

# Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

Supported by:  
**Rakamin Academy**  
Career Acceleration School  
[www.rakamin.com](http://www.rakamin.com)



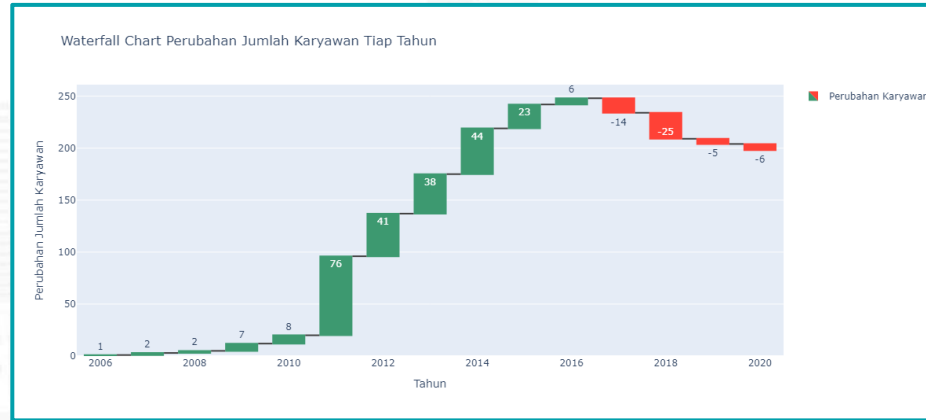
**Created by:**  
**Arieska Restu**  
[arieskarestu02@gmail.com](mailto:arieskarestu02@gmail.com)  
[linkedin.com/in/arieskarestu](https://linkedin.com/in/arieskarestu)

I am an Assistant Lecturer who has experience in the field of Data Science with a background in Informatics. Experienced in Data Analysis, Data Mining, and Machine Learning projects. Also experienced in extracting primary and secondary data, as well as developing and maintaining databases. Able to conduct in-depth data analysis to identify trends that are relevant to companies and clients, and proficient in creating analysis reports. I also have expertise in programming languages and data visualization.

“Sumber daya manusia (SDM) adalah aset utama yang perlu dikelola dengan baik oleh perusahaan agar tujuan bisnis dapat tercapai dengan efektif dan efisien. Pada kesempatan kali ini, kita akan menghadapi sebuah permasalahan tentang sumber daya manusia yang ada di perusahaan. Fokus kita adalah untuk mengetahui bagaimana cara menjaga karyawan agar tetap bertahan di perusahaan yang ada saat ini yang dapat mengakibatkan bengkaknya biaya untuk rekrutmen karyawan serta pelatihan untuk mereka yang baru masuk. Dengan mengetahui faktor utama yang menyebabkan karyawan tidak merasa, perusahaan dapat segera menanggulangnya dengan membuat program-program yang relevan dengan permasalahan karyawan.”

- Pada tahap data preprocessing, dilakukan beberapa proses yakni proses quick EDA, handling missing values, dan drop features.
- Untuk proses quick EDA dilakukan untuk melakukan pengecekan informasi dataset, kolom, statistical summaries, missing values, dan data duplicate.
- Dari proses quick EDA, didapatkan hasil bahwa terdapat missing value pada fitur SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, dan JumlahKetidakhadiran, serta fitur AlasanResign.
- Selain itu, dalam dataset tidak ada data yang duplicate dan terdapat fitur yang hanya memiliki satu unique value yaitu fitur PernahBekerja.

- Proses handling missing values dilakukan dengan beberapa cara sesuai dengan fiturnya.
- Missing value pada fitur SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, dan JumlahKetidakhadiran akan diisi menggunakan modus.
- Missing value pada fitur AlasanResign akan diisi nilainya menjadi masih\_bekerja karena pada fitur TanggalResign masih kosong (belum ada tanggalnya).
- Untuk proses drop features dilakukan untuk menghapus fitur-fitur yang tidak diperlukan dan fitur yang terlalu banyak missing valuenya yakni 'Username', 'EnterpriselD', 'PernahBekerja', dan 'IkutProgramLOP'.



Berdasarkan visualisasi ini, dapat disimpulkan bahwa kondisi perusahaan mengalami beberapa fase, yaitu sebagai berikut.

- Fase Kenaikan Tajam, pada tahun 2011-2012 terdapat kenaikan yang signifikan pada jumlah karyawan. Peningkatan ini mungkin disebabkan oleh ekspansi perusahaan atau kebutuhan operasional yang lebih tinggi, yang memerlukan banyak karyawan tambahan.
- Fase Periode Stabil, pada tahun 2013-2015, jumlah karyawan masih mengalami kenaikan, namun kenaikannya lebih stabil. Ini menunjukkan stabilitas perusahaan, di mana pertumbuhan karyawan masih terjaga meskipun tidak sebesar tahun-tahun sebelumnya.
- Fase Penurunan, pada tahun 2016-2020 terjadi penurunan jumlah karyawan yang cukup besar, terutama pada tahun 2017. Hal ini menunjukkan adanya permasalahan yang lebih serius, seperti penurunan produktivitas, kepuasan kerja, atau faktor eksternal yang memengaruhi. Walaupun pada tahun 2018-2020 penurunannya lebih kecil, tren ini tetap menunjukkan arah yang mengkhawatirkan.



