

Analisa Komparasi Kinerja Model *Logistic Regression* dan *Random Forest* dalam Memprediksi Risiko *Turnover* Karyawan

Omar Pahlevi¹, Yuni Fitriani², Dewi Ayu Nur Wulandari³, Handini Widyastuti⁴, Sri Utami⁵, Astriana Mulyani⁶

^{1,2,3,4,5} Universitas Bina Sarana Informatika

Jalan Kramat Raya No. 98, Kec. Senen, Jakarta Pusat, Indonesia

e-mail: ¹omar.opi@bsi.ac.id, ²yuni.yfi@bsi.ac.id, ³dewi.dan@bsi.ac.id, ⁴handini.hwy@bsi.ac.id, ⁵sri.sut@bsi.ac.id

⁶ Universitas Nusa Mandiri

Jalan Jatiwaringin Raya No. 2, Cipinang Melayu, Kec. Makasar, Jakarta Timur, Indonesia

e-mail: ⁶astriana.atm@nusamandiri.ac.id

Artikel Info : Diterima : 15-11-2025 | Direvisi : 25-11-2025 | Disetujui : 30-11-2025

Abstrak - *Turnover* karyawan merupakan salah satu nilai tolak ukur bagi keberhasilan suatu perusahaan dalam menjalankan kegiatan bisnisnya. Memprediksi *turnover* karyawan merupakan kegiatan signifikan yang penting bagi setiap perusahaan yang berkelanjutan, dimana informasi dini tentang status *turnover* karyawan memungkinkan organisasi untuk mengambil langkah-langkah pencegahan. Pada penelitian ini peneliti akan mengaplikasikan dan membandingkan dua buah model algoritma *supervised learning*, diantaranya model algoritma *Logistic Regression* dan *Random Forest* untuk memprediksi risiko *turnover* karyawan, kemudian membandingkan mana dari dua model algoritma tersebut yang paling akurat. Tujuan dilakukannya penelitian ini adalah untuk mengetahui model algoritma *Supervised Learning* yang akurat dan presisi dalam memprediksi risiko *turnover* karyawan. Berdasarkan hasil pengukuran kinerja kedua model dengan menggunakan metode *Confusion Matrix*. Berdasarkan hasil penelitian, diketahui bahwa model *Logistic Regression* memiliki tingkat akurasi sebesar 84,64% serta *F1-Score* yang baik dengan nilai sebesar 0,89, nilai presisi sebesar 0,82, dan nilai *recall* sebesar 0,96. Performa model *Random Forest* memiliki tingkat akurasi sebesar 80,12%, *F1-Score* sebesar 0,85 menunjukkan keseimbangan antara presisi sebesar 0,80 dan *recall* dengan nilai 0,92. Hal ini membuktikan bahwa model algoritma *Logistic Regression* adalah yang paling baik untuk untuk prediksi risiko *turnover* karyawan.

Kata Kunci : *Turnover* Karyawan, *Logistic Regression*, *Random Forest*, *Confusion Matrix*

Abstract - *Employee turnover* is one of the benchmarks for the success of a company in conducting its business activities. Predicting *employee turnover* is a significant activity that is important for every sustainable company, where early information about *employee turnover* status allows organizations to take preventive measures. In this study, the researcher will apply and compare two supervised learning algorithm models, namely the *Logistic Regression* and *Random Forest* algorithms, to predict *employee turnover* risk, then compare which of the two algorithm models is the most accurate. Based on the results of measuring the performance of both models using the *Confusion Matrix* method, it was found that the *Logistic Regression* model had an accuracy rate of 84.64% and a good *F1-Score* of 0.89, with a precision value of 0.82 and a recall value of 0.96. The *Random Forest* model had an accuracy rate of 80.12% and an *F1-Score* of 0.85, showing a balance between precision of 0.80 and recall of 0.92. This proves that the *Logistic Regression* algorithm model is the best for predicting *employee turnover* risk.

