

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

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



Created by:
Muhammad Rizdky Maulady
rizdkymaulady@gmail.com
[linkedin.com/in/mrizdkymaulady](https://www.linkedin.com/in/mrizdkymaulady)

I have 3+ years of experience in the retail industry, specializing in digital marketing, business analytics, and sales operations leveraging data to drive strategic decisions. Over time, I discovered my passion for Data Science.

I love extracting insights, optimizing workflows, and helping businesses make smarter, data-driven decisions. With my strong background in sales, marketing, and analytics, I'm excited to take on new challenges in the data world. Let's connect and build something awesome with data.

“Human resources (HR) are a key asset that must be managed well by companies to achieve business goals effectively and efficiently. In this instance, we will address an issue regarding the human resources within the company. Our focus is to understand how to retain employees in the current company, as this can lead to increased costs for recruiting and training new hires. By identifying the main factors causing employee dissatisfaction, the company can promptly address these issues by creating relevant programs to tackle employee concerns. “

Check info dataset

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 287 entries, 0 to 286
Data columns (total 26 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  SkorKepuasanPegawai                     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                           286 non-null    float64
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
25  Unnamed: 25                             0 non-null     float64
dtypes: float64(7), int64(2), object(17)
memory usage: 58.4+ KB
```

Check Duplicate

```
Jumlah baris duplikat: 0
Tidak ada baris duplikat.
```

Handling Missing Value

Missing values status: True

	Total Null Values	Percentage	Data Type
Unnamed: 25	287	100.000000	float64
IkutProgramLOP	258	89.895470	object
AlasanResign	66	22.996516	object
JumlahKetidakhadiran	6	2.090592	object
SkorKepuasanPegawai	5	1.742160	object
JumlahKeikutsertaanProjek	3	1.045296	object
PernahBekerja	1	0.348432	float64
JumlahKeterlambatanSebulanTerakhir	1	0.348432	float64
Username	0	0.000000	object
EnterpriseID	0	0.000000	int64

- **Unnamed: 25 (100% empty)** - Remove the column.
- **IkutProgramLOP (89.89%)** - Fill with '0'.
- **AlasanResign (22.99%)** - Fill with 'masih_bekerja'.
- **JumlahKetidakhadiran (2.09%)** - Fill with 0.
- **SkorKepuasanPegawai (1.74%)** - Fill with the median.
- **JumlahKeikutsertaanProjek (1.04%)** - Fill with 0.
- **PernahBekerja (0.34%)** - Check the unique values.
- **JumlahKeterlambatan Sebulan Terakhir (0.34%)** - Check the unique values.

Nilai unik pada kolom PernahBekerja: [1. nan]

```
Jumlah setiap nilai unik pada kolom PernahBekerja:
PernahBekerja
1.0    286
Name: count, dtype: int64
```

PernahBekerja - fill with mean

Remove data only 1 unique value

```
Kolom dengan satu unique value: ['PernahBekerja']
```

PernahBekerja - drop column

Handling Invalid Value

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 287 entries, 0 to 286
Data columns (total 26 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                           286 non-null    float64
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
25  Unnamed: 25                             0 non-null      float64
dtypes: float64(7), int64(2), object(17)
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 287 entries, 0 to 286
Data columns (total 24 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                    287 non-null    int64
12  JumlahKeikutsertaanProjek              287 non-null    int64
13  JumlahKeterlambatanSebulanTerakhir      287 non-null    int64
14  JumlahKetidakhadiran                   287 non-null    int64
15  NomorHP                                 287 non-null    object
16  Email                                   287 non-null    object
17  TingkatPendidikan                      287 non-null    object
18  IkutProgramLOP                          287 non-null    int64
19  AlasanResign                            287 non-null    object
20  TanggalLahir                           287 non-null    datetime64[ns]
21  TanggalHiring                           287 non-null    datetime64[ns]
22  TanggalPenilaianKaryawan                287 non-null    datetime64[ns]
23  TanggalResign                           89 non-null     datetime64[ns]
dtypes: datetime64[ns](4), int64(7), object(13)
```

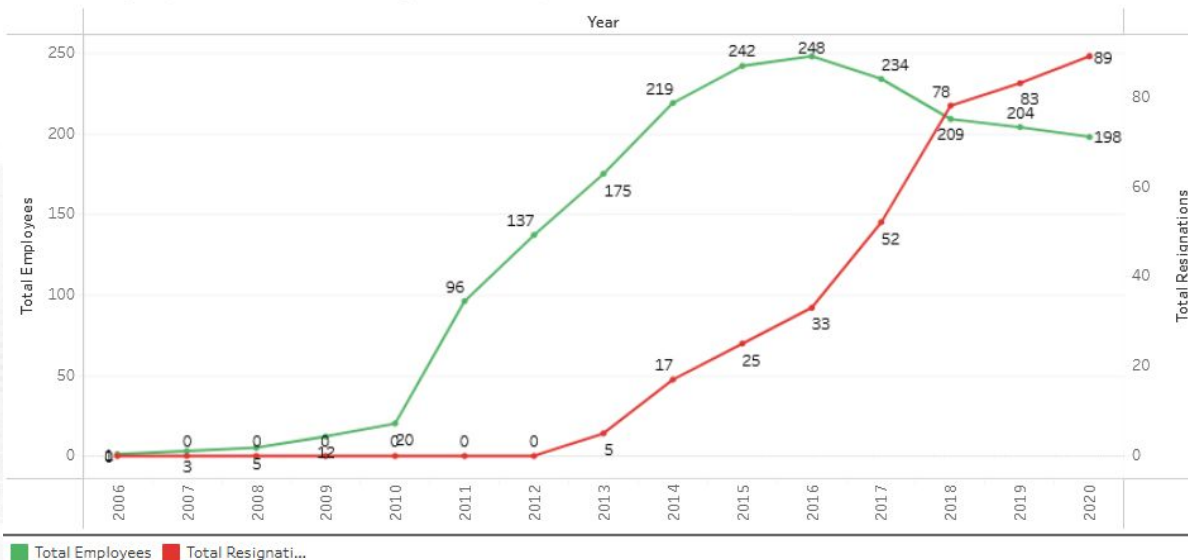
Columns containing dates must be converted to the datetime format.

