

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning



Created by:

Citta Mudita

citmud10@gmail.com

<https://www.linkedin.com/in/citta-mudita-40b91b22b/>

I have a strong interest in data analysis and data science. My experience in digital marketing has made me an expert in conducting exploratory data analysis and creating data visualizations to address business challenges. I have received awards as the Top 2 Student in the Data Science Bootcamp and as the Most Outstanding Student for the Final Project. Additionally, my team was recognized as the Best Final Project Team at Data Science Bootcamp.

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

“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.”

Handling Missing Value



- IkutProgramLOP sebesar 89,9% dipertimbangkan untuk di hapus karena banyak sekali memiliki nilai null
- AlasanResign sebesar 23%, mengisi dengan nilai "masihbekerja" karena yang tidak mengisi ini, adalah orang yang tidak memiliki tanggal resign/masih bekerja
- JumlahKetidakhadiran sebesar 2,09% mengisi dengan nilai median
- SkorKepuasanPegawai sebesar 1,74% mengisi dengan nilai median
- JumlahKeikutsertaanProjek sebesar 1,05% mengisi dengan nilai median
- JumlahKeterlambatanSebulanTerakhir sebesar 0,35% mengisi dengan nilai median

Handling Summary Aneh



- Terdapat nilai Yes di feature kolom Pernah Bekerja, sehingga perlu di ganti nilai tersebut dengan nilai 1 yang merepresentasikan Yes.

Dropped Data



- Membuang kolom yang memiliki satu unique value (konstanta) yaitu Pernah Bekerja

Source Code

```
[7]: #Menghapus Kolom IkutProgramLOP
df.drop(['IkutProgramLOP'], axis=1, inplace=True)

[8]: #Alasan Resign
fill_value_alasan_resign = 'masihbekerja'
df['AlasanResign'].fillna(fill_value_alasan_resign, inplace=True)

[9]: #Jumlah Ketidakhadiran
median_ketidakhadiran = df['JumlahKetidakhadiran'].median()
df['JumlahKetidakhadiran'].fillna(median_ketidakhadiran, inplace=True)

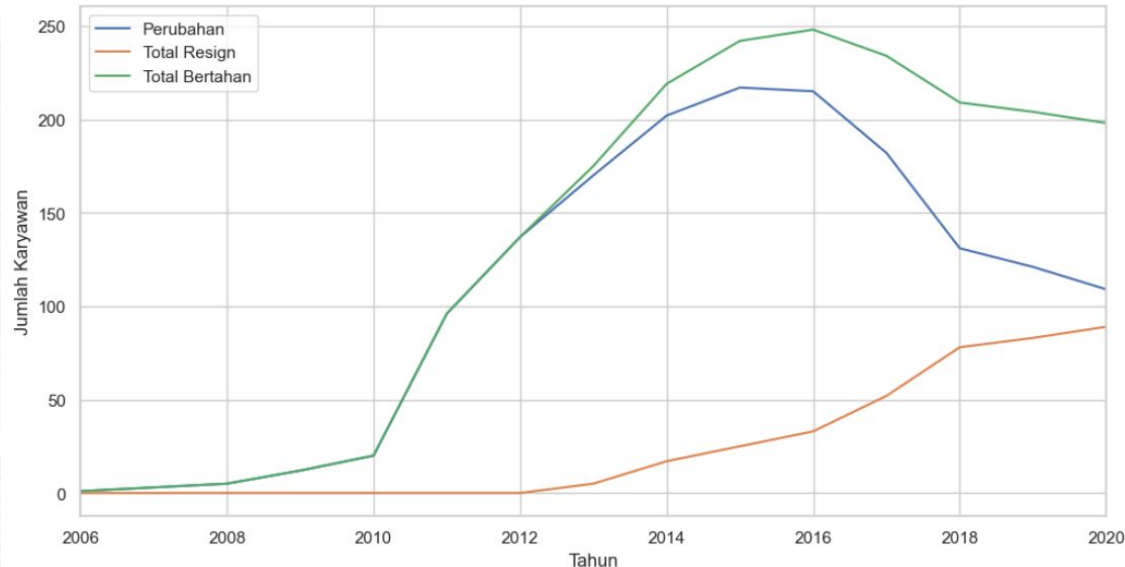
[10]: #SkorKepuasanPegawai
median_kepuasan_pegawai = df['SkorKepuasanPegawai'].median()
df['SkorKepuasanPegawai'].fillna(median_kepuasan_pegawai, inplace=True)

[11]: #JumlahKeikutsertaanProyek
median_keikutsertaan_proyek = df['JumlahKeikutsertaanProyek'].median()
df['JumlahKeikutsertaanProyek'].fillna(median_keikutsertaan_proyek, inplace=True)

[12]: #JumlahKeterlambatanSebulanTerakhir
median_keterlambatan = df['JumlahKeterlambatanSebulanTerakhir'].median()
df['JumlahKeterlambatanSebulanTerakhir'].fillna(median_keterlambatan, inplace=True)
```


