

Predictive Modeling for Employee Attrition

Introduction

Employee attrition is a critical challenge faced by organizations across various industries. Predicting employee attrition accurately can help organizations develop effective retention strategies and maintain a stable workforce. This report explores the process of predicting employee attrition using machine learning techniques.

Dataset Description

The dataset used in this analysis contains information about employees, including their demographics, job role, work environment satisfaction, performance ratings, and whether they have left the company (attrition). The dataset consists of X features and Y observations.

Below is a detailed report based on the provided code for data preprocessing, analysis, and machine learning model training and evaluation.

1. Data Loading and Preprocessing

Dataset:

The dataset used for this analysis is located at

C:/Users/sakss/Desktop/technocolabs/WA_Fn-UseC_-HR-Employee-Attrition.csv.

Data Cleaning:

- Removed irrelevant columns: EmployeeCount, StandardHours, Over18, and EmployeeNumber.
- Encoded categorical variables:
 - Attrition: 'Yes' -> 1, 'No' -> 0
 - Gender: 'Male' -> 1, 'Female' -> 0
 - OverTime: 'Yes' -> 1, 'No' -> 0

One-Hot Encoding:

Remaining categorical columns were one-hot encoded to convert them into a format suitable for machine learning algorithms.

Descriptive Statistics and Correlation Matrix:

Descriptive statistics were generated for the cleaned dataset, and a correlation matrix was created to understand the relationships between variables.

2. Data Visualization

Various plots were created to visualize the distribution of attrition across different features:

- **Department-wise Attrition:**
 - Count of employees by department and their attrition status.
- **Job Role and Attrition:**
 - Count of employees by job role and their attrition status.
- **Education Level and Attrition:**
 - Count of employees by education level and their attrition status.
- **Marital Status and Attrition:**
 - Count of employees by marital status and their attrition status.
- **Work-Life Balance and Attrition:**
 - Count of employees by work-life balance rating and their attrition status.
- **Job Satisfaction by Attrition:**
 - Count of employees by job satisfaction rating and their attrition status.
- **Age Distribution by Attrition:**
 - Distribution of employees' ages and their attrition status.
- **Monthly Income Distribution by Attrition:**
 - Distribution of employees' monthly income and their attrition status.
- **Years at Company and Attrition:**
 - Distribution of employees' years at the company and their attrition status.
- **Gender and Attrition:**
 - Count of employees by gender and their attrition status.
- **OverTime and Attrition:**
 - Count of employees by overtime status and their attrition status.
- **Age vs Monthly Income by Attrition:**
 - Scatter plot of employees' age versus monthly income colored by attrition status.
- **Correlation Matrix Heatmap:**
 - Heatmap showing the correlations between all numerical features.

3. Machine Learning Models

The dataset was split into training (70%) and testing (30%) sets. Features were standardized using a scaler.

Models Trained:

1. **Logistic Regression**
2. **Random Forest Classifier**
3. **Gradient Boosting Classifier**

Evaluation Metrics: For each model, the following metrics were computed:

- Confusion Matrix
- Classification Report (Precision, Recall, F1-Score)

- Accuracy Score