Columns containing numerical data should be converted to the integer format.

Annual Report on Employee Number Changes

Tableau Dashboard

Total Employees vs. Total Resignations by Year

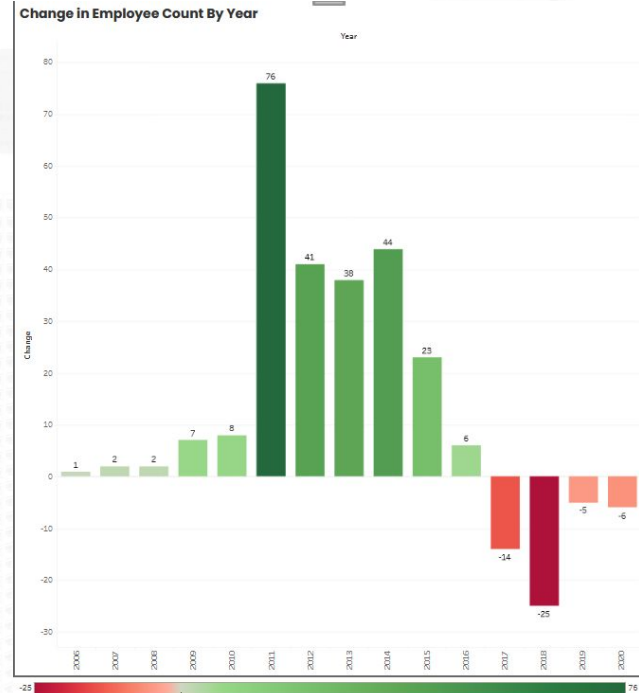
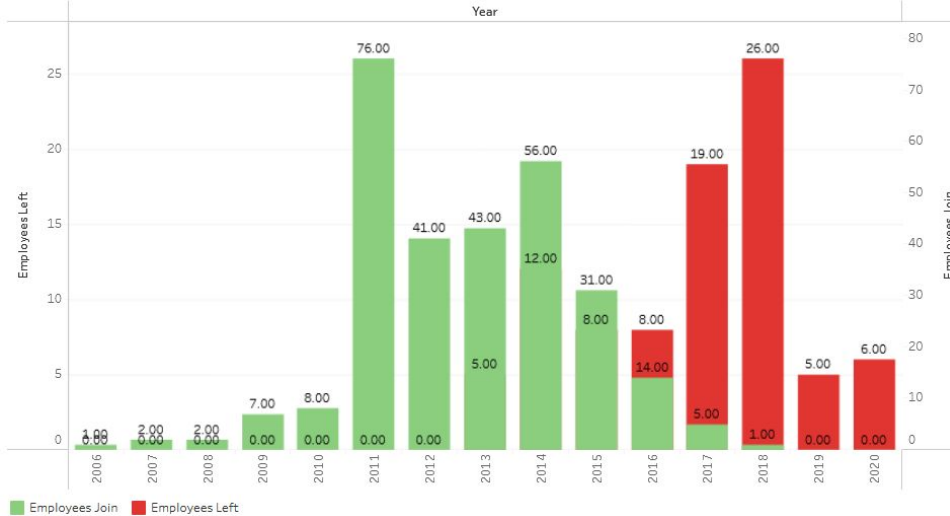


The company's condition has become concerning since 2017, particularly due to:

- **A sharp increase in resignation rates**, which may be caused by internal factors (such as job satisfaction, salary, work culture) or external factors (such as economic conditions, business competition).
- **No new recruitment since 2018**, indicating that the company is no longer expanding or is in cost-saving mode.
- **If this pattern continues**, the workforce may further decrease, posing a risk to business operations.

Annual Report on Employee Number Changes

Employees Joined vs. Left By Year



Periode 2010–2015, jumlah karyawan terus bertambah, puncaknya di 2011 dengan 76 rekrutmen. Namun, sejak 2016 tren berubah—resign meningkat, terutama di 2018 (26 resign, hanya 1 rekrut). Akibatnya, sejak 2017 pertumbuhan karyawan negatif, terburuk di 2018 (-25), dan terus menurun hingga 2020 (-6). Ini menunjukkan kesulitan perusahaan dalam mempertahankan tenaga kerja.

Business Insights:

- ✓ **Prevent Resignation Surge:** Identify reasons for resignations since 2017 and improve employee well-being.
- ✓ **Maintain Employee Stability:** Evaluate salaries, benefits, and work culture to reduce the decline since 2018.
- ✓ **Resume Recruitment:** Since there has been no recruitment since 2018, initiate hiring or consider outsourcing.
- ✓ **Financial Analysis:** Ensure financial conditions support recruitment and operational efficiency.