Perubahan Kenaikan dan Penurunan Jumlah Karyawan

Dimulai tahun 2017 perusahaan mengalami penurunan karyawan yang signifikan, yang mana lebih besar daripada penambahan karyawan, yang menandakan perusahaan sedang tidak baik-baik saja

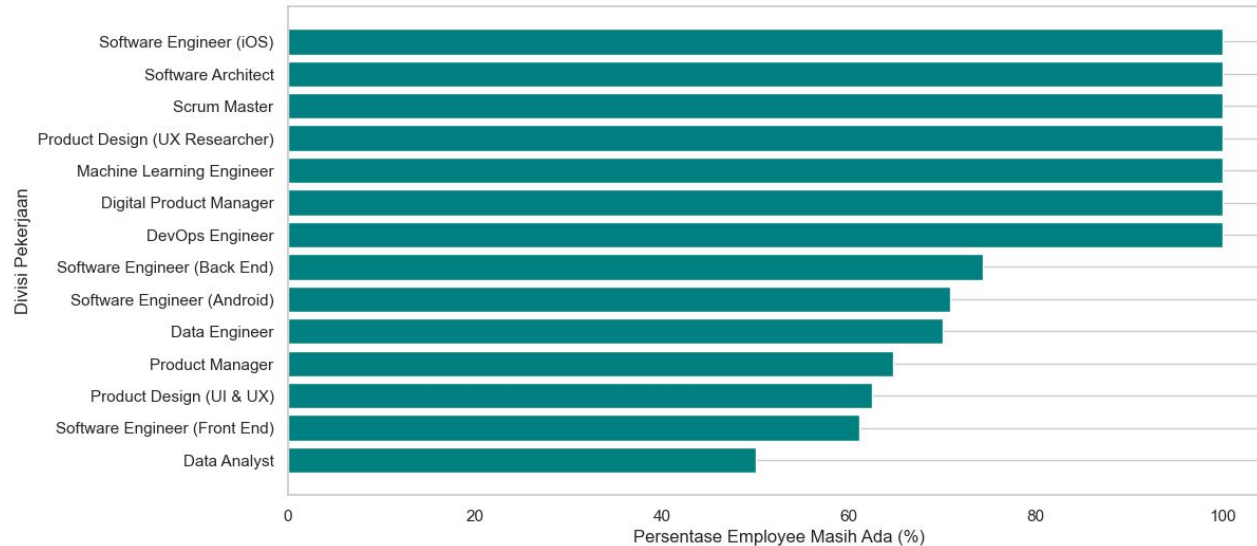


Pada tahun 2017, perusahaan mengalami penurunan karyawan yang signifikan. Karyawan yang resign jauh lebih tinggi daripada karyawan yang masuk. Hal ini menunjukkan bahwa perusahaan mengalami fluktuasi karyawan yang besar pada tahun ini, dan penurunan menjadi tanda bahwa perusahaan sedang menghadapi tantangan atau masalah yang perlu segera diatasi.

