

Predicting Employee Attrition

Introduction

Employee attrition is defined as the natural process by which employees leave the workforce – for example, through resignation for personal reasons or retirement – and are not immediately replaced.' Attrition is an inevitable part of any business. But when attrition crosses a particular threshold, it becomes a cause for concern. There will come a time when an employee wants to leave your company – for either personal or professional reasons. HR Analytics helps human resources to interpret data, find out the trends & help take required steps to keep the organization running smoothly & profitably. It has helped human resources to be more active and gain involved role in an organization's planning & objectives. The proposed research project on employee attrition using machine learning emphasizes on uncovering the factors that lead to employee attrition and explore important questions such as 'show me a breakdown of distance from home by job role and attrition' or 'compare average monthly income by education and attrition'.

About the Dataset

The dataset is about 1470 rows and 35 attributes which consist of 8 categorical variables and 26 numerical variables and the target variable is 'attrition'.

- 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

Challenges

1. Perform Exploratory Data Analysis on employee attrition to identify key features that are responsible for attrition.
2. Build Classification models using Knn, Logistic Regression, SVM, Random Forest.
3. Compare metrics of various models to find the optimal model.