

BAB I

PENDAHULUAN

1.1 Latar Belakang

Turnover karyawan menjadi salah satu tantangan utama yang dihadapi organisasi dalam mencapai tujuan, karena berdampak langsung pada efektivitas dan efisiensi perusahaan. *Turnover* adalah keluarnya karyawan dari perusahaan, yang dapat memberikan dampak baik positif maupun negatif (Dessler, 2013). *Turnover* positif terjadi ketika karyawan dengan kinerja di bawah standar meninggalkan perusahaan, memberikan peluang bagi perusahaan untuk merekrut pengganti yang lebih kompeten. Namun, secara umum, *turnover* lebih sering merugikan perusahaan, terutama ketika karyawan berkinerja tinggi yang keluar, sehingga perusahaan harus menanggung biaya perekrutan dan pelatihan kembali (Titah Ayu, 2022). Selain itu, *turnover* karyawan dapat menular, di mana kecenderungan karyawan untuk keluar dari organisasi bisa memicu rekan kerja lainnya untuk mengikuti, terutama apabila rekan yang keluar tersebut terlihat meraih kesuksesan di tempat kerja yang baru (Nasution, 2009).

Fenomena *turnover* karyawan menjadi tantangan signifikan bagi perusahaan, terutama dalam mempertahankan tenaga kerja dari generasi muda. Data dari (Deloitte GenZ Millennial Survey, 2025) menunjukkan bahwa 31% Gen Z dan 17% milenial secara aktif berencana untuk meninggalkan pekerjaannya dalam dua tahun ke depan. Angka ini mencerminkan tingkat *turnover* yang masih tinggi di kalangan generasi muda dan menuntut perhatian khusus dari organisasi.

Di tingkat nasional, temuan dari (Indonesia Millennial and Gen Z Report, 2025) juga mengindikasikan kondisi serupa. Laporan tersebut mencatat bahwa kenaikan biaya hidup menjadi tantangan utama bagi 66% responden, sementara 41% Gen Z mengalami kesulitan menabung dan 45% milenial mengaku pendapatannya tidak mencukupi untuk kebutuhan dasar. Tekanan ekonomi tersebut berdampak pada kepuasan kerja dan memicu keinginan untuk berpindah pekerjaan, khususnya jika tidak disertai dengan peluang pengembangan diri atau lingkungan kerja yang mendukung.

Selaras dengan temuan Deloitte, faktor-faktor utama yang mendorong niat *resign* dari kalangan Gen Z dan milenial mencakup aspek finansial (*money*), makna dalam pekerjaan (*meaning*), serta kesejahteraan fisik dan mental (*well-being*). Ketidakpuasan terhadap keseimbangan hidup, stres kerja yang tinggi, serta kurangnya prospek pengembangan karier menjadi pemicu utama dari keputusan tersebut. Oleh karena itu, penting bagi perusahaan untuk memahami kebutuhan dan preferensi generasi muda agar dapat merancang strategi retensi yang adaptif dan berkelanjutan.

Turnover karyawan yang tinggi dapat memiliki dampak serius bagi perusahaan jika masalah ini dibiarkan. Pertama, perusahaan harus mengalokasikan dana lebih besar untuk perekrutan dan pelatihan karyawan baru, yang mungkin menimbulkan dampak negatif pada anggaran dan sumber daya yang tersedia (Abiyyu, 2023). Selain itu, *turnover* yang tinggi dapat mengganggu produktivitas tim, mengakibatkan penurunan moral di antara karyawan yang tersisa, dan menciptakan atmosfer kerja yang tidak stabil (Anusha & Rajesh, 2024). Hal ini juga bisa

berdampak pada reputasi perusahaan di mata calon karyawan dan membuat perusahaan sulit dalam menarik bakat terbaik ke depannya. Dengan demikian, mengabaikan masalah *turnover* dapat membawa konsekuensi jangka panjang yang merugikan bagi pertumbuhan dan keberlanjutan organisasi.