Keywords: *Employee Turnover* *Logistic Regression*, *Random Forest*, *Confusion Matrix*.



PENDAHULUAN

Dalam menyikapi tren pola mengenai pasar tenaga kerja yang berubah secara drastis dan signifikan, para pimpinan perusahaan semakin menyadari adanya kerumitan dalam hal mempertahankan karyawan. Pada masa ini dimana ketidakpastian ekonomi, kemajuan teknologi, dan globalisasi mempengaruhi dinamika tenaga kerja, mempertahankan karyawan yang memiliki kinerja yang baik telah menjadi komponen utama dalam kesuksesan bisnis (Agustini, 2024). Konsekuensinya selain dapat mengganggu proses operasional yang telah terintegrasi dengan teknologi, hal ini dapat menyebabkan kehilangan pengetahuan dan keterampilan yang telah dikumpulkan (Setiyadi et al., 2024).

Pada dunia bisnis, pergantian karyawan merupakan masalah serius karena dapat berdampak langsung pada pertumbuhan dan tingkat keuntungan yang diperoleh oleh perusahaan (Aditama et al., 2025). Selain meningkatkan biaya operasional, seperti biaya perekrutan dan orientasi karyawan baru, tingkat pergantian karyawan yang tinggi dapat menurunkan produktivitas. Tingkat pergantian karyawan yang tinggi dapat menurunkan kualitas kerja secara keseluruhan dan mengganggu kekompakan *teamwork*. Rasa ketidakpuasan karyawan terhadap tempat kerja, yang mendorong mereka mencari peluang di tempat lain, merupakan salah satu dari banyak penyebab yang sering menyebabkan masalah ini.

Indikator terkuat dari tingkat *turnover* karyawan yang sebenarnya adalah niat *turnover*, yang didefinisikan sebagai keinginan karyawan untuk meninggalkan perusahaan dalam jangka waktu tertentu (Lazzari et al., 2022). Meskipun dipertanyakan indikator ini masih sering muncul perihal data pengunduran diri yang spesifik terkadang tidak tersedia karena hal-hal seperti kebijakan privasi. Selain itu, karena satu hal terjadi sebelum yang lain, peramalan yang akurat tentang niat *turnover* memungkinkan pembuat kebijakan dan pemberi kerja untuk mengambil tindakan dan mencegah *turnover* yang sebenarnya.

Tingkat prosentase karyawan yang meninggalkan perusahaan dan digantikan oleh karyawan baru dikenal sebagai tingkat pergantian karyawan. Pergantian karyawan yang tinggi merupakan masalah yang dapat memberikan dampak negatif bagi suatu perusahaan (Atef et al., 2022). Hal ini dapat berdampak pada pemutusan hubungan kerja, kekosongan posisi, perekrutan, pelatihan, dan penggantian karyawan cukup mahal. Selain itu, perusahaan dengan tingkat *turnover* yang tinggi pada akhirnya kehabisan karyawan, yang membuatnya tidak produktif dan menghambat pertumbuhannya. Oleh karena itu, memprediksi *turnover* karyawan merupakan proses yang penting bagi setiap perusahaan yang berkelanjutan, dan informasi dini tentang status *turnover* karyawan memungkinkan organisasi untuk mengambil langkah-langkah pencegahan.

Turnover karyawan memiliki dampak langsung terhadap rencana strategis organisasi dan seringkali merupakan hasil dari keputusan perekrutan dan manajemen yang buruk. Hal ini berdampak pada manajemen yang buruk dan kesenjangan kinerja manajerial dalam pengambilan keputusan dapat meningkatkan tingkat perputaran karyawan (Pourkhodabakhsh et al., 2023). Dalam menghadapi tingkat perputaran karyawan, banyak perusahaan harus memodifikasi, meningkatkan, dan memodernisasi sistem manajemen mereka. Untuk mengatasi masalah ini, pemerintah dan perusahaan komersial menghabiskan miliaran dolar setiap tahun. Dapat dikatakan bahwa menjaga dan mengoptimalkan sumber daya yang mahal dan terbatas seperti sumber daya manusia adalah salah satu perhatian utama manajer, yang memerlukan keunggulan lebih dan alat inovatif untuk meningkatkan keputusan.

