

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

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



Created by:

Lana Fuadi

lanafuadi@gmail.com

linkedin.com/lanafuadi

“I am a results-driven data scientist with a passion for leveraging data to drive business success. Through rigorous training at [Rakamin](#) Bootcamp, I have honed my skills in Python, ML, SQL, and Business Intelligence. I am committed to continuous learning and have completed additional certifications and course to further enhance my expertise. With strong collaboration, critical thinking, and problem solving skills, I am confident in my ability to help your company make data-driven decisions that propel your business to new heights.”

“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. “

1. Handle Unrelevant Columns:

- For efficiency, I dropped neither important nor relevant columns, such as Username, EnterpriseID, NomorHP, Email.
- I dropped dropped IkutProgramLOP because it has too many missing values, I guess imputation will not help. 258 is missing out of 287.
- I dropped PernahBekerja because it has only one unique value, it will not affect modelling.

2. Handle Missing Values:

- There are total 81 missing values spread in AlasanResign (66), JumlahKehadiran (6), SkorKepuasanPegawai (5), JumlahKeikutsertaanProjek (3), and JumlahKeterlambatanSebulanTerakhir (1).
- I use mode imputation for all of them because AlasanResign has string values, it has no mean attribute. Also the rest's means are float with many decimal while their original values are float without decimal other than 0. mean imputation will add unneeded unique value.

3. Handle Duplicate Values:

- There are no any duplicate value.