

Analisis Perbandingan Model Random Forest dan XGBoost dalam Memprediksi Turnover Karyawan

Muhamad Bayu Setiawan, Alam Rahmatulloh*

Program Studi Informatika, Fakultas Teknik

Universitas Siliwangi

217006083@student.unsil.ac.id, alam@unsil.ac.id*

Abstrak

Turnover karyawan merupakan tantangan besar bagi organisasi, karena memengaruhi produktivitas, stabilitas tim, dan menimbulkan biaya besar terkait perekrutan, orientasi, dan pelatihan. Untuk mengatasi masalah ini, analisis prediktif menggunakan *machine learning* menawarkan pendekatan yang ampuh untuk mengetahui faktor-faktor yang memengaruhi *turnover* karyawan dan memungkinkan strategi berbasis data untuk retensi karyawan. Studi ini mengevaluasi kinerja prediktif dua model *machine learning* tingkat lanjut, *Random Forest* dan *XGBoost*, menggunakan dataset karyawan. Model-model ini dipilih karena kemampuannya menangani data berdimensi tinggi, menangkap hubungan *non-linier*, dan memberikan hasil klasifikasi yang andal. Performa model dinilai menggunakan metrik utama seperti *accuracy*, *precision*, *recall*, dan *f1-score*, dengan peningkatan tambahan yang dicapai melalui *tuning hyperparameter*. Hasilnya menunjukkan bahwa *XGBoost* sedikit mengungguli *Random Forest*, dengan mencapai akurasi 98,83% dibandingkan 98,80%, yang menunjukkan ketangguhannya dalam tugas prediktif. Meskipun terdapat perbedaan kecil, kedua model menunjukkan potensi yang kuat dalam mengatasi prediksi *turnover*. Hal ini menunjukkan pentingnya pendekatan berbasis data dalam pengambilan keputusan organisasi. Dengan memanfaatkan model ini, perusahaan dapat mengembangkan strategi yang ditargetkan untuk meminimalkan risiko pergantian karyawan, meningkatkan retensi karyawan, dan meningkatkan keberhasilan organisasi.

Kata Kunci : *Turnover Karyawan, Hyperparameter Tuning, Prediksi, Random Forest, XGBoost.*

Abstract

Employee turnover is a major challenge for organizations, as it impacts productivity, team stability, and incurs huge costs related to recruiting, onboarding, and training. To address this issue, predictive analytics using machine learning offers a powerful approach to understand the factors that influence employee turnover and enable data-driven strategies for employee retention. This study evaluates the predictive performance of two advanced machine learning models, Random Forest and XGBoost, using employee datasets. These models were chosen for their ability to handle high-dimensional data, capture non-linear relationships, and provide reliable classification results. Model performance is assessed using key metrics such as accuracy, precision, recall, and f1-score, with additional improvements achieved through hyperparameter tuning. The results show that XGBoost slightly outperforms Random Forest, achieving an accuracy of 98.83% versus 98.80%, indicating its robustness in predictive tasks. Despite small differences, both models show strong potential in addressing turnover prediction. This shows the importance of a data-driven approach in organizational decision making. By leveraging this model, companies can develop targeted strategies to minimize the risk of employee turnover, increase employee retention, and increase organizational success.

Keywords: *Employee Turnover, Hyperparameter Tuning, Prediction, Random Forest, XGBoost.*

1. Pendahuluan

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 Gary, 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 (Kusumaningtyas 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 bila mereka melihat rekan yang keluar meraih kesuksesan di tempat baru (Nasution, 2009). Fenomena ini menjadi semakin relevan di era modern, di mana mobilitas karyawan semakin tinggi dan persaingan untuk mendapatkan talenta terbaik semakin ketat.

