Predicting Employee Attrition using HR Analytics

Name: Shrishti Mishra

Tools Used: Python, Power BI, Scikit-learn

Date: 22-06-2025

# Abstract-

This project aims to identify the key drivers of employee attrition using HR analytics and predictive modelling. Attrition, or employee turnover, poses a significant challenge to organizations in terms of knowledge loss, rehiring costs, and productivity gaps. We apply machine learning algorithms on HR data to predict which employees are most likely to leave and generate actionable insights. The project combines Python for model building and Power BI for interactive dashboards, helping HR teams to make data-driven decisions for employee retention. The final solution provides a foundation for developing an early warning system for resignations.

# Problem Statement

High employee attrition impacts team performance, increases operational costs, and slows down project timelines. Traditional HR practices rely on reactive responses rather than preventive strategies. This project aims to bridge that gap by analyzing past employee data to uncover patterns related to voluntary exits. We address questions such as: “Which departments have the highest attrition?”, “What employee characteristics influence resignations?”, and “Can we predict who might leave next?”. The goal is to turn raw HR data into insights that can guide proactive interventions.

# 3. Tools & Technologies Used

|  |  |
| --- | --- |
| Tool | Purpose |
| Python (Pandas, Seaborn, Scikit-learn) | Data cleaning, EDA, model building |
| Power BI | Dashboard & visual storytelling |
| Jupyter Notebook | Code execution |
| Decision Tree / Logistic Regression | Prediction models |

# 4. Data Overview

The dataset contains 1,470 employee records across multiple departments. Each record includes demographic, professional, and performance-related attributes such as Age, Gender, Department, Monthly Income, OverTime status, Job Role, Distance from Home, Years at Company, and more.

The target column is Attrition with binary values: "Yes" (employee left) and "No" (employee stayed). The dataset was clean with minimal missing values, allowing efficient model development. One-hot encoding and standard scaling were applied during preprocessing.

# 5. Exploratory Data Analysis (EDA)

* High attrition in Sales and HR departments
* Employees with OverTime and low Monthly Income showed higher resignation rates
* Younger and newer employees were more likely to leave

# 6. Model Building

We used Decision Tree Classifier (and Logistic Regression). Train-Test split: 80:20. Accuracy achieved: ~85%. Feature Importance showed: OverTime, MonthlyIncome, JobSatisfaction as top factors.

# 7. Dashboard Summary (Power BI)

* Included **KPI cards** for:
  + Total Employees
  + Total Attrition
  + Attrition Rate (%)
* Used **stacked bar charts** to compare Attrition across:
  + Departments
  + Gender
* Added visualizations for:
  + Monthly Income vs Attrition
  + Age vs Attrition (bucketed age groups)
* Integrated **slicers (filters)** for:
  + Gender.
* Imported feature importance chart (from model) as an image visual
* Enabled interactive exploration for HR teams to filter and compare attrition patterns in real time

# 8. Recommendations

* Reduce Overtime in high-risk departments
* Improve job satisfaction for new joiners
* Monitor low-income employees for proactive engagement

# 9. Conclusion

By integrating machine learning and business intelligence tools, this project demonstrates how HR data can be transformed into meaningful insights. Predictive models help identify high-risk employees, while dashboards allow dynamic interaction with attrition trends. This solution enables data-driven HR decisions, reduces talent loss, and improves employee satisfaction. Future work may include integrating real-time data feeds, automating alerts for risky profiles, or extending the analysis to include external factors like economic shifts or industry trends.