Salah satu potensi penggunaan pembelajaran mesin adalah penerapan analitik prediktif untuk mengidentifikasi karyawan yang paling mungkin meninggalkan pekerjaan mereka. Studi tentang sumber daya manusia dengan menganalisis data tentang variabel seperti tingkat keterlibatan, kemajuan karir, dan kepuasan kerja, dan menerapkan algoritma pembelajaran mesin dapat menentukan karyawan mana yang paling mungkin meninggalkan perusahaan. Hal ini memberikan kesempatan bagi staf sumber daya manusia untuk bertindak dan mengurangi tingkat *turnover* karyawan (Paigude et al., 2023). Hal ini dapat dicapai dengan menerapkan intervensi yang terfokus, seperti menawarkan kesempatan pelatihan dan pengembangan kepada karyawan atau menangani masalah spesifik yang mungkin mereka hadapi.

Pada penelitian ini, peneliti mencoba menggali permasalahan mengenai prediksi risiko *turnover* karyawan dengan mengimplementasikan dua buah model algoritma *supervised learning* dan dikomparasi agar dapat mengetahui model algoritma yang terbaik. Model algoritma *supervised learning* yang pertama yakni model *Logistic Regression*. Model algoritma ini cukup banyak diimplementasikan dan efektif pada berbagai bidang penelitian (Wardhani & Lhaksmama, 2022). Kemudian model algoritma *supervised learning* yang kedua yang digunakan pada penelitian ini yakni model algoritma *Random Forest*. Peneliti menggunakan model algoritma ini dikarenakan memiliki kehandalan dalam *handling* pada data *imbalance* (Kurniawan et al., 2025).

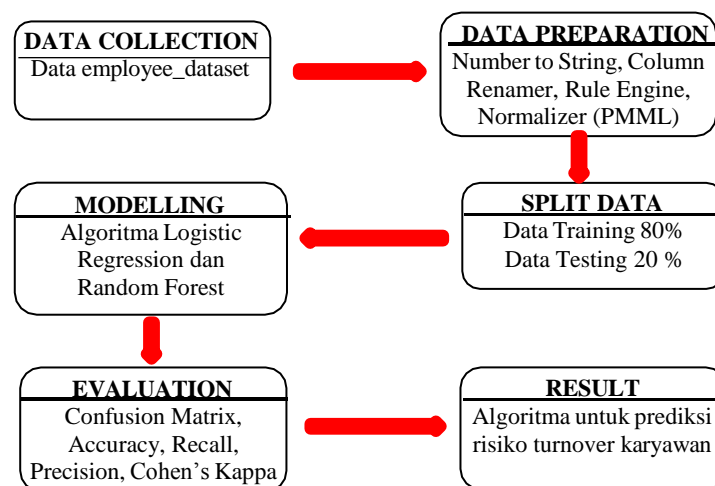
Penelitian yang terkait dengan studi ini, antara lain kegiatan penelitian yang dilakukan oleh (Chakraborty et al., 2021) dimana pada penelitian ini berbagai metode klasifikasi diperkenalkan, seperti *logistic regression*, analisis diskriminan linier, klasifikasi ridge, klasifikasi lasso, *decision tree*, dan *random forest*, untuk memprediksi dan mengukur secara bersamaan kemungkinan *turnover* setiap karyawan baru. Data dari departemen sumber daya manusia perusahaan yang tersedia di kaggle digunakan untuk memperkirakan *turnover* karyawan. Penelitian terkait yang berikutnya, yakni penelitian yang dilakukan oleh (Chung et al., 2023). Dalam studi ini, peneliti mengusulkan model untuk memprediksi tingkat *turnover* karyawan agar dapat mengambil langkah-langkah manajemen talenta yang sebelumnya dilakukan secara retrospektif. Model prediktif dibangun