Fenomena ini menjadi masalah bagi setiap perusahaan, Data Deloitte dari tahun 2019 mengungkapkan bahwa tingkat *turnover* global untuk karyawan milenial berada di angka 21%, yang secara signifikan lebih tinggi daripada Generasi X (13%) dan Baby Boomers (7%) (Manggabarani et al., 2020), (Zulaikha et al., 2021). Menurut penelitian dari *IDN Research Institute* (2019), 35% karyawan milenial di Indonesia memiliki niat untuk mengundurkan diri dari posisi mereka saat ini dalam kurun waktu dua tahun. Angka-angka ini mengindikasikan bahwa generasi milenial memiliki karakteristik unik yang memengaruhi tingkat *turnover*, seperti preferensi terhadap fleksibilitas kerja, budaya organisasi, dan peluang pengembangan karier. Statistik ini menekankan pentingnya bagi perusahaan untuk memahami variabel-variabel yang berkontribusi terhadap tingginya tingkat *turnover* karyawan milenial dan menciptakan metode yang efisien untuk mempertahankan mereka.

Turnover karyawan dapat menimbulkan dampak signifikan terhadap produktivitas dan stabilitas perusahaan (Maria, 2023). Tingginya tingkat *turnover* dapat mengakibatkan biaya yang sangat besar, baik dalam hal rekrutmen dan pelatihan karyawan baru, maupun dalam hilangnya pengetahuan dan pengalaman yang dimiliki karyawan yang pergi (Abiyyu, 2023a). Selain itu, *turnover* dapat memengaruhi moral karyawan yang tersisa, menciptakan lingkungan kerja yang tidak stabil dan menurunkan semangat kerja. Oleh karena itu, pemahaman yang lebih baik mengenai faktor-faktor yang mempengaruhi *turnover* karyawan sangat penting bagi manajemen sumber daya manusia. Penelitian ini bertujuan untuk memberikan wawasan yang lebih dalam mengenai perilaku *turnover* karyawan dan bagaimana teknologi analitik dapat dimanfaatkan untuk memprediksi pergerakan tersebut.

Meskipun banyak perusahaan telah menyadari pentingnya memprediksi *turnover* karyawan, banyak yang masih menghadapi kesulitan dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap keputusan karyawan untuk bertahan atau keluar dari organisasi. Berbagai metode analisis data telah digunakan, namun tidak semua metode dapat menangani kompleksitas data yang tinggi dan hubungan non-linear antar variabel (Zhao et al., 2019). Penggunaan metode analitik yang kurang tepat sering kali menghasilkan prediksi yang tidak akurat, sehingga perusahaan gagal mengambil langkah-langkah pencegahan yang efektif. Dalam konteks ini, pemilihan model yang tepat menjadi krusial untuk meningkatkan akurasi prediksi *turnover* (Louak, 2021).

Penelitian-penelitian sebelumnya telah mengeksplorasi penggunaan model pembelajaran mesin untuk memprediksi *turnover* karyawan, salah satunya adalah *Support Vector Machine (SVM)*. Model SVM terbukti mampu memberikan prediksi *turnover* dengan akurasi dasar sebesar 0,56 untuk berbagai metrik evaluasi, seperti akurasi, *presisi*, *recall*, dan *f1-score* tanpa seleksi fitur. Seleksi fitur melalui metode *wrapper* menunjukkan peningkatan akurasi hingga 0,60, sementara metode filter dan *embedded* mengalami penurunan performa