Pada tahun 2018 hingga 2020, terlihat bahwa tidak ada penambahan karyawan yang signifikan, sementara jumlah karyawan yang resign tetap tinggi. Ini mengindikasikan bahwa perusahaan belum mampu membalikkan tren negatif yang dimulai pada tahun 2017. Akibat hal tersebut, banyak proyek-proyek yang tidak selesai, yang dapat mempengaruhi produktivitas dan efisiensi perusahaan.

Persentase Karyawan yang Masih Bekerja Berdasarkan Divisi Pekerjaan

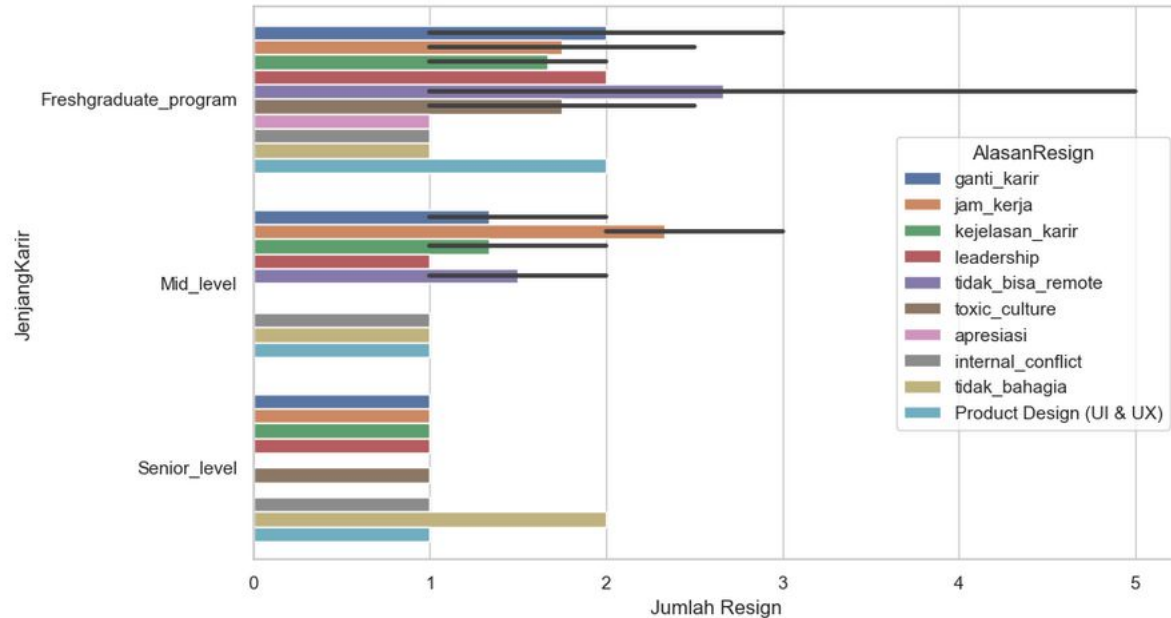
Divisi yang tidak memiliki karyawan yang telah mengundurkan diri termasuk Software Engineer, Software Architect, Scrum Master, Product Design, Machine Learning Engineer, Digital Product Manager, dan DevOps. Sementara itu, Data Analyst adalah divisi dengan tingkat pengunduran diri karyawan sebesar 50%.



Grafik tersebut memberikan gambaran tentang persentase karyawan yang masih bekerja berdasarkan divisi pekerjaan. Terdapat 7 Divisi tidak memiliki pengunduran diri, sementara Data Analyst memiliki tingkat pengunduran diri sebesar 50%, yang perlu di analisis lebih lanjut terkait alasan apa yang menyebabkan karyawan mengundurkan diri.