Berdasarkan hasil visualisasi didapatkan insight-insight, yaitu sebagai berikut.

- Penurunan jumlah karyawan yang konsisten selama beberapa tahun menunjukkan bahwa perusahaan mungkin sedang menghadapi tantangan internal atau eksternal. Ini bisa berupa masalah dalam manajemen sumber daya manusia, lingkungan kerja yang kurang mendukung, atau tekanan dari kompetitor di pasar.
- Setelah melihat penurunan karyawan yang signifikan, terutama pada tahun 2017, perusahaan harus mengevaluasi ulang kebijakan retensi karyawan. Program peningkatan kepuasan kerja, penyesuaian remunerasi, atau peninjauan kembali budaya kerja mungkin perlu dipertimbangkan untuk mencegah kehilangan lebih banyak karyawan di masa mendatang.
- Pertumbuhan besar yang terjadi di tahun 2011-2012 kemungkinan berkaitan dengan ekspansi atau proyek besar yang dijalankan perusahaan pada saat itu. Namun, perusahaan harus berhati-hati dalam menjaga keseimbangan antara ekspansi dan mempertahankan sumber daya manusia yang ada.
- Dengan tren penurunan karyawan sejak tahun 2016, kondisi perusahaan dapat dikatakan mengkhawatirkan. Hal ini bisa mempengaruhi stabilitas operasional perusahaan dalam jangka panjang. Manajemen perlu fokus pada pemulihan jumlah karyawan serta memperbaiki kondisi internal perusahaan.

Sunburst Chart: Resign Berdasarkan Jenjang Karir, Performa Karyawan, dan Alasan Resign di Divisi Data Analyst



Berdasarkan hasil visualisasi tersebut, insight-insight yang bisa didapatkan yaitu sebagai berikut.

- Mayoritas pegawai di divisi Data Analyst yang resign berasal dari jenjang karir Sangat Bagus dan Bagus, sedangkan yang memiliki jenjang karir Biasa juga menyumbang jumlah resign yang signifikan. Sangat Bagus mungkin adalah karyawan senior yang sudah lama bekerja dan bisa jadi ada alasan khusus mengapa mereka resign.
- Karyawan dengan performa Bagus dan Sangat Bagus justru banyak yang resign. Ini bisa menunjukkan bahwa masalahnya bukan terletak pada performa karyawan, melainkan pada faktor lain di luar kemampuan kerja.
- Toxic Culture dan Internal Conflict menjadi alasan dominan di setiap level performa, khususnya di tingkat Bagus dan Biasa. Ini menandakan bahwa lingkungan kerja yang kurang kondusif menjadi masalah besar di divisi ini.

Berdasarkan hasil visualisasi dan insight tersebut, berikut merupakan rekomendasi-rekomendasi yang dapat digunakan.

- Karena toxic culture menjadi alasan resign yang signifikan, perusahaan harus segera melakukan evaluasi mendalam mengenai budaya kerja di divisi ini. Ini bisa mencakup pelatihan bagi manajer untuk menciptakan lingkungan kerja yang lebih inklusif dan mengurangi friksi antara tim.
- Internal Conflict juga menjadi alasan utama. Ini menunjukkan adanya masalah komunikasi atau konflik interpersonal yang belum terselesaikan. Perusahaan dapat melakukan pelatihan manajemen konflik dan membentuk tim HR yang lebih aktif dalam menangani permasalahan karyawan.
- Karena banyak karyawan dengan performa Sangat Bagus dan Bagus yang resign, perusahaan perlu mempertimbangkan langkah-langkah retensi khusus seperti program loyalitas karyawan atau peningkatan reward bagi mereka yang berprestasi untuk memastikan mereka tetap bertahan di perusahaan.
- Resign dari karyawan di jenjang karir Sangat Bagus bisa jadi menandakan bahwa peluang pengembangan karir lebih lanjut kurang tersedia. Memberikan lebih banyak kesempatan promosi atau rotasi kerja bisa membantu mengurangi angka resign dari karyawan senior.

Dengan langkah-langkah tersebut, diharapkan perusahaan dapat mengurangi jumlah resign, terutama di divisi Data Analyst dan menjaga karyawan yang memiliki performa dan kontribusi tinggi.