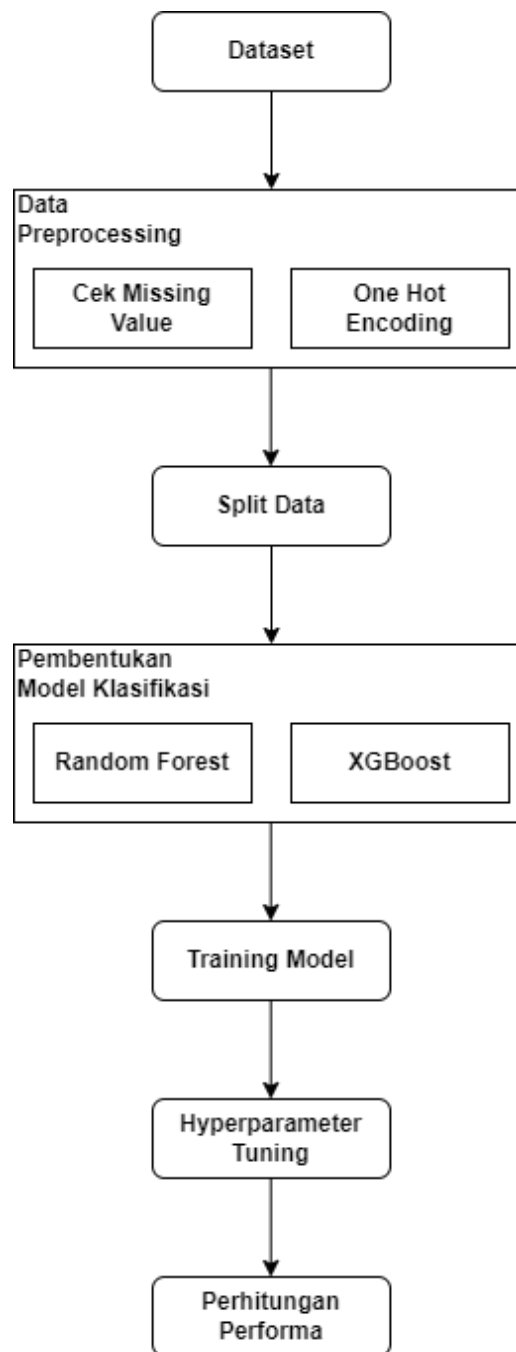
hingga 0,55. Variasi ini menunjukkan pentingnya pemilihan metode seleksi fitur yang tepat dalam meningkatkan performa model prediksi *turnover* (Abiyyu, 2023b). Namun, hasil tersebut masih memiliki ruang untuk perbaikan, terutama dengan penerapan teknik-teknik canggih seperti *hyperparameter tuning* yang belum sepenuhnya diterapkan dalam penelitian tersebut.

Dalam penelitian ini, metode *Random Forest* dan *XGBoost* digunakan dan dikombinasikan dengan proses *hyperparameter tuning* menggunakan *GridSearch* untuk memperoleh konfigurasi parameter optimal. Kedua metode ini dipilih karena kemampuannya dalam menangani data berdimensi tinggi serta hubungan non-linear antar variabel, yang diharapkan dapat meningkatkan akurasi prediksi *turnover*. Dataset yang digunakan mencakup variabel-variabel seperti tingkat kepuasan kerja, jumlah proyek, waktu kerja bulanan rata-rata, dan lama bekerja. Proses analisis meliputi tahapan persiapan data, pembagian data menjadi set pelatihan dan pengujian, pembentukan model dengan *GridSearch* untuk tuning parameter, serta evaluasi model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*.

Permasalahan dalam penulisan ini adalah adanya dampak negatif dan ketidakstabilan yang dialami oleh organisasi yang disebabkan oleh *turnover* yaitu permasalahan biaya pengeluaran yang semakin besar dan karyawan yang melakukan *turnover* mengajak karyawan yang lainnya yang masih bertahan. Maka tujuan dari penulisan ini agar perusahaan dapat merancang strategi yang lebih baik untuk meningkatkan retensi karyawan dan kinerja organisasi secara keseluruhan serta mengetahui faktor yang mempengaruhi terjadinya *turnover* dan strategi apa yang dilakukan untuk mencegah terjadinya *turnover*.

2. Metode Penelitian

Penelitian ini mengikuti beberapa tahapan penting, dimulai dari proses pengumpulan dataset hingga perhitungan performa. Gambar 1 menunjukkan alur metodologi penelitian.



Gambar 1. Alur metode penelitian

a. Dataset

Dataset yang digunakan dalam penelitian ini berisi 14,999 data pegawai dengan 10 fitur yang menggambarkan karakteristik individu, informasi pekerjaan, serta variabel target yang terkait dengan *turnover* karyawan. Variabel-variabel yang terdapat dalam dataset ini mencakup tingkat kepuasan kerja karyawan (*satisfaction_level*) dan skor evaluasi terakhir mereka (*last_evaluation*). Kedua variabel ini

penting karena tingkat kepuasan yang rendah atau skor evaluasi yang ekstrem bisa menjadi alasan seorang karyawan mempertimbangkan untuk keluar dari perusahaan. Selain itu, jumlah proyek yang diikuti karyawan (*number_project*) dan jam kerja rata-rata per bulan (*average_monthly_hours*) memberikan gambaran mengenai beban kerja, yang jika terlalu tinggi atau rendah dapat menyebabkan ketidakpuasan dan keinginan untuk mencari pekerjaan lain.

