

Predicting Employee Retention using Logistic Regression

Objective:

As started in the problem, the objective is to develop a logistic regression model to predict employee retention based on demographic details, job satisfaction scores, performance metrics and tenure. By leveraging these insights, the company can focus on the improvement areas it needs to work on to retain more employees.

Dataset:

We used the employee retention dataset provided in the case study. It included features such as satisfaction level, last evaluation, number of projects, average monthly hours, time spent at the company, work accidents, promotion in the last 5 years, department, and salary.

Approach:

Exploratory Data Analysis (EDA): We analyzed distributions, correlations, and patterns in the data to understand relationships between features and attrition.

Data Preprocessing: Handled missing values (if any), encoded categorical variables, scaled numerical features, and checked for multicollinearity using VIF.

Model Building: Used Logistic Regression to build a binary classification model. We dropped variables causing high multicollinearity to improve model performance.

Model Evaluation: Evaluated the model using accuracy, precision, recall, F1-score, and confusion matrix. The model performed well and gave us useful insights into which factors influence employee attrition.

Conclusion:

Logistic regression proved effective in predicting employee attrition. Key influencing factors included satisfaction level, average monthly hours, and number of projects. The model can help HR teams proactively identify at-risk employees.

Tools & Libraries Used:

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels