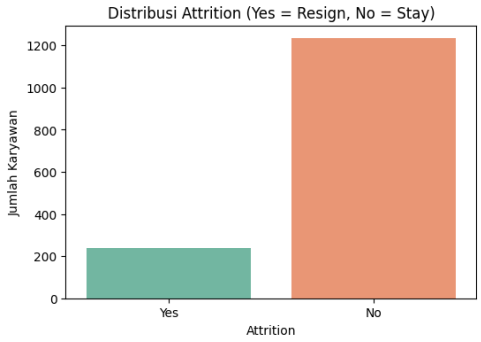
Penerapan metode ini bertujuan untuk mengekstrak alasan spesifik pemicu *turnover* karyawan secara transparan, sehingga menghasilkan visualisasi yang

3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil eksperimen komputasi yang telah dilakukan, mulai dari penanganan ketidakseimbangan data, optimasi hyperparameter menggunakan Bayesian Optimization, perbandingan kinerja metrik klasifikasi, hingga interpretasi model prediktif menggunakan Explainable AI (SHAP).

3.1. Hasil Prapemrosesan Data

Eksplorasi kelas target pada dataset asli menunjukkan rasio ketimpangan yang signifikan (1233 sampel kelas mayoritas/bertahan berbanding 237 sampel kelas minoritas/resign), seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Grafik Batang Distribusi Attrition

Implementasi algoritma SMOTE secara eksklusif pada set data latih berhasil mensintesis jumlah observasi kelas minoritas. Hasil akhirnya mendistribusikan set latih secara seimbang dengan rasio observasi 986:986, mengeliminasi bias mayoritas sebelum tahap training.

3.2. Optimasi Model (Bayesian Tuning)

Proses Bayesian Optimization dijalankan selama 30 iterasi (trials). Konvergensi parameter optimal yang ditemukan oleh Optuna meliputi: n_estimators = 164, max_depth = 5, learning_rate = 0.083, subsample = 0.505, dan colsample_bytree = 0.757. Penggunaan parameter tersebut terbukti mereduksi kompleksitas komputasi sekaligus meningkatkan kualitas segmentasi pohon pada XGBoost.

3.3. Evaluasi Kinerja Klasifikasi

Pengujian dilakukan terhadap data uji (testing set) yang tidak diseimbangkan, guna menguji daya generalisasi model secara riil. Tabel 1 memaparkan komparasi performa antara XGBoost baseline dan XGBoost yang telah dioptimasi.

TABEL I. PERBANDINGAN KINERJA MODEL XGBOOST

Metrik (Kelas Resign)	XGBoost Baseline	XGBoost (Bayesian)
Akurasi Global	85%	84%
Precision	54%	48%
Recall	30%	34%
F1-Score	38%	40%
False Negative	33 Observasi	31 Observasi

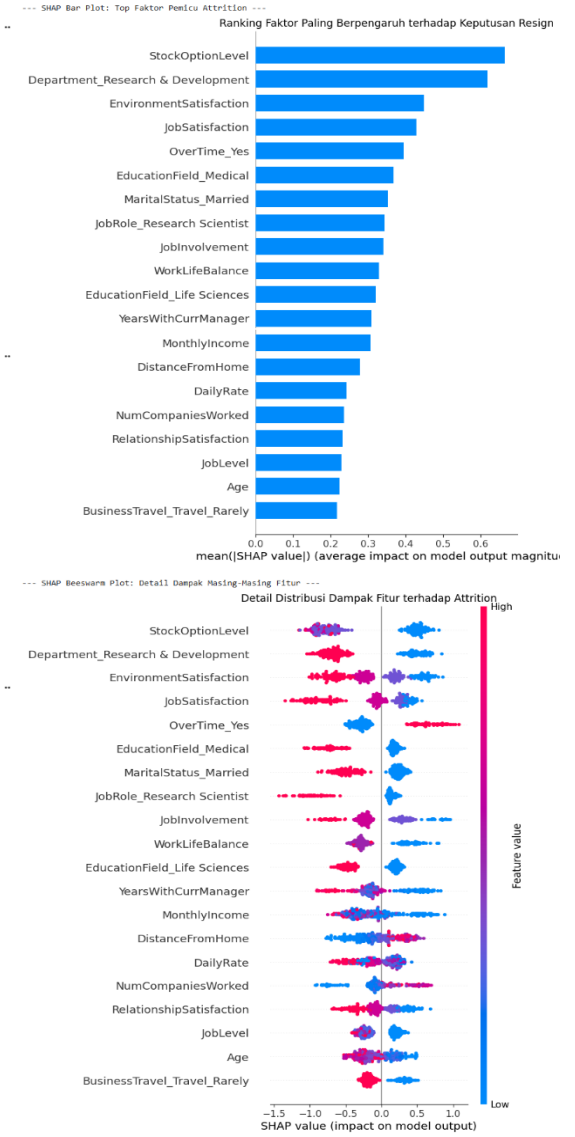
Berdasarkan Tabel 1, model teroptimasi dengan Bayesian Tuning mencatatkan peningkatan Recall dari 30% menjadi 34% dan F1-Score dari 38% menjadi 40% dibandingkan model baseline. Meskipun secara kuantitatif persentase peningkatan ini bersifat marjinal, penurunan jumlah False Negative dari 33 menjadi 31 observasi menegaskan efektivitas metode optimasi dalam menggeser batas deteksi untuk lebih peka terhadap kelas minoritas (karyawan yang akan resign). Dalam konteks analitik HR, mereduksi False Negative sangat krusial karena biaya kerugian akibat kegagalan mendeteksi karyawan bertalenta yang akan keluar umumnya jauh lebih besar daripada risiko salah memprediksi karyawan yang bertahan (False Positive).