Jumlah Karyawan Resign berdasarkan Jenjang Karir dan Alasan Resign

Fresh graduate cenderung mengundurkan diri karena keterbatasan pekerjaan yang tidak dapat dilakukan secara remote. Di tingkat mid level, alasan pengunduran diri lebih sering terkait dengan jam kerja yang sangat sibuk. Sementara di level senior, banyak yang memutuskan untuk pergi karena kurangnya kebahagiaan di tempat kerja.



Rekomendasi :

Fresh Graduate

disarankan untuk memberikan kesempatan untuk pekerjaan jarak jauh kepada karyawan yang memungkinkan, terutama bagi mereka yang bekerja di bidang IT seperti software engineer, yang dapat bekerja dari berbagai lokasi.

Middle level

Alasan pengunduran diri berkaitan dengan jam kerja yang terlalu padat, perusahaan bisa mempertimbangkan untuk menambah jumlah karyawan di tim atau departemen terkait. Dengan cara ini, beban kerja dapat lebih merata dan mengurangi tekanan individual.

Senior level

perusahaan dapat fokus pada fasilitas dan program kesejahteraan karyawan, termasuk program kesehatan, kegiatan sosial, pengembangan diri, dan keseimbangan kerja-kehidupan pribadi. Hal ini akan membantu menciptakan lingkungan kerja yang lebih seimbang dan menyenangkan.

Build an Automated Resignation Behavior Prediction using Machine Learning

Handle Missing Value

No missing value

Handle Duplicate Value

No Duplicate Data

Feature Extraction

Membuat feature baru resign dari tanggal resign

Feature Encoding

Label Encoding

JenjangKarir
PerformancePegawai
TingkatPendidikan
SkorKepuasanPegawai

One Hot Encoding

StatusKepegawaian
StatusPernikahan
Pekerjaan
AlasanResign

Feature Selection

Drop feature
Username, JenisKelamin,
'EnterpriseID', 'NomorHP', 'Email',
AsalDaerah, TanggalLahir,
'HiringPlatform', TanggalHiring,
TanggalPenilaianKaryawan,
AlasanResign_masih_bekerja

Modeling

Data Train

70%

- Handling Imbalance Data (handling imbalance data using SMOTE)

Split Data Train and Data Test

Data Test

30%

Handle Outlier

Tidak melakukan handle outlier
karna menggunakan model
yang robust terhadap outlier

■ Model Comparison

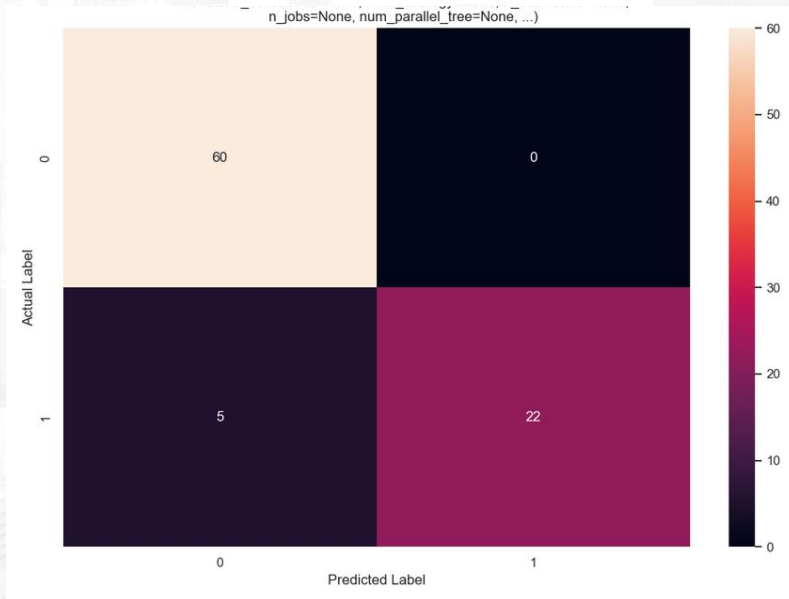
Model	Recall (Train)	Recall (Test)	Accuracy (Train)	Accuracy (Test)	ROC AUC (Train)	ROC AUC (Test)
Logistic Regression	0.822581	0.629630	0.855	0.678161	0.913628	0.770370
AdaBoost	0.854839	0.962963	0.955	0.988506	0.938990	0.986420
Random Forest	0.983871	0.925926	0.995	0.977011	1.000000	1.000000
XGBoost	0.806452	0.814815	0.940	0.942529	0.938172	0.980864