Dataset ini juga mencakup lama waktu yang telah dihabiskan karyawan di perusahaan (*time_spend_company*), yang bisa menjadi indikator potensi kejenuhan atau stagnasi, serta riwayat kecelakaan kerja (*Work_accident*) yang dapat memengaruhi kenyamanan dan keamanan kerja. Variabel target yang menjadi fokus penelitian adalah *left*, yaitu apakah seorang karyawan telah keluar atau masih bekerja di perusahaan, yang merupakan sasaran prediksi untuk mengidentifikasi potensi *turnover*. Selain itu, dataset ini mencatat apakah karyawan menerima promosi dalam lima tahun terakhir (*promotion_last_5years*), di mana kurangnya promosi bisa menyebabkan ketidakpuasan dan mendorong keinginan untuk pindah kerja.

Variabel tambahan lainnya adalah *sales*, yang menunjukkan departemen kerja karyawan, serta *salary*, yang menggambarkan tingkat gaji (*low, medium, high*). Kedua variabel ini relevan karena beberapa departemen dan tingkat gaji mungkin menunjukkan pola *turnover* yang lebih tinggi, mencerminkan kondisi atau kepuasan finansial yang memengaruhi retensi karyawan. Secara keseluruhan, variabel-variabel ini dipilih karena relevansinya dalam menganalisis faktor-faktor yang berkontribusi terhadap *turnover* karyawan, sehingga model prediksi dapat mengidentifikasi pola-pola dari data untuk membantu organisasi dalam strategi retensi.

b. Data Preprocessing

Dataset kemudian diimpor dan dilakukan serangkaian langkah preprocessing untuk memastikan kualitas data yang optimal sebelum digunakan dalam analisis. Salah satu langkah penting dalam preprocessing adalah

pengecekan nilai-nilai yang hilang (*missing values*), yang dapat memengaruhi akurasi model jika tidak ditangani dengan baik. Setelah itu, transformasi variabel kategorikal, seperti “*sales*” dan “*salary*,” dilakukan menggunakan metode *one-hot encoding*.

sales_IT	sales_RandD	...	sales_hr	sales_management	sales_marketing
0.0	0.0	...	0.0	0.0	0.0
0.0	0.0	...	0.0	0.0	0.0
0.0	0.0	...	0.0	0.0	0.0
0.0	0.0	...	0.0	0.0	0.0
0.0	0.0	...	0.0	0.0	0.0

Gambar 2. Hasil One-hot encoding.

Transformasi ini mengonversi setiap kategori dalam variabel menjadi kolom biner (*dummy variables*), sehingga data dapat dengan mudah diolah oleh algoritma machine learning, yang umumnya hanya menerima input berupa nilai numerik. Langkah ini sangat penting karena variabel kategorikal yang tidak diubah menjadi bentuk numerik tidak dapat dimengerti oleh model machine learning dan dapat menyebabkan kesalahan dalam prediksi (광 호 et al., 2019).

c. Split Data

Setelah tahap pra-pemrosesan, dataset dibagi menjadi dua subset yaitu 80% digunakan untuk data pelatihan dan 20% untuk data pengujian. Pembagian ini bertujuan untuk melatih model pada satu subset dan menguji kinerjanya pada subset yang lain, sehingga dapat menghindari *overfitting* (Ursenbach et al., 2019).

d. Pembentukan Model Klasifikasi

Pada tahap ini dilakukan Pembentukan model dengan menginisialisasi model *Random Forest* dan *XGBoost* dengan parameter *default*. Model *Random Forest* diinisialisasi menggunakan *RandomForestClassifier()*, sementara model *XGBoost* diinisialisasi menggunakan *XGBClassifier()*.

