

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning



Created by:

Muhamad Zamzam Istimaqom

zamzamistimaqom@gmail.com

<https://www.linkedin.com/in/zamzamistimaqom>

Hi. I am Zamzam Istimaqom,

This report is eda, visualization, and extract insight from user behaviour data

“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.”

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 287 entries, 0 to 286
```

```
Data columns (total 25 columns):
```

#	Column	Non-Null Count	Dtype
0	Username	287 non-null	object
1	EnterpriseID	287 non-null	int64
2	StatusPernikahan	287 non-null	object
3	JenisKelamin	287 non-null	object
4	StatusKepegawaian	287 non-null	object
5	Pekerjaan	287 non-null	object
6	JenjangKarir	287 non-null	object
7	PerformancePegawai	287 non-null	object
8	AsalDaerah	287 non-null	object
9	HiringPlatform	287 non-null	object
10	SkorSurveyEngagement	287 non-null	int64
11	SkorKepuasanPegawai	282 non-null	float64
12	JumlahKeikutsertaanProjek	284 non-null	float64
13	JumlahKeterlambatanSebulanTerakhir	286 non-null	float64
14	JumlahKetidakhadiran	281 non-null	float64
15	NomorHP	287 non-null	object
16	Email	287 non-null	object
17	TingkatPendidikan	287 non-null	object
18	PernahBekerja	287 non-null	object
19	IkutProgramLOP	29 non-null	float64
20	AlasanResign	221 non-null	object
21	TanggalLahir	287 non-null	object
22	TanggalHiring	287 non-null	object
23	TanggalPenilaianKaryawan	287 non-null	object
24	TanggalResign	287 non-null	object

