

Employee Attrition Analysis

Project Overview

With the use of Logistic Regression, the aim of this project is to identify key factors contributing to employee attrition and understanding the factors that can help in developing targeted strategies to retain valuable employees.

Objectives

1. Identify key factors contributing to employee attrition.
2. Develop a logistic regression model to predict attrition.
3. Evaluate the model's performance and interpret its coefficients.

Data

The dataset contains various features such as demographic information, job characteristics, compensation details, and satisfaction metrics which are as follows:

- Age
- Attrition
- BusinessTravel
- DailyRate
- Department
- DistanceFromHome
- Education
- EducationField
- EmployeeCount
- EmployeeNumber
- EnvironmentSatisfaction
- Gender
- HourlyRate
- JobInvolvement
- JobLevel

- JobRole
- JobSatisfaction
- MaritalStatus
- MonthlyIncome
- MonthlyRate
- NumCompaniesWorked
- Over18
- OverTime
- PercentSalaryHike
- PerformanceRating
- RelationshipSatisfaction
- StandardHours
- StockOptionLevel
- TotalWorkingYears
- TrainingTimesLastYear
- WorkLifeBalance
- YearsAtCompany
- YearsInCurrentRole
- YearsSinceLastPromotion
- YearsWithCurrManager

Methodology

1. **Data Preprocessing:** Check for and handle missing values, encode categorical variables, and scale numerical features.
2. **Exploratory Data Analysis:** Understand distributions and relationships between variables.
3. **Modeling:** Train a logistic regression model to predict employee attrition.
4. **Evaluation:** Evaluate the model's performance using accuracy, confusion matrix, and classification report.

Results

The logistic regression model achieved an accuracy of approximately **89.80%**. This model allows us to:

1. Identify significant predictors of employee attrition.
2. Understand the direction and magnitude of the relationship between predictors and attrition.

Key predictors of attrition included:

1. **OverTime**: Employees working overtime are much more likely to leave.
2. **YearsAtCompany**: Longer tenure at the company increased the likelihood of attrition.
3. **Department**: Certain departments had higher attrition rates.
4. **MaritalStatus**: Marital status influenced the likelihood of attrition.

Visuals

Visualizations are included in the `images` directory.

Usage

1. Clone the repository.
2. Install the required dependencies.
3. Run the scripts in the Jupyter notebooks in the `notebooks` directory.

Source

Data used for this project is sourced from Kaggle. This is a fictional data set created by IBM data scientists. (<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>).