ELEVATE LABS

HR ANALYTICS- ATTRITION PREDICTION

REPORT

BY SRUTISREEYA JENA

MAILTO:SRUTISREEYA@GMAIL.COM

HTTPS://GITHUB.COM/SRUTI-SREEYA

Introduction

In today's competitive talent landscape, retaining skilled employees is critical to organizational success. Employee attrition—when staff voluntarily or involuntarily leave—can disrupt productivity and increase costs. This project addresses attrition using HR analytics, by building a machine learning model in Python to predict at-risk employees and visualizing trends through Power BI dashboards.

Abstract

The goal of this project is to identify key drivers of employee attrition and develop proactive retention strategies. The workflow begins with data cleaning and preprocessing of an HR dataset, followed by exploratory analysis and predictive modeling in Python using Random Forest and Logistic Regression. Post-modeling, the cleaned data is fed into Power BI to build interactive dashboards for stakeholders to visualize and monitor attrition patterns. The combination of Python and Power BI provides both analytical depth and business-friendly visualization.

Tools Used

Python (Jupyter Notebook): For data preprocessing, visualization, and machine learning modeling.
Python Libraries: pandas, numpy, seaborn, matplotlib, sklearn.
PowerBI: To build interactive dashboards using cleaned HR data.
Dataset: "WA_Fn-UseC_-HR-Employee-Attrition.csv", preprocessed in Python and used in Power BI.
Models Implemented: Random Forest Classification Model, XGBoost Classification Model

Steps Involved in Building the Project

Step 1: Data Preprocessing:
Missing values and inconsistent entries were handled.
Categorical variables like BusinessTravel, MaritalStatus, OverTime were label-encoded.
Data imbalance was considered, and feature scaling was applied for numerical features.

Step 2: Exploratory Data Analysis (EDA)
Correlation heatmaps, bar charts, and box plots showed key attrition drivers:
Low job/environment satisfaction
Low monthly income
Excessive overtime
Greater distance from home
Low work-life balance

Step 3: Model Building (Python)
Used train-test split (80–20) to evaluate performance.
Random Forest had the best accuracy (~87%) and ROC-AUC score among tested models.
Feature importance revealed OverTime, JobSatisfaction, and MonthlyIncome as top predictors.

Step 4: Dashboard Creation (Power BI)
Imported cleaned data (hr_cleaned_for_powerbi.csv) into Power BI.
Visuals created:
Attrition by Department, Gender, Job Role
Attrition by Income Level, Overtime, Distance from Home
Filtered views using slicers (Department, Education, etc.)
Dashboard highlights included high attrition among Entry-Level, Sales, and overtime-working employees.

Conclusion

This project successfully demonstrates how HR analytics can be leveraged to reduce employee attrition. The machine learning model identifies high-risk employees, while Power BI dashboards help HR teams monitor trends and make informed decisions. The insights suggest focusing on improving work-life balance, fair compensation, and recognition to retain talent. This dual-approach project is a scalable framework for real-world HR departments aiming to reduce turnover using data.