Namun demikian, capaian metrik Recall yang tertahan pada angka 34% menyoroti sebuah batasan fundamental (limitation) dalam pemodelan prediktif ini. Nilai tersebut secara obyektif mengindikasikan bahwa model masih belum berhasil mendeteksi sekitar 66% dari total karyawan yang secara aktual mengundurkan diri. Keterbatasan ini mengonfirmasi bahwa fenomena employee attrition melibatkan dinamika psikososial yang sangat kompleks. Atribut tabular struktural yang tersedia di dalam dataset (seperti jarak tempuh, peran pekerjaan, atau tingkat pendidikan) memiliki batas maksimal (puncak empiris) dalam memetakan niat dan motivasi seorang manusia. Banyak faktor penentu pengunduran diri yang bersifat tidak terukur secara kuantitatif (unquantifiable variables), seperti dinamika politik internal, konflik antar-personal, masalah keluarga, hingga tawaran kompensasi tak terduga dari kompetitor, yang tidak terekam dalam basis data historis.

Oleh karena itu, efektivitas model XGBoost teroptimasi ini harus diposisikan secara proporsional. Model ini tidak dirancang sebagai prediktor deterministik absolut, melainkan sebagai sistem peringatan dini (early warning system) berbantuan Explainable AI yang menyaring dan menyoroti anomali pola data. Implementasi model ini tetap membutuhkan intervensi dan validasi kualitatif dari pihak manajemen HR sebelum keputusan retensi dieksekusi.

3.4. Analisis Explainable AI (SHAP)

Ekstraksi nilai dari model menggunakan TreeExplainer menghasilkan pemetaan visual seperti pada Gambar 3.



Gambar 3. Visualisasi SHAP Bar Plot dan Beeswarm Plot

Berdasarkan analisis SHAP summary plot, probabilitas pengunduran diri karyawan pada IBM secara komputasional diregulasi oleh lima fitur dominan:

a. StockOptionLevel: Fitur ini menempati peringkat urgensi tertinggi. Ketiadaan alokasi opsi saham (StockOptionLevel = 0) berkorelasi linear dengan tingginya nilai prediktif resign.
b. Department: Berada di luar departemen riset (Research & Development) meningkatkan kecenderungan turnover.
c. Environment & JobSatisfaction: Skor numerik kepuasan yang rendah secara logis mendorong kelas prediksi ke arah positif (resign).
d. OverTime: Frekuensi lembur yang tinggi terdeteksi sebagai anomali beban kerja yang secara langsung memperbesar metrik probabilitas pengunduran diri.

3.5. Analisis Kelebihan dan Kelemahan Sistem

Integrasi XGBoost, SMOTE, dan Bayesian Tuning menawarkan efisiensi komputasi yang tinggi tanpa mengorbankan integritas data. Hadirnya modul SHAP menggeser paradigma model dari sekadar alat klasifikasi probabilitas menjadi sistem diagnostik yang actionable bagi pihak manajemen.

Kelemahan matematis dari pemodelan ini terletak pada Recall yang terhenti di angka 34%. Hal ini mengindikasikan bahwa fitur-fitur struktural dalam basis data IBM (tabular) tidak secara komprehensif menangkap seluruh probabilitas heuristik dari perilaku manusia, yang kemungkinan dipengaruhi oleh parameter eksternal di luar atribut dataset.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Penelitian ini berhasil mengembangkan model prediktif klasifikasi *employee attrition* yang efisien, tangguh, dan transparan. Pendekatan pada level data menggunakan SMOTE terbukti krusial dalam menangani ketimpangan kelas, sementara optimasi *hyperparameter* berbasis *Bayesian Tuning* terbukti jauh lebih efisien dan mampu meningkatkan sensitivitas algoritma XGBoost dengan capaian akhir *Recall* 34% dan *F1-Score* 40%. Di luar pencapaian performa akurasi klasifikasi, kebaruan dan kontribusi utama

penelitian ini terletak pada integrasi *Explainable AI* (SHAP) yang berhasil membedah mekanisme *black-box* algoritma *ensemble*. Sistem secara presisi mampu memetakan atribusi fitur, menyoroti ketiadaan opsi kepemilikan saham, ketidakpuasan terhadap lingkungan kerja, dan tingginya beban jam lembur sebagai akar masalah utama pemicu pengunduran diri. Pendekatan ini memberikan landasan diagnostik berbasis data yang konkret bagi pihak manajemen.

