

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

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



Created by:

Haolia

haoliaaul@gmail.com

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

“Geophysical engineering graduates who diligently want to fulfil a role where intellectual, integrity, and curiosity are highly valued. Motivated, able to research, design, implement new features and learn various software. Skill handling problems with unique ways to develop innovative solutions. Proficient using Python, SQL, Tableau and other statistical tools for data multi purposes. Looking for opportunities in data analyst, data science, data engineer and Business Intelligence.”

“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
```

Menampilkan ringkasan data secara keseluruhan.

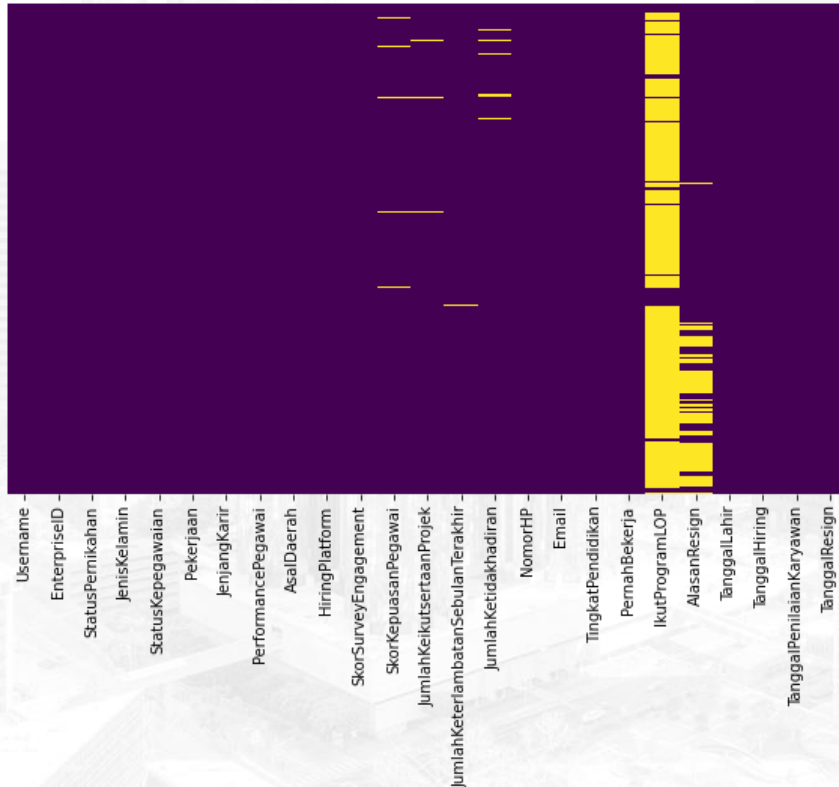
> Terdapat Missing Value pada :

- SkorKepuasanPegawai
- JumlahKeikutsertaanProjek
- JumlahKeterlibatanSebulanTerakhir
- JumlahKetidakhadiran
- IkutProgramLOP
- AlasanResign

- Tipe data yang tidak sesuai
- Melakukan Handling missing value dengan drop baris feature yang memiliki nilai missing value lebih dari 50 % dan melakukan input nilai median.

Untuk selengkapnya, dapat melihat jupyter notebook disini

Heatmap of missing values



Data Cleaning

```
# Drop 'IkutProgramLOP'
df = df.drop(['IkutProgramLOP'], axis=1)
✓ 0.4s

# AlasanResign will be filled with 'masih_bekerja' because the employees are still there
df['AlasanResign'] = df['AlasanResign'].fillna('masih_bekerja')
✓ 0.4s

import statistics
✓ 0.5s

#Input ScoreKepuasanPegawai with the mode because it is ordinal data
df['SkorKepuasanPegawai'] = df['SkorKepuasanPegawai'].fillna(statistics.mode(df['SkorKepuasanPegawai']))
✓ 0.5s

#Input other missing variables with the median
df['JumlahKeikutsertaanProjek'] = df['JumlahKeikutsertaanProjek'].fillna(df['JumlahKeikutsertaanProjek'].median())
df['JumlahKeterlambatanSebulanTerakhir'] = df['JumlahKeterlambatanSebulanTerakhir'].fillna(\
    df['JumlahKeterlambatanSebulanTerakhir'].median())
df['JumlahKetidakhadiran'] = df['JumlahKetidakhadiran'].fillna(df['JumlahKetidakhadiran'].median())
✓ 0.3s
```

Untuk selengkapnya, dapat melihat jupyter notebook disini