Sebagai respons terhadap permasalahan ini, berbagai penelitian mulai memanfaatkan algoritma *machine learning* untuk memprediksi potensi *turnover* karyawan secara lebih akurat dan proaktif. Salah satu algoritma yang banyak digunakan adalah *Random Forest*, yang terbukti mampu mencapai akurasi tinggi hingga 93% dalam skenario prediksi *turnover* karyawan, serta menunjukkan performa yang lebih unggul dibandingkan algoritma lain seperti KNN dalam hal *precision*, *recall*, dan *F1-score* (Jamroni dkk., 2025). Keunggulan *Random Forest* terletak pada kemampuannya dalam mengenali pola kompleks dan mengidentifikasi fitur penting yang berpengaruh terhadap keputusan karyawan untuk keluar. Selain itu, algoritma *XGBoost* juga menunjukkan performa yang sangat baik dalam menangani data tidak seimbang, terutama setelah penerapan teknik *resampling* seperti SMOTE, sebagaimana dilaporkan oleh (Maahiroh, 2024). Temuan-temuan tersebut menguatkan alasan penggunaan kedua algoritma ini karena masing-masing memiliki keunggulan yang saling melengkapi, *Random Forest* pada stabilitas model dan interpretabilitas fitur, sedangkan *XGBoost* pada akurasi dan efisiensi dalam mempelajari interaksi fitur yang kompleks.

Namun demikian, penerapan algoritma *machine learning* dalam permasalahan prediksi *turnover* karyawan tidak terlepas dari tantangan ketidakseimbangan kelas (*imbalanced data*). Menurut (Kumar V dkk., 2024), proses klasifikasi bertujuan

membangun model prediktif untuk menentukan label kelas suatu data, namun dalam praktiknya distribusi kelas sering kali tidak seimbang, sehingga model cenderung bias terhadap kelas mayoritas dan kurang optimal dalam memprediksi kelas minoritas. Kondisi ini menjadi krusial dalam konteks *turnover* karyawan, karena kegagalan mengidentifikasi karyawan yang berpotensi keluar dapat berdampak besar terhadap pengambilan keputusan manajerial. Permasalahan tersebut menunjukkan perlunya teknik penyeimbangan data (*data balancing*) yang tidak hanya mengurangi bias model, tetapi juga meningkatkan representasi kelas minoritas agar pola data dapat dipelajari secara lebih proporsional. Salah satu pendekatan yang umum digunakan adalah dengan menambahkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Penelitian terdahulu menunjukkan bahwa *Synthetic Minority Over-sampling Technique* (SMOTE) efektif digunakan untuk mengatasi permasalahan tersebut, karena mampu membentuk sampel sintetis di sekitar ruang fitur kelas minoritas, bukan sekadar melakukan duplikasi data (Maahiroh, 2024). Oleh karena itu, penelitian ini menerapkan SMOTE pada tahap pra-pemrosesan data sebagai dasar untuk meningkatkan kualitas pembelajaran model *machine learning* yang digunakan dalam prediksi *turnover* karyawan.

Penerapan SMOTE mampu meningkatkan kualitas data latih dan membantu model dalam menangani ketidakseimbangan kelas, namun performa prediksi tetap sangat dipengaruhi oleh karakteristik algoritma yang digunakan. Kompleksitas data *turnover* karyawan yang melibatkan banyak faktor dengan hubungan non-linear serta distribusi kelas yang tidak seimbang menyebabkan model tunggal sering kali

belum mampu memberikan hasil yang optimal secara konsisten. Dalam konteks ini, algoritma *Random Forest* dan *XGBoost* memang telah menunjukkan kinerja yang baik secara individual, tetapi keduanya tetap memiliki keterbatasan tertentu. *Random Forest* mampu membangun model yang stabil dan mengurangi risiko *overfitting*, namun terkadang kurang mampu menangkap pola yang sangat kompleks. Di sisi lain, *XGBoost* unggul dalam akurasi dan efisiensi komputasi, tetapi masih berpotensi menghasilkan bias apabila digunakan secara tunggal. Perbedaan karakteristik tersebut menunjukkan bahwa tidak ada satu algoritma yang secara konsisten unggul dalam seluruh kondisi data, sehingga diperlukan pendekatan yang mampu mengombinasikan keunggulan dari berbagai model. Oleh karena itu, strategi *ensemble learning* menjadi penting untuk meningkatkan kemampuan generalisasi model serta menghasilkan prediksi yang lebih stabil dan akurat dalam menghadapi kompleksitas permasalahan *turnover* karyawan.