```
dtypes: float64(5), int64(2), object(18)
```

```
memory usage: 56.2+ KB
```

- Data terdiri dari 287 baris
- Terdapat beberapa missing value pada kolom Skor Kepuasan Pegawai, Jumlah Keikutsertaan Project, Jumlah Keterlambatan Sebulan Terakhir, Ikut Program LOP, dan alasan Resign

Skor Kepuasan Pegawai,
Jumlah Keikutsertaan Project,
Jumlah keterlambatan ,
Jumlah Ketidakhadiran :

Handling :
Fillna.median()

Alasan :
Data numerik dan lebih robust
terhadap outlier

Ikut Program LOP

Handling :
Drop()

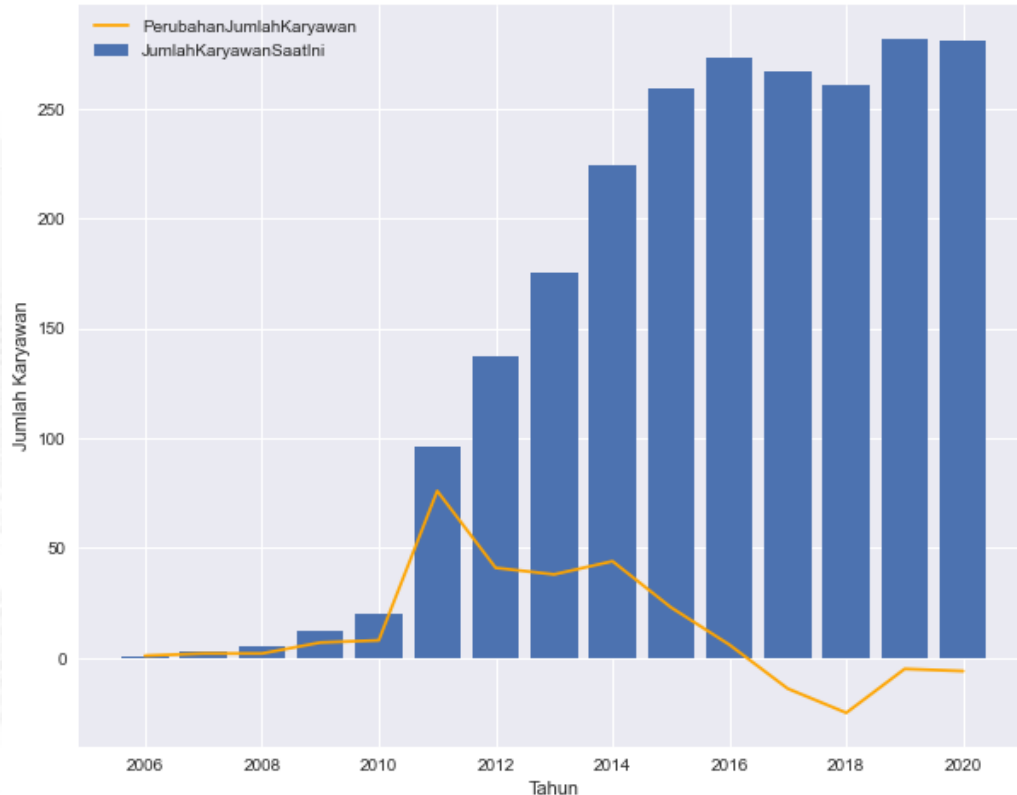
Alasan :
Data missing value lebih
besar daripada 50%

Alasan Resign

Handling :
Drop()

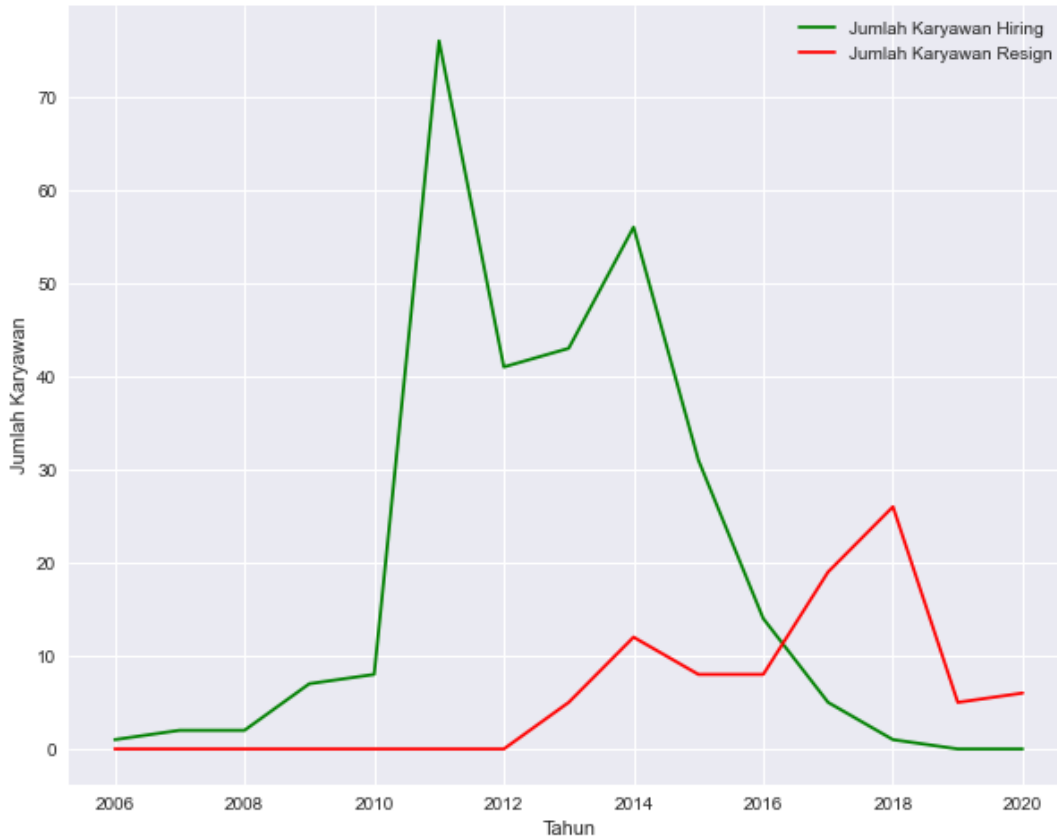
Alasan :
Data missing value lebih
besar daripada 50%

Perbandingan jumlah existing karyawan dan jumlah karyawan resign setiap tahun



Pada tahun 2016, terjadi perubahan jumlah karyawan adalah dibawah 0, artinya jumlah karyawan yang resign lebih banyak daripada karyawan yang masuk. Hal ini terus berlangsung selama 4 tahun terakhir dan akan menyebabkan kondisi perusahaan menjadi tidak sehat

Perbandingan Jumlah Hiring dan Jumlah Resign setiap tahun



Jumlah karyawan mengalami penurunan pertama kali pada tahun 2016.

Hal ini akan berdampak pada performance perusahaan karena ada beberapa posisi yang tidak akan terisi atau kosong. Sehingga diperlukan pendalaman untuk mengetahui alasan apa saja karyawan resign sehingga perusahaan dapat berbenah. Kondisi ini cukup mengkhawatirkan mengingat jumlah karyawan resign lebih tinggi daripada karyawan hirinng(terjadi ketidakseimbangan) sehingga membuat perusahaan akan tidak sehat.