e. Training Model

Setelah model dibentuk, dilakukan pelatihan model. Pada tahap ini, masing-masing model dilatih menggunakan data *training*. Model *Random Forest* dilatih dengan *rf_model.fit(X_train, y_train)*, sementara

model *XGBoost* dilatih dengan $xgb_model.fit(X_train, y_train)$. Setelah pelatihan, dilakukan evaluasi awal model berdasarkan akurasi untuk mengetahui kinerja awal masing-masing model (Guerra et al., 2023).

f. Hyperparameter Tuning

Tahap berikutnya adalah *hyperparameter tuning*, di mana dilakukan optimasi parameter untuk meningkatkan performa model (Schratz et al., 2019). Untuk *Random Forest*, *tuning* dilakukan menggunakan *GridSearchCV*, yang menguji berbagai parameter seperti $n_estimators$, max_depth , $min_samples_split$, dan $min_samples_leaf$. Proses yang sama dilakukan untuk model *XGBoost*, di mana parameter yang diuji meliputi $learning_rate$, $n_estimators$, max_depth , $subsample$, dan $colsample_bytree$. Hasil dari *tuning* ini menghasilkan model dengan parameter terbaik untuk masing-masing algoritma.

g. Perhitungan Performa

Pada tahap ini, dilakukan evaluasi performa untuk algoritma *Random Forest* dan *XGBoost*. Metrik performa yang dihitung meliputi akurasi, *precision*, *recall*, dan *F1 Score*. Berikut adalah masing-masing rumus untuk menghitung akurasi, *precision*, *recall*, dan *F1 Score*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Metrik evaluasi *accuracy* dalam klasifikasi menghitung proporsi prediksi benar yang dibuat oleh model, yang didefinisikan sebagai satu dikurangi tingkat kesalahan standar (ER). Ini mencerminkan efektivitas sistem klasifikasi secara keseluruhan dalam mengidentifikasi kelas dengan benar.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Metrik evaluasi *precision* dalam klasifikasi menghitung rasio prediksi positif sebenarnya terhadap total prediksi positif yang dibuat oleh pengklasifikasi. Ini mengukur keakuratan prediksi positif, menunjukkan berapa banyak prediksi yang benar.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Metrik evaluasi *recall* dalam klasifikasi

menghitung proporsi kejadian positif sebenarnya yang diidentifikasi dengan benar oleh model dari semua kejadian positif aktual. Ini mengukur kemampuan model untuk menangkap semua kasus yang relevan di kelas minoritas.

$$F1\ Score = \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

F1 Score dalam klasifikasi mengevaluasi keseimbangan antara *precision* dan *recall*, memberikan metrik tunggal yang mencerminkan keakuratan prediksi positif dan kemampuan untuk mengidentifikasi semua kejadian yang relevan, yang penting untuk menilai performa model.

Dalam hal ini, TP merujuk pada *True Positive*, TN pada *True Negative*, FP pada *False Positive*, dan FN pada *False Negative* (Chicco et al., 2021).

3. Hasil dan Pembahasan

a. Klasifikasi

Model *machine learning* dievaluasi menggunakan empat metode pengukuran yaitu akurasi, presisi, *recall*, dan *f1-score*. Berikut hasil evaluasi yang dilakukan sebelum dan setelah *hyperparameter tuning* yang ditunjukkan oleh tabel 1 dan 2.

Tabel 1. Sebelum di *hyperparameter tuning*

	Accuracy	Precision	Recall	F1 Score
<i>Random Forest</i>	0.9880	0.9855	0.9631	0.9742
<i>XGBoost</i>	0.9863	0.9770	0.9645	0.9707

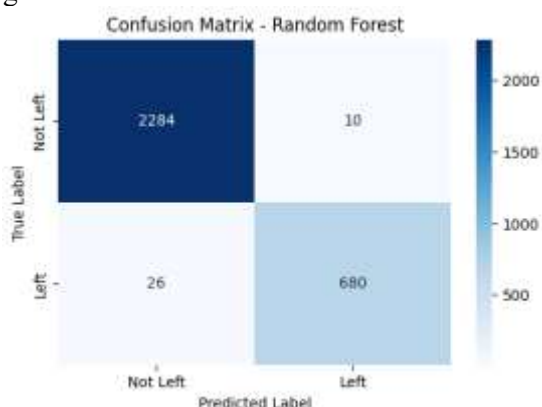
Tabel 2. Sebelum di *hyperparameter tuning*