4.2. Saran

Berdasarkan temuan dan keterbatasan metrik *Recall* yang dicapai, penelitian ini merekomendasikan beberapa arah pengembangan di masa depan. Pertama, mengingat data tabular terstruktur memiliki batas empiris dalam memetakan dinamika psikososial manusia, penelitian selanjutnya disarankan untuk mengintegrasikan pendekatan data kualitatif dan tidak terstruktur (seperti analisis sentimen dari ulasan teks karyawan, transkrip wawancara *exit interview*, atau rekam jejak komunikasi internal) guna memperkaya fitur prediktif. Kedua, penerapan kerangka kerja *Explainable AI* dapat diperluas menggunakan algoritma interpretasi alternatif (seperti LIME atau *TreeSHAP* tingkat lanjut) untuk melakukan validasi silang terhadap konsistensi fitur dominan yang dihasilkan.

5. UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas PGRI Jombang atas dukungan finansial dan fasilitas administrasi yang telah diberikan. Pelaksanaan dan penyelesaian penelitian ini dapat terwujud secara optimal berkat dukungan pendanaan penuh yang bersumber dari program Hibah Penelitian Lembaga Tahun 2026. Kontribusi lembaga tersebut sangat berperan krusial dalam menunjang seluruh tahapan teknis operasional penelitian, mulai dari proses komputasi, evaluasi algoritma, hingga tahap diseminasi dan publikasi hasil akhir penelitian.

Daftar Pustaka:

[1] P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *Algorithms*, vol. 4, no. 5, p. C5, 2016.

[2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[3] M. Al-Ali, M. Alwateer, S. A. Alsaedi, H. M. Balaha, M. Badawy, and M. A. Elhosseini, "Integrating machine learning and explainable AI for employee attrition prediction in HR analytics: M. AL-Ali et al.," *Sci. Rep.*, vol. 16, no. 1, p. 6344, 2026.

[4] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. fusion*, vol. 58, pp. 82–115, 2020.

[5] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Adv. Neural Inf. Process. Syst.*, vol. 24, 2011.

[6] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[7] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[8] D. Chung, J. Yun, J. Lee, and Y. Jeon, "Predictive model of employee attrition based on stacking ensemble learning," *Expert Syst. Appl.*, vol. 215, p. 119364, 2023.

[9] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 17, no. 3, p. 37, 1996.

[10] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv Prepr. arXiv1807.02811*, 2018.

[11] S. Garg, S. Sinha, A. K. Kar, and M. Mani, "A review of machine learning applications in human resource management," *Int. J. Product. Perform. Manag.*, vol. 71, no. 5, pp. 1590–1610, 2022.

[12] M. Haris, "Analysis of The Application Of Hyperparameter Tuning in Machine Learning to Increase The Accuracy of Sales-Level Prediction," *J. Inform. Rekayasa Elektron.*, vol. 7, no. 1, pp. 98–106, 2024.

[13] S. M. Lundberg and S.-I. Lee, "A unified

approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[14] W. Narkbunnum and K. Hinthaw, "Interpretable gradient boosted modeling of employee attrition: A SHAP-based framework for HR analytics," *Int. J. Anal. Appl.*, vol. 23, p. 236, 2025.

[15] M. I. Rifatama, M. R. Faisal, R. Herteno, I. Budiman, and M. I. Mazdadi, "Optimasi Algoritma K-Nearest Neighbor Dengan Seleksi Fitur Menggunakan Xgboost," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 64–72, 2023.

[16] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

[17] D. Taye and B. Getnet, "The Impact of Employee Turnover on Organizational Performance: A Case Study of Mada Walabu University, Bale Robe, Ethiopia," *Am. J. Pure Appl. Biosci.*, vol. 2, no. 3, pp. 51–63, 2020.

[18] N. R. P. Wardoyo, B. C. Perdana, and A. Nabilah, "KLASIFIKASI TINGKAT STRES PADA DATASET AFFECTNET MENGGUNAKAN HOG, SVM, DAN PSO," *J. Manaj. Inform. dan Sist. Inf.*, vol. 9, no. 2, pp. 248–261, 2026.

[19] N. R. P. Wardoyo, J. Santoso, and E. I. Setiawan, "Klasifikasi Micro-Expression Menggunakan K-Nearest Neighbors Menggunakan Fitur CAS dan HOG," *J. Intell. Syst. Comput.*, vol. 5, no. 2, pp. 96–103, 2023.

[20] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019.