Pendekatan *ensemble learning* sendiri terdiri dari tiga kategori utama, yaitu *Bagging*, *Boosting*, dan *Stacking*, yang masing-masing menggabungkan beberapa model untuk meningkatkan performa prediksi. *Bagging* mengurangi variansi dengan membangun model secara paralel, sedangkan *Boosting* menurunkan bias melalui pelatihan model secara berurutan dengan fokus pada kesalahan sebelumnya. Berbeda dari keduanya, *Stacking* mengombinasikan keluaran beberapa model dasar melalui *meta-learner* yang mempelajari pola prediksi lintas model sehingga mampu menghasilkan keputusan akhir yang lebih akurat. Sejumlah penelitian juga menunjukkan bahwa *Stacking* memberikan performa lebih unggul, sebagaimana ditunjukkan oleh (Sudarto & Kusrini, 2024) dan (Biswas dkk., 2023)

yang melaporkan bahwa *Stacking* melampaui *Random Forest*, *Decision Tree*, dan model individual lainnya. Oleh karena itu, *Stacking* dipilih dalam penelitian ini karena mampu menggabungkan kelebihan *Random Forest* dan *XGBoost* secara lebih komprehensif dibandingkan *Bagging* dan *Boosting* yang hanya mengintegrasikan model pada satu lapisan.

Dalam arsitektur *Stacking*, *meta-learner* berfungsi menggabungkan *output* dari model dasar untuk menghasilkan prediksi akhir yang lebih akurat. Pemilihan jenis *meta-learner* menjadi krusial, karena model inilah yang menentukan bagaimana pola dari *Random Forest* dan *XGBoost* disintesis. Berbagai penelitian menunjukkan bahwa *meta-learner* linear seperti *Logistic Regression* kerap kurang optimal ketika menghadapi pola hubungan non-linear antar fitur atau antar prediksi model dasar. Temuan ini diperkuat oleh penelitian (Hossen dkk., 2021) yang membandingkan berbagai algoritma diantaranya *Logistic Regression*, SVM, KNN, *Naive Bayes*, *Decision Tree*, dan MLP, di mana hasilnya menunjukkan bahwa MLP memberikan performa paling tinggi dalam kasus klasifikasi. Selain itu, studi (Intharasompong dkk., 2025) juga menunjukkan bahwa dari beberapa *meta-learner* yang diuji, termasuk *Logistic Regression*, *Gradient Boosting*, dan *XGBoost*, MLP secara konsisten menjadi *meta-learner* terbaik karena kemampuannya menangkap interaksi non-linear secara lebih efektif. Berdasarkan bukti empiris tersebut, penelitian ini memilih MLP sebagai *meta-learner* karena memiliki kemampuan generalisasi yang lebih kuat dibandingkan model linear dan terbukti memberikan performa unggul pada penelitian terdahulu.

Meskipun kombinasi *Random Forest* dan *XGBoost* dalam metode *Stacking* telah terbukti memberikan performa yang baik pada berbagai studi, masih terdapat ruang pengembangan pada aspek *Hyperparameter Tuning* dan pemilihan *meta-learner*. Penelitian sebelumnya umumnya menggunakan metode *tuning* konvensional seperti *GridSearch* atau *RandomSearch*, serta *meta-learner* berbasis *regresi linear* yang kurang mampu menangkap hubungan non-linear antar *base learner*. Sementara itu, meskipun studi (Khurshid dkk., 2025) dan (Intharasompong dkk., 2025) membuktikan bahwa *Bayesian Optimization* dapat meningkatkan performa model, penerapannya pada arsitektur *Stacking* dengan kombinasi *Random Forest*, *XGBoost* dan *meta-learner* MLP belum dieksplorasi secara spesifik. Oleh karena itu, *novelty* utama penelitian ini terletak pada integrasi *Stacking Random Forest*, *XGBoost* dengan *meta-learner* MLP yang dioptimalkan menggunakan *Bayesian Optimization*, sehingga menghasilkan pendekatan baru yang lebih efektif untuk meningkatkan akurasi dan stabilitas prediksi *turnover* karyawan dibandingkan model *Stacking* pada penelitian terdahulu.

