Problem Statement - HR Attrition
Case Study: HR Attrition using Ensemble Techniques

Context:

- McCurr Consultancy is an MNC that has thousands of employees spread across the globe. The company believes in hiring the best talent available and retaining them for as long as possible. A huge amount of resources is spent on retaining existing employees through various initiatives. The Head of People Operations wants to bring down the cost of retaining employees. For this, he proposes limiting the incentives to only those employees who are at risk of attrition. As a recently hired Data Scientist in the People Operations Department, you have been asked to identify patterns in characteristics of employees who leave the organization. Also, you have to use this information to predict if an employee is at risk of attrition. This information will be used to target them with incentives.

Problem:

The data-set aims to answer the following key questions:

- What are the different factors that can help in identifying attriting employees?

- Can we build a model to predict the attrition of employees? What should be the metric of choice to evaluate such a model?

Attribute Information:

- EmployeeNumber - Employee Identifier

- Attrition - Did the employee attrite?
- Age - Age of the employee
- BusinessTravel - Travel commitments for the job
- DailyRate - Data description not available
- Department - Employee Department
- DistanceFromHome - Distance from work to home (in km)
- Education - 1-Below College, 2-College, 3-Bachelor, 4-Master,5-Doctor
- EducationField - Field of Education
- EmployeeCount - Employee Count in a row
- EnvironmentSatisfaction - 1-Low, 2-Medium, 3-High, 4-Very High
- Gender - Employee's gender
- HourlyRate - Data description not available
- JobInvolvement - 1-Low, 2-Medium, 3-High, 4-Very High

- JobLevel - Level of job (1 to 5)
- JobRole - Job Roles
- JobSatisfaction - 1-Low, 2-Medium, 3-High, 4-Very High
- MaritalStatus - Marital Status
- MonthlyIncome - Monthly Salary
- MonthlyRate - Data description not available
- NumCompaniesWorked - Number of companies worked at
- Over18 - Over 18 years of age?
- OverTime - Overtime?
- PercentSalaryHike - The percentage increase in salary last year
- PerformanceRating - 1-Low, 2-Good, 3-Excellent, 4-Outstanding
- RelationshipSatisfaction - 1-Low, 2-Medium, 3-High, 4-Very High
- StandardHours - Standard Hours

- StockOptionLevel - Stock Option Level
- TotalWorkingYears - Total years worked
- TrainingTimesLastYear - Number of training attended last year
- WorkLifeBalance - 1-Low, 2-Good, 3-Excellent, 4-Outstanding
- YearsAtCompany - Years at Company
- YearsInCurrentRole - Years in the current role
- YearsSinceLastPromotion - Years since the last promotion
- YearsWithCurrManager - Years with the current manager

Learning Outcomes:

- Exploratory Data Analysis
- Preparing the data to train a model
- Training and understanding of data using ensemble models
- Model evaluation

Steps and Tasks:

- Import Libraries and Load Dataset
- Overview of data
- Data Visualization
- Data preparation
- Choose Model, Train, and Evaluate

- Conclusion