berdasarkan 30 variabel yang memengaruhi tingkat *turnover* karyawan, yang diambil dari data IBM HR Analytics Employee Attrition & Performance, yang terdiri dari 1.470 record. Hasil akhir penelitian ini, total delapan model prediktif, termasuk *Logistic Regression*, *Random Forest*, *XGBoost*, *SVM*, *Neural Network*, dan model ensemble, dibangun dan kinerjanya dievaluasi.

Penelitian yang dilakukan oleh (Raza et al., 2022) dengan topik penelitian *Predicting Employee Attrition Using Machine Learning Approaches*, dimana memiliki kemiripan dalam hal mengidentifikasi penyebab tingginya tingkat *turnover* karyawan dan kerangka kerja pembelajaran untuk memprediksi tingkat *turnover* karyawan. Studi penelitian ini bertujuan untuk menganalisis faktor-faktor organisasi yang menyebabkan tingginya tingkat *turnover* karyawan dan memprediksi tingkat *turnover* karyawan menggunakan teknik pembelajaran mesin. Penelitian berikutnya yang berkaitan dengan topik penelitian ini, yaitu penelitian yang dilakukan oleh (Guerranti & Dimitri, 2022) dengan topik penelitian *A Comparison of Machine Learning Approaches for Predicting Employee Attrition*. Pembahasan pada penelitian ini mengenai pengumpulan himpunan data dengan jumlah observasi yang lebih besar, fitur tambahan, dan mungkin struktur relasional dapat menghasilkan hasil yang lebih baik. Meskipun demikian, penerapan model pembelajaran mesin untuk memprediksi tingkat *turnover* karyawan dapat memainkan peran fundamental dalam konteks pengambilan keputusan manajerial. Perusahaan dapat memutuskan untuk mengambil tindakan seperti menawarkan peluang kerja baru serta memperbaiki kondisi kerja yang buruk, dengan tujuan mengurangi tingkat *turnover* karyawan.

METODE PENELITIAN

Tahapan pada penelitian ini terdiri dari beberapa tahap yang tertuang pada gambar 1 kerangka pemikiran penelitian. Permasalahan pada penelitian ini mengacu pada komparasi dua buah model algoritma *supervised learning* yang presisi dalam memprediksi risiko *turnover* karyawan. Model algoritma yang dibandingkan yakni model algoritma *Logistic Regression* dan *Random Forest*. Berikut ini gambar 1 kerangka pemikiran penelitian.



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Pemikiran Penelitian

Berikut ini penjabaran mengenai tahapan-tahapan yang ada pada kerangka pemikiran:

- A. *Data collection* merupakan tahapan pertama pada penelitian ini, dimana peneliti dalam melakukan riset ini menggunakan data `employee_dataset.csv` yang diperoleh secara terbuka dan bersifat data sekunder pada laman <https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset/data>. *Dataset* ini mencakup informasi tentang karyawan suatu perusahaan, termasuk data demografis, karakteristik terkait pekerjaan, riwayat pekerjaan, dan latar belakang pendidikan.
- B. Tahapan selanjutnya dalam penelitian ini yakni *Data Preparation* merupakan tahapan yang berisikan *node Number to String* yang berfungsi untuk mentransformasikan angka dalam kolom (atau sekelompok kolom) menjadi teks, *node Column Renamer* yang digunakan untuk mengubah nama kolom sesuai dengan skema penamaan yang ditentukan dalam dialog. *Node* ketiga yang digunakan pada tahapan ini *node Rule Engine* berfungsi untuk mengambil daftar aturan yang didefinisikan pengguna dan mencoba mencocokkannya dengan setiap baris dalam tabel input. *Node* yang ke empat, yaitu *node Normalizer (PMML)* yang digunakan untuk memilih semua nilai kolom numerik agar dinormalisasi.