Model yang dipilih adalah XGBoost karena memiliki *accuracy* yang cukup memadai dan nilai *recall* yang tertinggi diantara semua model

Untuk XGBoost digunakan Hyperparameter sebagai berikut : eta = 0.05, max_depth = 1, scale_pos_weight = 1.3

Hyperparameter nthread, tree_method dan predictor dipilih sedemikian rupa agar komputasi dan fitting model memakan waktu yang sesedikit mungkin. Nilai untuk Hyperparameter eta dan scale_pos_weight dipilih karena menghasilkan nilai *recall* tertinggi tanpa mengorbankan nilai *accuracy* begitu besar, serta model tidak mengalami overfitting dengan menggunakan Hyperparameter tersebut

Confusion Matrix

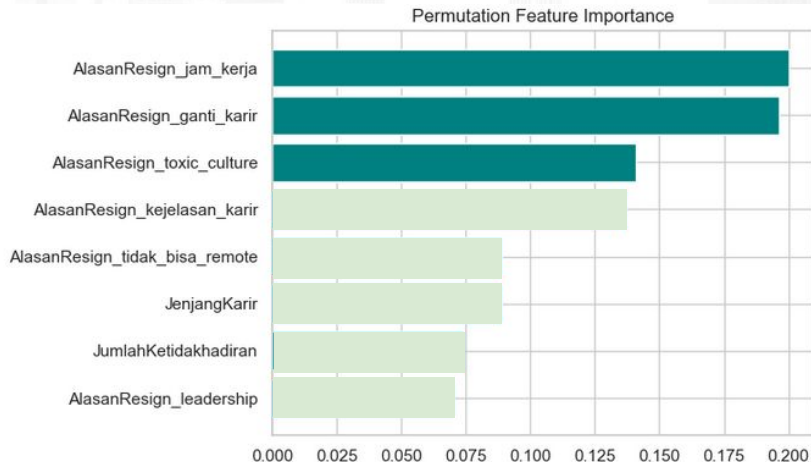


Dengan metric utama pada modeling adalah recall, dimana berfungsi untuk mendeteksi sebanyak-banyaknya karyawan yang resign dan benar-benar resign (TP).

Dari confusion matrix terlihat jumlah karyawan yang resign dan benar-benar resign (TP) sebanyak 22, sedangkan karyawan yang tidak resign sebanyak (TN) 60.

Dan karyawan yang terdeteksi tidak resign tetapi pada kenyataannya resign adalah 5 karyawan (FN) dan sedangkan karyawan yang sebenarnya tidak resign tetapi terdeteksi resign (FP) sebanyak 0 karyawan.

Feature Importance



Feature yang paling berpengaruh adalah feature AlasanResign, dengan top 3 alasan resign yang paling berpengaruh adalah

- Jam Kerja
- Ganti Karir
- Toxic Culture

Alasan Resign ini seperti pada tahap EDA, didominasi oleh fresgraduate program dan mid level.

Recommendation Business

Alasan Resign Jam Kerja

Selidiki penyebab, pantau dan evaluasi beban kerja karyawan supaya jam kerja yang terlalu padat tidak menjadi hambatan bagi karyawan menjaga keseimbangan hidup. Pertimbangkan untuk memberikan opsi kerja fleksibel atau kebijakan kerja dari jarak jauh. Ini dapat membantu mengurangi tekanan yang mungkin disebabkan oleh jam kerja yang tidak fleksibel.