Masalah penelitian yang menjadi fokus penelitian ini mengarah pada bagaimana perbandingan performa antara model dasar *Random Forest* dan *XGBoost*, model gabungan *Stacking Ensemble* dengan *meta-learner* MLP, serta model *Stacking* yang telah dioptimasi menggunakan *Bayesian Optimization* dalam memprediksi *turnover* karyawan. Melalui pendekatan ini, penelitian ini bertujuan menghasilkan model yang mampu memberikan prediksi yang lebih akurat dan stabil, sekaligus menawarkan wawasan komprehensif mengenai efektivitas

kombinasi *Random Forest* dan *XGBoost* dalam kerangka *Stacking* yang diperkuat dengan optimasi *hyperparameter* berbasis *Bayesian Optimization*.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, dapat dirumuskan masalah pada penelitian ini, yaitu:

- a. Bagaimana performa model *Random Forest*, *XGBoost*, serta *Ensemble Stacking* sebelum dan sesudah dilakukan *hyperparameter tuning* dengan *Bayesian Optimization* dalam memprediksi *turnover* karyawan?
- b. Faktor-faktor apa saja yang paling signifikan dalam memengaruhi keputusan karyawan untuk meninggalkan perusahaan?

1.3 Tujuan Penelitian

Dari uraian masalah yang sudah dirumuskan, didapat tujuan penelitian sebagai berikut:

- a. Membandingkan performa model *Random Forest*, *XGBoost*, serta *Ensemble Stacking* sebelum dan sesudah dilakukan *hyperparameter tuning* dengan *Bayesian Optimization* dalam memprediksi *turnover* karyawan.
- b. Mengidentifikasi faktor-faktor yang paling signifikan yang memengaruhi keputusan karyawan untuk meninggalkan perusahaan.

1.4 Manfaat Penelitian

Adapun beberapa manfaat yang didapatkan dari penelitian ini, sebagai berikut:

- a. Bagi Perusahaan

- Membantu perusahaan mengidentifikasi karyawan yang berpotensi untuk keluar.
- Memungkinkan perusahaan untuk mengambil tindakan preventif untuk meningkatkan retensi karyawan.
- Mengoptimalkan alokasi sumber daya untuk program retensi karyawan.

b. Bagi Akademisi

- Memperkaya literatur mengenai penerapan *machine learning* dalam bidang manajemen sumber daya manusia.
- Memberikan kontribusi terhadap pengembangan model prediksi yang lebih baik untuk fenomena *turnover* karyawan.

1.5 Batasan Masalah

Berikut merupakan batasan masalah yang terdapat pada penelitian ini:

- a. Tipe *turnover* yang diteliti adalah *voluntary turnover* atau *turnover* sukarela.
- b. Dataset yang digunakan adalah *Employee Attrition Dataset* (IBM HR Analytics) yang bersumber dari *Kaggle*, dengan jumlah data sebanyak 1.470 karyawan dan 35 variabel. Variabel target berupa *Attrition (Yes/No)*, sedangkan variabel prediktor meliputi aspek demografi (seperti usia dan jenis kelamin), kompensasi (misalnya gaji bulanan), beban kerja (antara lain rata-rata jam kerja bulanan), serta berbagai variabel lain yang berkaitan dengan karakteristik karyawan dan faktor-faktor yang memengaruhi keputusan untuk bertahan atau meninggalkan perusahaan.

- c. Model prediksi dalam penelitian ini dibatasi pada penggunaan algoritma *Random Forest*, *XGBoost*, serta metode *Stacking Ensemble* yang menggabungkan kedua algoritma tersebut. Model di luar ketiga pendekatan ini tidak digunakan.
- d. Proses *hyperparameter* tuning hanya memanfaatkan metode *Bayesian Optimization*, tanpa melibatkan teknik optimasi lainnya seperti *Random Search* maupun *Grid Search*.
- e. *Meta learner* dalam penelitian ini secara khusus dibatasi pada penggunaan *Multi-Layer Perceptron* (MLP) sebagai model tingkat lanjut yang menggabungkan keluaran dari kedua *base learner*.
- f. Pendekatan *Stacking Ensemble* dalam penelitian ini hanya menggunakan kombinasi dua *base learner*, yaitu *Random Forest* dan *XGBoost*, serta menerapkan MLP sebagai satu-satunya *meta learner* tanpa melibatkan variasi metode lain.
- g. Ruang lingkup penelitian hanya mencakup analisis dan perbandingan performa model, tanpa mengembangkan hasil pemodelan menjadi sebuah sistem prediksi yang siap diimplementasikan secara operasional.