- C. Tahapan yang ketiga yaitu *Split Data*, dimana pada tahapan ini berisikan *node Table Partitioner* yang berfungsi untuk membagi *data training* dan *data testing*, dibuat dengan membagi tabel input secara baris demi baris.
- D. Tahapan selanjutnya pada Kerangka Pemikiran Penelitian ini yaitu *Modelling*, dimana berisikan dua buah model algoritma antara lain *Logistic Regression* dan *Random Forest*. Peneliti dalam riset ini menerapkan model algoritma *Logistic Regression*, dikarenakan model ini memiliki kemampuan dalam mengidentifikasi hubungan antara beberapa variabel (Anggara et al., 2022). Model algoritma lain yang digunakan pada penelitian ini yaitu algoritma *Random Forest*. Model algoritma ini digunakan pada penelitian ini, dikarenakan kemampuannya dalam hal mengurangi data berdimensi tinggi (Hossen et al., 2021). Tahapan berikutnya, yakni *Evaluation*, dimana berupa *Confusion Matrix*, *Accuracy*, *Recall*, *Precision* dan *Cohen's Kappa*.
- E. Kemudian tahapan terakhir yakni, *Result* merupakan hasil akhir komparasi algoritma *supervised learning* untuk prediksi *turnover* karyawan.

HASIL DAN PEMBAHASAN

Hasil dari komparasi dua buah algoritma pembelajaran mesin yakni model algoritma *Logistic Regression* dan *Random Forest* disajikan berupa hasil implementasi antarmuka dan hasil algoritma. Data masukan *employee data* dalam jumlah besar, yang disebut dataset, biasanya menjadi langkah pertama dalam siklus kinerja sistem. Himpunan data ini dibagi menjadi *data training* dan *data testing*. *Data training* digunakan untuk pembuatan dan pelatihan model, sedangkan *data testing* digunakan untuk proses klasifikasi model.

A. Data Collection

Pada penelitian ini menggunakan himpunan data *employee_dataset.csv* yang bersumber pada laman <https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset/data>. Himpunan data ini terdiri dari delapan fitur yang bersifat independent, antara lain *Joining Year*, *City*, *Payment Tier*, *Age*, *Gender*, *Ever Benched*, *Experience in Current Domain*. Fitur yang menjadi *class* atau labelnya yakni *LeaveOrNot* dengan nilai 0 untuk tetap dan nilai 1 untuk mengindikasikan keluar. Berikut ini tampilan gambar himpunan datanya.

#	RowID	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceIn...	LeaveOrNot
1	Row0	Bachelors	2017	Bangalore	3	34	Male	No	0	0
2	Row1	Bachelors	2013	Pune	1	28	Female	No	3	1
3	Row2	Bachelors	2014	New Delhi	3	38	Female	No	2	0
4	Row3	Masters	2016	Bangalore	3	27	Male	No	5	1
5	Row4	Masters	2017	Pune	3	24	Male	Yes	2	1
6	Row5	Bachelors	2016	Bangalore	3	22	Male	No	0	0
7	Row6	Bachelors	2015	New Delhi	3	38	Male	No	0	0
8	Row7	Bachelors	2016	Bangalore	3	34	Female	No	2	1
9	Row8	Bachelors	2016	Pune	3	23	Male	No	1	0
10	Row9	Masters	2017	New Delhi	2	37	Male	No	2	0
11	Row10	Masters	2012	Bangalore	3	27	Male	No	5	1
12	Row11	Bachelors	2016	Pune	3	34	Male	No	3	0
13	Row12	Bachelors	2018	Pune	3	32	Male	Yes	5	1
14	Row13	Bachelors	2016	Bangalore	3	39	Male	No	2	0
15	Row14	Bachelors	2012	Bangalore	3	37	Male	No	4	0
16	Row15	Bachelors	2017	Bangalore	1	29	Male	No	3	0
17	Row16	Bachelors	2014	Bangalore	3	34	Female	No	2	0
18	Row17	Bachelors	2014	Pune	3	34	Male	No	4	0