Penelitian menunjukkan beban kerja yang terlalu padat berdampak negatif. Selidiki penyebab dan berikan opsi kerja fleksibel." - Dr. Jane Doe.

Alasan Resign Ganti Karir

Sediakan jalur karir yang terdefinisi dengan baik dan jelas agar pegawai memiliki visi dan motivasi untuk tetap berkembang di perusahaan. Ini akan membantu mempertahankan karyawan di semua tingkatan, termasuk Middle dan Senior.

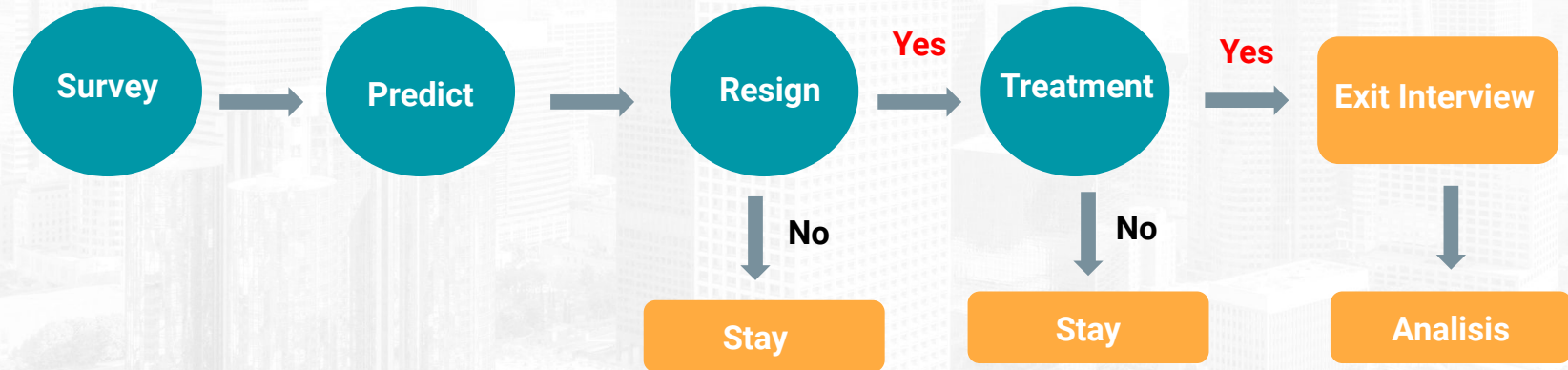
Jalur karir yang jelas adalah motivasi utama. Penelitian menegaskan perusahaan dengan jalur karir yang jelas mempertahankan karyawan." - Prof. John Smith.

Alasan Resign Toxic Culture

Selidiki penyebab toksisitas dalam budaya perusahaan. Ini bisa termasuk evaluasi hubungan antara atasan dan bawahan, tingkat stres di lingkungan kerja, atau masalah komunikasi yang perlu diatasi. Fokus pada pengembangan keterampilan kepemimpinan manajer untuk menciptakan lingkungan kerja yang positif, di mana karyawan merasa dihargai dan didukung. Selalu berikan saluran komunikasi yang terbuka agar karyawan dapat melaporkan masalah dan memberikan masukan untuk perbaikan budaya perusahaan.

Penelitian psikologi organisasi menunjukkan budaya toksik merugikan kesejahteraan karyawan. Selidiki dan atasi penyebabnya." - Dr. Emily Johnson.

Recommendation Software Development



True Positive (TP) = 22
False Negative (FN) = 5
Total = 87

Assumption 80%
resign rate

$$(0.8(TP+FN))/TOTAL$$

89

Before Model



67

After Model

A Decrease of 25%



22

Resign Rate

31%

89

Before Model



Resign Rate

8%

22

Decrease



Resign Rate

23%

67

After Model