	Accuracy	Precision	Recall	F1 Score
<i>Random Forest</i>	0.9880	0.9855	0.9631	0.9742
<i>XGBoost</i>	0.9883	0.9841	0.9660	0.9749

Berdasarkan dua tabel diatas *hyperparameter tuning* secara signifikan meningkatkan performa model *Random Forest* dan *XGBoost* dalam memprediksi *turnover* karyawan. Sebelum *tuning*, *Random*

Forest mencapai akurasi 0.9767, *precision* 0.9667, *recall* 0.9481, dan *F1-score* 0.9573. Setelah *tuning*, akurasi meningkat menjadi 0.9883 (+1.19%), *precision* menjadi 0.9878 (+2.19%), *recall* menjadi 0.9655 (+1.84%), dan *F1-score* menjadi 0.9765 (+1.92%). *XGBoost* juga menunjukkan peningkatan performa setelah *tuning*. Akurasi meningkat dari 0.9893 menjadi 0.9907 (+0.14%), *precision* dari 0.9844 menjadi 0.9870 (+0.26%), *recall* dari 0.9714 menjadi 0.9762 (+0.48%), dan *F1-score* dari 0.9779 menjadi 0.9816 (+0.37%).

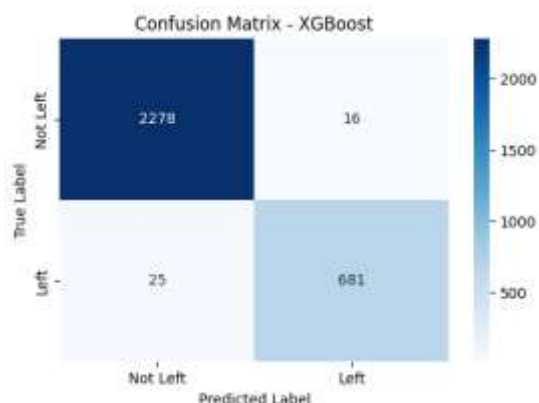
Berikut *confusion matrix* untuk model *Random Forest* yang ditunjukkan dengan gambar 2.



Gambar 3. Confusion Matrix Random Forest

Pada gambar 3, model *Random Forest* menunjukkan hasil prediksi dengan 2284 kasus "Not Left" yang diprediksi dengan benar (*True Negative*) dan hanya 10 kasus "Not Left" yang salah diprediksi sebagai "Left" (*False Positive*). Untuk kelas "Left," *Random Forest* berhasil memprediksi 680 kasus dengan benar (*True Positive*), namun masih salah memprediksi 26 kasus "Left" sebagai "Not Left" (*False Negative*). Secara keseluruhan, *Random Forest* memiliki akurasi yang tinggi dalam mengklasifikasikan kedua kelas, dengan kesalahan minor di kedua jenis kesalahan (*False Positive* dan *False Negative*).

Selain itu berikut adalah *confusion matrix* untuk model *XGBoost* yang ditunjukkan dengan gambar 3.



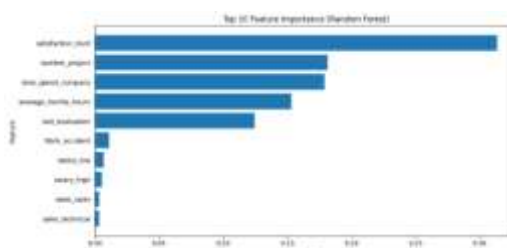
Gambar 4. Confusion Matrix XGBoost

Pada gambar 4, model *XGBoost* memiliki hasil prediksi yang juga sangat baik dan mirip dengan *Random Forest*. Model ini berhasil memprediksi 2278 kasus "Not Left" dengan benar, tetapi salah mengklasifikasikan 16 kasus "Not Left" sebagai "Left". Untuk kelas "Left," *XGBoost* mampu memprediksi 681 kasus dengan benar dan membuat kesalahan pada 25 kasus, yang salah diklasifikasikan sebagai "Not Left".