Sumber: Hasil Penelitian (2025)

Gambar 2. Tampilan Himpunan Data

B. Data Preparation

Pada tahapan ini terdapat tiga buah *node* yang digunakan. *Node* yang pertama yakni *node Number to String* digunakan untuk memilih fitur-fitur dengan tipe numerik menjadi string diantaranya *JoiningYear*, *PaymentTier*, *Age*, *ExperienceCurrentDomain*, dan *LeaveOrNot*. *Node* yang kedua yaitu *node Column Renamer* digunakan untuk mengubah nama fitur *LeaveOrNot* menjadi *status_karyawan*. *Node* yang ketiga yaitu *node Rule Engine* yang digunakan untuk mengubah data. Sintaks yang digunakan $\$status_karyawan > 0 \Rightarrow \text{"Tetap"}$. $\text{NOT } \$status_karyawan > 0 \Rightarrow \text{"Keluar"}$. Kemudian *node* yang ke empat yakni *node Normalizer (PMML)* untuk normalisasi data menggunakan *Z-Score*.

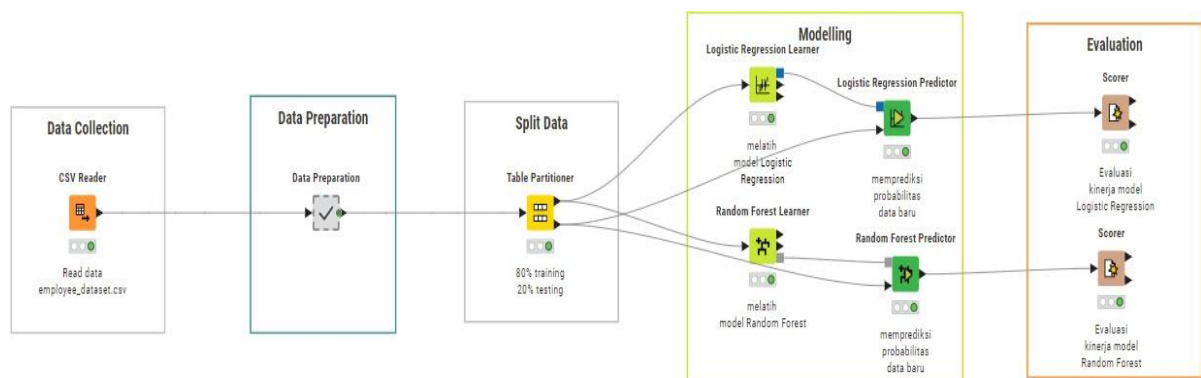
C. Split Data

Node yang digunakan pada tahapan ini yakni *node Table Partitioner*. Pada tahapan ini dilakukan pembagian

data latih sebesar 80 % dengan jumlah record sebanyak 3722, sedangkan data uji sebesar 20 % dengan jumlah record sebesar 931. *Overfitting* dapat dihindari dalam praktiknya dengan menggunakan berbagai teknik validasi model dan sebaiknya dibagi menjadi dua kategori: *dataset* pelatihan dan pengujian. *Hold-out validation* sebesar 80:20 merupakan metode yang paling sederhana dan paling cepat secara komputasi (Burzykowski et al., 2023).