Resign Reason Analysis for Employee Attrition Management Strategy

	Pekerjaan	JumlahActive	JumlahResign	JumlahKaryawan	PresentaseActive
0	Data Analyst	8	8.0	16.0	50.000000
1	Data Engineer	7	3.0	10.0	70.000000
2	DevOps Engineer	3	0.0	3.0	100.000000
3	Digital Product Manager	2	0.0	2.0	100.000000
4	Machine Learning Engineer	2	0.0	2.0	100.000000
5	Product Design (UI & UX)	15	9.0	24.0	62.500000
6	Product Design (UX Researcher)	1	0.0	1.0	100.000000
7	Product Manager	11	6.0	17.0	64.705882
8	Scrum Master	3	0.0	3.0	100.000000
9	Software Architect	1	0.0	1.0	100.000000
10	Software Engineer (Android)	17	7.0	24.0	70.833333
11	Software Engineer (Back End)	81	28.0	109.0	74.311927
12	Software Engineer (Front End)	44	28.0	72.0	61.111111
13	Software Engineer (iOS)	3	0.0	3.0	100.000000

Interpretasi :

Jenis pekerjaan yang memiliki resign tertinggi adalah pekerjaan Software Engineer yang memiliki jenjang karir Fresh Graduate. Kebanyakan alasan adalah alasan yang berkaitan pengembangan perusahaan dan culture perusahaan seperti kejelasan karir, toxic culture, dan jam kerja.

Rekomendasi :

Perusahaan bisa mengakomodir alasan-alasan dengan sesekali mengadakan kegiatan yang sifatnya pengembangan diri, sistem kejelasan karir, dan handle toxic culture pada karyawan sehingga alasan-alasan pribadi semacam ini tidak muncul di kemudian hari.

Label Encoding :
Jenjang Karir
Performance Pegawai
PernahBekerja,
TingkatPendidikan

Label Encoding :
['StatusPernikahan',
'StatusKepegawaian', 'Pekerjaan',
'JenjangKarir',
'PerformancePegawai',
'SkorSurveyEngagement',
'SkorKepuasanPegawai',
'JumlahKeikutsertaanProjek',