Dari hasil *Confusion Matrix* ini, terlihat bahwa kedua model memiliki performa yang sangat mirip, dengan perbedaan yang sangat kecil dalam jumlah kesalahan prediksi pada masing-masing kelas. *Random Forest* tampaknya memiliki sedikit lebih banyak *True Negatives*, sedangkan *XGBoost* sedikit lebih unggul dalam *True Positives* untuk kelas "Left." Secara keseluruhan, kedua model ini cocok digunakan untuk tugas klasifikasi dengan perbedaan performa yang cukup tipis, sehingga pemilihan model bisa bergantung pada faktor lain, seperti interpretabilitas, kecepatan, atau preferensi aplikasi tertentu.

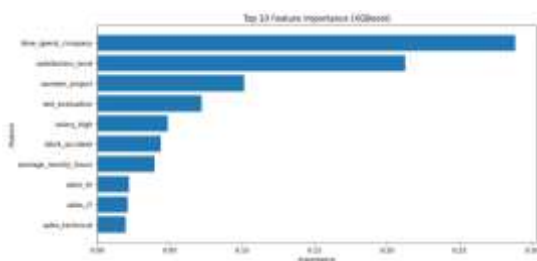
b. Fitur Penting

Visualisasi *feature importance* bertujuan untuk mengidentifikasi dan memahami faktor-faktor yang paling berpengaruh terhadap tingkat keluarnya karyawan.



Gambar 5. Top 10 Fitur Penting Random Forest

Pada gambar 5 Diagram batang untuk *Random Forest* menunjukkan bahwa fitur *satisfaction_level* memiliki nilai *importance* tertinggi, yang mengindikasikan bahwa tingkat kepuasan karyawan merupakan faktor paling dominan dalam memprediksi *turnover* karyawan. Fitur penting lainnya, seperti *number_project*, *time_spend_company*, *average_monthly_hours*, *last_evaluation*, dan *Work_accident* juga memiliki pengaruh yang signifikan.



Gambar 6. Top 10 Fitur Penting XGBoost

Pada gambar 6 Diagram batang untuk *XGBoost* juga menunjukkan bahwa *satisfaction_level* merupakan fitur terpenting. Namun, urutan dan nilai *importance* untuk beberapa fitur lainnya sedikit berbeda dibandingkan dengan *Random Forest*. Misalnya, fitur *last_evaluation* memiliki *importance* yang lebih tinggi pada *XGBoost*, sedangkan *number_project* memiliki *importance* yang lebih rendah.

4. Kesimpulan

Penelitian ini membandingkan dua model *machine learning*, yaitu *Random Forest* dan *XGBoost*, untuk memprediksi *turnover* karyawan menggunakan dataset kepegawaian dari *Kaggle*. Berdasarkan hasil evaluasi, kedua model menunjukkan performa yang baik dalam memprediksi keluarnya karyawan, dengan *XGBoost* sedikit lebih unggul dalam hal akurasi, yaitu 98.83% dibandingkan *Random Forest* yang mencapai 98.80%. Metrik lainnya, seperti *precision*, *recall*, dan *F1-score*, juga menunjukkan bahwa kedua model mampu menghasilkan prediksi yang akurat, namun *XGBoost* lebih konsisten dalam mempertahankan keseimbangan antara *precision* dan *recall*. Faktor yang paling berpengaruh terhadap keputusan karyawan

untuk keluar adalah *satisfaction_level* dan *time_spend_company*, sehingga hal ini dapat menjadi fokus bagi manajemen dalam merancang strategi retensi.

Meskipun penelitian ini memberikan wawasan yang berharga, terdapat beberapa limitasi yang perlu dicatat. Pertama, dataset yang digunakan hanya mencakup data dari satu sumber, yaitu *Kaggle*, yang mungkin tidak sepenuhnya mewakili kondisi di perusahaan lain. Kedua, analisis ini tidak mempertimbangkan faktor eksternal yang mungkin memengaruhi *turnover*, seperti kondisi ekonomi atau perubahan dalam industri. Ketiga, meskipun *hyperparameter tuning* dilakukan, kemungkinan masih ada parameter lain yang dapat dioptimalkan lebih lanjut untuk meningkatkan akurasi model.