D. *Modelling*

Pada tahapan *Modelling* ini menggunakan aplikasi *KNIME Analytics Platform* versi 5.5.2 mengimplementasikan dua buah model algoritma *Supervised Learning* yang memiliki *node* atau komponen yang terinci dalam *platform* tersebut. Model algoritma seperti *Logistic Regression* dan *Random Forest* memiliki *node* yang berbeda untuk setiap tahap dalam proses pengembangan model. Berikut ini gambar 2 menampilkan dua alur kerja model prediksi risiko *turnover* karyawan



Sumber: Hasil Penelitian (2025)

Gambar 3. Dua Alur Kerja Model Prediksi Risiko *Turnover* Karyawan menggunakan KNIME

Pada alur kerja, data *employee_dataset.csv* dibaca menggunakan *node csv reader*. Selanjutnya, himpunan data tersebut di filter untuk memisahkan data yang akan digunakan. Parameter ini memiliki sifat kualitatif dan kuantitatif yang beragam. Selanjutnya, ke tahap pengujian algoritma, dimana algoritma menggunakan *node* spesifik, yaitu, *Random Forest* menggunakan *node Random Forest Learner* dan *Predictor*, *Logistic Regression* menggunakan *node Logistic Regression Learner* dan *Predictor*. Hasil dari prediksi dapat diperoleh dengan menambahkan *node* *scorer* pada *node Predictor*.

E. *Evaluation*

Implementasi pembelajaran mesin dilakukan dengan variasi parameter untuk mencapai nilai akurasi tertinggi. Dimana hal ini membantu mengoptimalkan model serta mencegah *overfitting*. Kegiatan *evaluation* ini bertujuan memastikan bahwa model yang dihasilkan memiliki performa optimal dan mampu memberikan prediksi yang akurat pada data baru. Berikut ini tabel 1 mengenai hasil komparasi performa dua model algoritma pada pembelajaran mesin.

Tabel 1. Hasil Komparasi Performa Dua Model Algoritma

Algoritma	Accuracy	Recall	Precision	F1-Score	Cohens's Kappa
Random Forest	80,12%	0,92	0,80	0,85	0,53
Logistic Regression	84,64%	0,96	0,82	0,89	0,63

Sumber: Hasil Penelitian (2025)

Hasil pengujian dua buah model algoritma pembelajaran mesin menunjukkan performa yang berbeda dimana model algoritma *Logistic Regression* memiliki performa yang lebih baik dibandingkan dengan model algoritma *Random Forest*. Hasil pengujian tersebut dijabarkan sebagai berikut:

1. *Random Forest* memiliki tingkat akurasi sebesar 80,12%, yang menunjukkan tingkat keakuratan tinggi dalam melakukan klasifikasi. *F1-Score* yang tinggi (0,85) menunjukkan keseimbangan antara presisi (0,80) dan *recall* (0,92).

Logistic Regression juga memberikan hasil yang baik dengan akurasi sebesar 84,64%, serta *F1-Score* yang baik (0,89), presisi (0,82), dan *recall* (0,96) yang lebih baik dibandingkan model algoritma *Random Forest*.

KESIMPULAN

Pada hasil penelitian dan pengujian, performa model algoritma *Logistic Regression* untuk prediksi risiko turnover karyawan memberikan tingkat akurasi sebesar 84,64% serta *F1-Score* yang baik dengan nilai sebesar 0,89, nilai presisi sebesar 0,82, dan nilai *recall* sebesar 0,96. Performa model *Random Forest* memiliki tingkat akurasi sebesar 80,12%, nilai *F1-Score* sebesar 0,85 menunjukkan keseimbangan antara presisi sebesar 0,80 dan *recall* dengan nilai 0,92. Berdasarkan tingkat akurasi, *F1-Score*, nilai presisi, dan nilai *recall*, maka performa model algoritma *Logistic Regression* adalah yang paling baik untuk untuk prediksi risiko turnover karyawan. Kemudian usulan pengembangan pada penelitian ini, bisa dengan menerapkan model algoritma supervised learning yang lainnya misalnya model algoritma *Support Vector Machine* dan *Naive Bayes* serta dioptimalkan dengan model *Particle Swarm Optimization*.