'JumlahKeterlambatanSebulanTerakh
ir',
'JumlahKetidakhadiran',
'TingkatPendidikan', 'AlasanResign',
'TanggalLahir', 'TanggalHiring',
'TanggalResign', 'TahunHiring',
'TahunResign', 'is_resign']

Feature Engineering :
Tahun Resign, Tahun
Hiring, Usia Hired,
Lama Bekerja

Data Split :

Data Training :

80 %

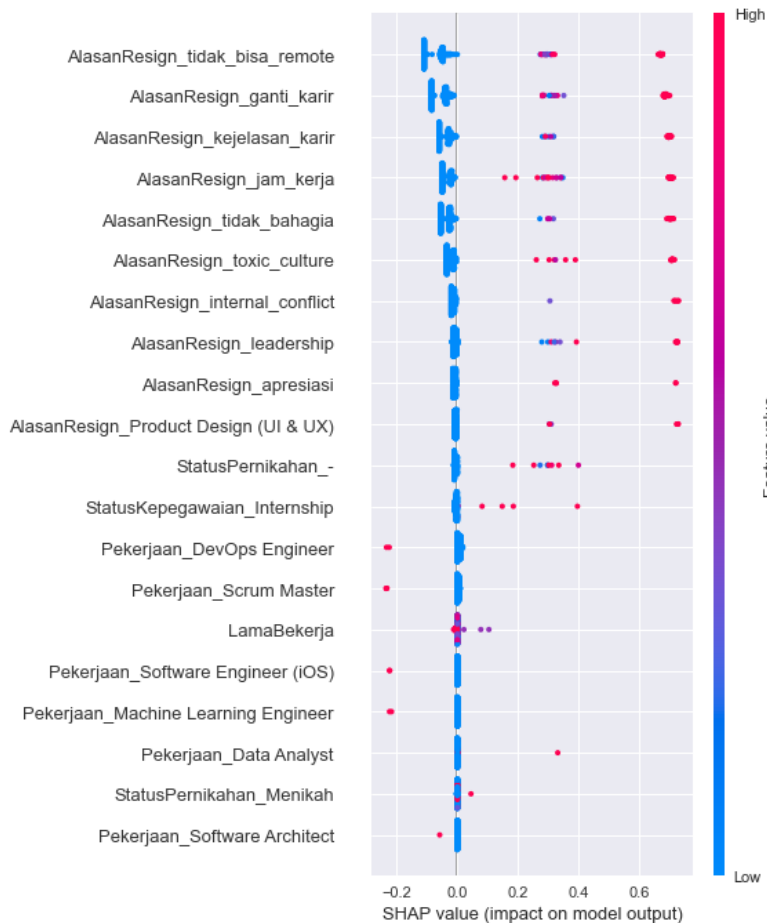
Data Test:

20 %

Imbalance Data : SMOTE

COMPARASION MODEL

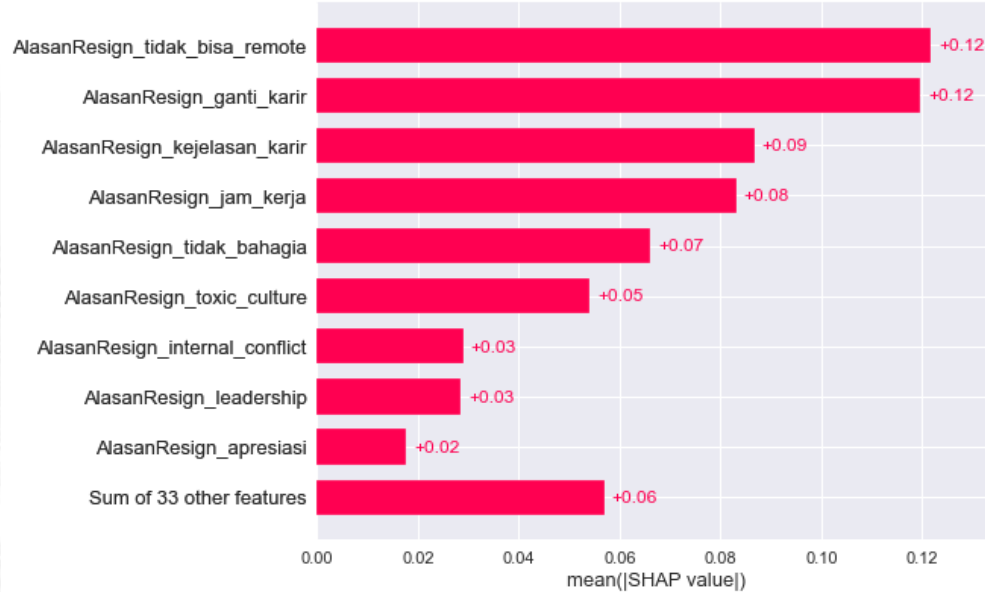
MODEL	PRECISION	RECALL	F1 SCORE	Running Time
Logistic Regression	1	1	1	0.00327
KNN	0.9	0.95	0.93	0.03109
SVM	1	1	1	0.01113
Naïve Bayes	1	1	1	0.00398
Random Forest	1	1	1	0.01849



Para karyawan melakukan resign karena alasan tidak bisa remote, ganti karir, kejelasan karir dan alasan jam kerja berpengaruh cukup besar terhadap keputusan karyawan untuk melakukan resign.

Secara status pekerjaan, internship juga mempengaruhi karyawan melakukan resign dibanding karyawan yang bersatus fulltime dan outsource.

Presenting Machine Learning Products to the Business Users



- Alasan karyawan melakukan Resign secara umum berkaitan dengan ketentuan perusahaan yang terlalu kaku di era saat ini yaitu tidak bisa remote, terikat jam kerja, dan kejelasan karir. Secara status pekerjaan, yang terjadi adalah karyawan fresh Graduate lebih besar melakukan resign. Kemungkinan Fresh Graduate adalah Gen Z yang memiliki pandangan berbeda dengan Pekerja Fulltime yang rata rata umurnya diatas 30 tahhun.

Rekomendasinya : Perusahaan bisa membuat ketentuan untuk membuat kebijakan pekerjaan dapat dilakukan secara remote dan membuat kejelasan jenjang karir yang jelas kepada karyawan