Untuk penelitian selanjutnya, disarankan untuk melakukan analisis dengan menggunakan dataset yang lebih beragam dari berbagai industri untuk mendapatkan hasil yang lebih *generalizable*. Selain itu, penggabungan algoritma lain, seperti *Neural Networks* atau *ensemble methods* yang lebih kompleks, dapat dieksplorasi untuk melihat apakah dapat meningkatkan akurasi prediksi. Penelitian juga dapat mempertimbangkan variabel eksternal dan interaksi antar variabel untuk memodelkan faktor-faktor yang lebih komprehensif dalam memprediksi *turnover* karyawan.

Daftar Pustaka

- Abiyyu, A. S. (2023a). Perbandingan Metode Seleksi Fitur untuk Mengoptimasi Model Support Vector Machine dalam Memprediksi Turnover Pegawai. *EProceedings of Engineering*, 10(2), 1921. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20040>
- Abiyyu, A. S. (2023b). *Perbandingan Metode Seleksi Fitur untuk Mengoptimasi Model Support Vector Machine dalam Memprediksi Turnover Pegawai* (Vol. 10, Issue 2).
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable

- than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14, 1–22. <https://doi.org/10.1186/s13040-021-00244-z>
- Dessler Gary. (2013). *Human Resource Management, 13th Edition* (13th ed.). Boston : Pearson Education., 2013.
- Guerra, E., Wilhelmi, F., Miozzo, M., & Dini, P. (2023). The Cost of Training Machine Learning Models Over Distributed Data Sources. *IEEE Open Journal of the Communications Society*, 4, 1111–1126. <https://doi.org/10.1109/OJCOMS.2023.3274394>
- Kusumaningtyas Titah Ayu. (2022). *Pengaruh Kepuasan Kerja Terhadap Turnover Intention Karyawan Generasi Milenial (Studi Pada Industri Finance Di Kecamatan Jombang)*. SEKOLAH TINGGI ILMU EKONOMI (STIE) PGRI DEWANTARA JOMBANG.
- Louak, M. J. (2021). *Employee Turnover Prediction with Machine Learning Algorithms*.
- Manggabarani, A. S., Marzuki, F., & Mahendro. (2020). The Effect Of Milenial Generation Characteristics Of Job Satisfaction Relations And Employee Engagement. *Ilomata International Journal of Management*, 1(4), 239–248. <https://doi.org/10.52728/ijjm.v1i4.166>
- Maria, R. (2023). Dampak Turnover Karyawan pada Bisnis dan Cara Mengatasinya. *Greatdayhr*.
- Nasution, W. A. (2009). Pengaruh Kepuasan Kerja Karyawan Terhadap Intensi Turnover Pada Call Center Telkomsel di Medan. *Jurnal MANDIRI*, 4.
- Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, 109–120. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>
- Ursenbach, J., O’Connell, M. E., Neiser, J., Tierney, M. C., Morgan, D., Kosteniuk, J., & Spiteri, R. J. (2019). Scoring algorithms for a computer-based cognitive screening tool: An illustrative example of overfitting machine learning approaches and the impact on estimates of classification accuracy. *Psychological Assessment*, 31(11), 1377–1382. <https://doi.org/10.1037/pas0000764>
- Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). *Employee Turnover Prediction with Machine Learning: A Reliable Approach* (pp. 737–758). https://doi.org/10.1007/978-3-030-01057-7_56
- Zulaikha, S., Nasrudin, M., & Lestari, E. P. (2021). Dialectic of Public Sphere vs Private Sphere in Millennial Families in Lampung Customary Law Community. *Samarah*, 5(2), 928–956. <https://doi.org/10.22373/sjhc.v5i2.9150>
- 광 호김, 병 훈장, 황 규최, Ho Kim, K., Chang, B., & Kyu Choi, H. (2019). Deep Learning Based Short-Term Electric Load Forecasting Models using One-Hot Encoding. *Online j.Inst.Korean.Electr.Electron.Eng*, 23(3), 19–22. <https://doi.org/10.7471/ikeee.2019.23.3.852>