- Sebelum melakukan proses feature encoding dan sebagainya, dilakukan proses handling outlier terlebih dahulu.
- Dimana pada proses handling outlier, mengganti nilai outliernya dengan batas maksimum.
- Untuk proses feature encoding, dilakukan label encoding untuk fitur StatusPernikahan, JenisKelamin, StatusKepegawaian, Pekerjaan, JenjangKarir, PerformancePegawai, AsalDaerah, HiringPlatform, TingkatPendidikan, AlasanResign.
- Lalu untuk proses feature engineering, dilakukan pembuatan fitur baru seperti LamaBekerja dan Resign.
- Kemudian melakukan feature selection, untuk memilih fitur-fitur yang akan digunakan dalam proses pelatihan model machine learning. Fitur-fitur tersebut yakni Pekerjaan, JenjangKarir, PerformancePegawai, LamaBekerja, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, JumlahKetidakhadiran, SkorKepuasanPegawai, StatusPernikahan.

- Selanjutnya adalah melakukan proses pemisahan untuk feature dan target. Dimana target yang dipakai adalah Resign yang bernilai 1 dan 0.
- Setelah itu, melakukan proses standarisasi untuk setiap feature numerical.
- Lalu melakukan proses split data untuk data training dan testing, dimana perbandingannya yaitu sebesar 80% : 20%.
- Kemudian melakukan proses handling data imbalanced dengan menggunakan SMOTE.
- Lalu melakukan proses modeling, dimana algoritma machine learning yang digunakan yakni Logistic Regression, Random Forest, Decision Tree, dan K-Nearest Neighbors.

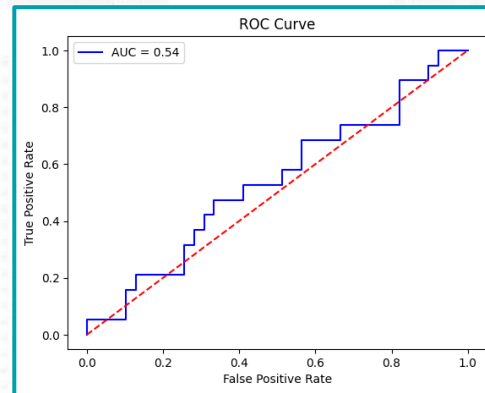
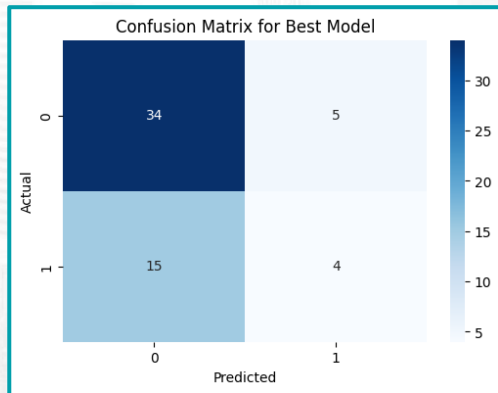
- Berikut merupakan tabel perbandingan dari hasil training untuk setiap model machine learning.

Tabel Perbandingan Hasil Training

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.482759	0.238095	0.263158	0.250000
Random Forest	0.637931	0.375000	0.157895	0.222222
Decision Tree	0.517241	0.235294	0.210526	0.222222
K-Nearest Neighbors	0.517241	0.333333	0.473684	0.391304

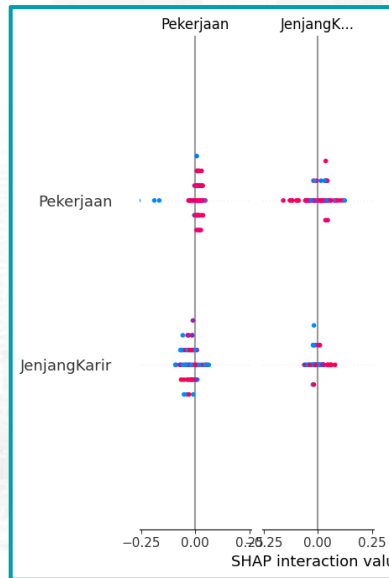
- Berdasarkan tabel tersebut, didapatkan bahwa model yang terbaik yakni Random Forest.
- Model terbaik ini akan dilakukan hyperparameter tuning dengan menggunakan GridSearchCV.

- Berikut merupakan confusion matrix dan ROC plot dari hasil hyperparameter tuning.



- Berdasarkan hasil evaluasi model ini cukup baik dalam melakukan prediksi karyawan yang tidak resign, tetapi lemah dalam mendeteksi karyawan yang resign.

- Berikut merupakan visualisasi dari SHAP summary.



- Berdasarkan SHAP summary plot, variabel seperti JenjangKarir dan Pekerjaan adalah yang paling berpengaruh. Hal ini bisa menjadi fokus HR untuk meningkatkan retensi karyawan.



A faded, light-colored background image of a city skyline with various skyscrapers and buildings.

# THANK YOU

Have a nice day!