REFERENSI

- Aditama, Praditya, A., Zein, M. R., Nadriani, M., Ghifari, O., Trisnawati, & Lubis, B. O. (2025). *Klasifikasi Potensi Turnover Karyawan Berdasarkan Data Kinerja Menggunakan Algoritma C4.5 Studi Kasus PT. Republic Technology Nusantara*. 9(2), 2723–2730.
- Agustini. (2024). Analisis Strategi Retensi Karyawan untuk Mengurangi Turnover dan Meningkatkan Loyalitas Karyawan. *Nobel Management Review*, 5(3), 306–320.
- Anggara, E. D., Widjaja, A., & Suteja, B. R. (2022). *Prediksi Kinerja sebagai Rekomendasi Kenaikan Golongan dengan Decision Tree dan Regresi Logistik*. 8(April), 218–234.
- Atef, M., Elzanfaly, D. S., & Ouf, S. (2022). Early Prediction of Employee Turnover Using Machine Learning Algorithms Original Scientific Paper. *International Journal of Electrical and Computer Engineering Systems*, 13(2), 135–144.
- Burzykowski, T., Geubbelmans, M., Rousseau, A., & Valkenborg, D. (2023). Validation of machine learning algorithms. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(2), 295–297. <https://doi.org/10.1016/j.ajodo.2023.05.007>
- Chakraborty, R., Mridha, K., Shaw, R. N., & 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>
- Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. *Expert Systems with Applications*, 215, 119364. <https://doi.org/https://doi.org/10.1016/j.eswa.2022.119364>
- Guerranti, F., & Dimitri, G. M. (2022). A Comparison of Machine Learning Approaches for Predicting Employee Attrition. *Applied Sciences*, 13(1), 267. <https://doi.org/10.3390/app13010267>
- Hossen, M. A., Hossain, E., Khalib, Z. I. A., & Siddika, F. (2021). Ensemble method based architecture using random forest importance to predict employee 's turn over Ensemble method based architecture using random forest importance to predict employee 's turn over. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1755/1/012039>
- Kurniawan, S., Nugroho, A., & Suherman. (2025). Analisis Faktor yang Mempengaruhi Promosi Karyawan Menggunakan Random Forest pada Dataset Employee Promotion. *Jurnal Pustaka AI*, 5(2), 177–187. <https://doi.org/https://doi.org/10.55382/jurnalpustakaai.v5i2.115>
- Lazzari, M., Alvarez, J. M., & Ruggieri, S. (2022). Predicting and Explaining Employee Turnover Intention. *International Journal of Data Science and Analytics*, 279–292. <https://doi.org/10.1007/s41060-022-00329>
- Paigude, S., Pangarkar, S. C., Hundekari, S., Wanjale, K., & Dongre, Y. (2023). Potential of Artificial Intelligence in Boosting Employee Retention in the Human Resource Industry. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(3).
- Pourkhodabakhsh, N., Mamoudan, M. M., & Bozorgi-amiri, A. (2023). Effective Machine Learning, Meta-Heuristic Algorithms and Multi-Criteria Decision Making to Minimizing Human Resource Turnover. *Applied Intelligence*, 16309–16331. <https://doi.org/https://doi.org/10.1007/s10489-022-04294-6>
- Raza, A., Munir, K., Almutairi, M., Younas, F., & Muhammad, M. F. S. (2022). Predicting Employee Attrition Using Machine Learning Approaches. *Applied Sciences*. <https://doi.org/https://doi.org/10.3390/app12136424>
- Setiyadi, P., Prayogi, M. N., & Solichin, A. (2024). *Optimalisasi Prediksi Kehilangan Karyawan menggunakan*

Teknik RFE, SMOTE, dan ADABOOST. 9(4), 2131–2145.

Wardhani, F. H., & Lhaksana, K. M. (2022). Predicting Employee Attrition Using Logistic Regression with Feature Selection. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 6(4), 2214–2222. <https://doi.org/https://doi.org/10.33395/sinkron.v7i4.11783>