HR Analytics – Employee Attrition Prediction Using Machine Learning

Introduction:

Organizations often struggle to identify which employees are likely to resign and why. Traditional methods such as surveys and manual reviews are limited, subjective, and slow to analyze. In this project, a real-world HR dataset is examined to uncover behavioural and demographic patterns that correlate with employee attrition. The study focuses on multiple factors such as salary levels, working conditions, overtime workload, travel obligations, job roles, and tenure. A machine learning classifier is used to predict employees at risk of leaving, while a Power BI dashboard presents the results in an interactive format. The ultimate objective is to transform HR decision-making from reactive to proactive using data analytics.

Abstract:

Employee attrition not only increases recruitment expenses but also causes the loss of organizational knowledge and affects overall productivity. This project uses HR Analytics and Machine Learning to analyse key drivers behind employee resignation and forecast future attrition risks. By performing data preprocessing, exploratory data analysis, model building, and visualization, the project delivers data-backed insights that support HR decision-making. The predictive results and dashboard visualizations can guide targeted retention strategies, allowing organizations to proactively reduce turnover.
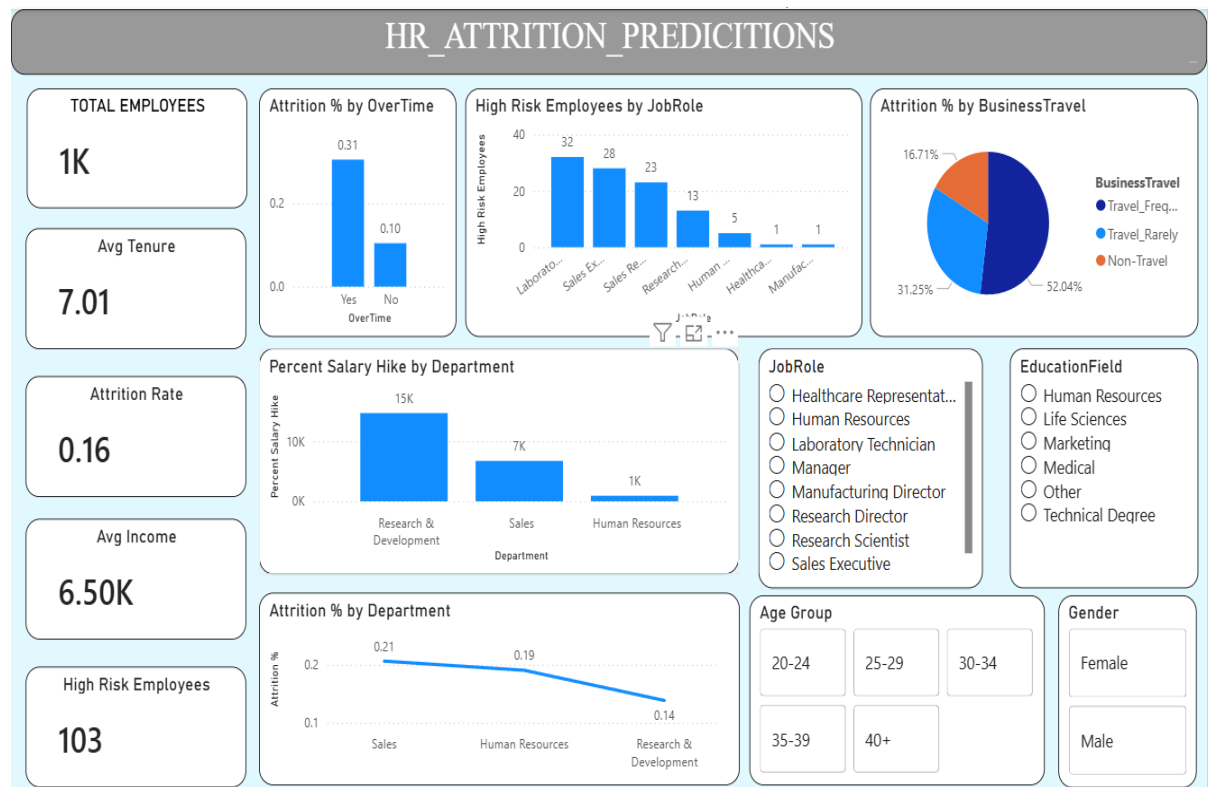
Tools Used:

Tools	Purpose
Python (Pandas, Seaborn)	Data cleaning & EDA
Scikit-Learn (Logistic Regression)	Predictive modeling
SHAP	Feature interpretation
Power BI	Dashboards & KPI visualization

Steps Involved in Building the Project:

1. **Dataset Preparation:** Loaded the HR employee attrition dataset, checked missing values, and removed irrelevant attributes such as EmployeeCount, Over18, and StandardHours.
2. **Exploratory Data Analysis (EDA):** Analyzed attrition trends by department, salary band, business travel, job role, promotions, and overtime through graphical distributions and comparisons.
3. **Model Development:** Constructed a Logistic Regression classifier using a pipeline with one-hot encoding for categorical variables and standard scaling for numerical features.
4. **Model Evaluation:** Measured performance using accuracy score and confusion matrix to assess prediction capability for identifying employees at risk.
5. **Explainable AI:** Applied SHAP value analysis to determine the key factors influencing attrition, highlighting overtime, salary level, and job role as major contributors.

6. **Dashboard Creation:** Built an interactive Power BI dashboard showcasing attrition KPIs, high-risk roles, salary insights, tenure risk, and business travel impact using slicers and visual reports.

DASHBOARD SCREENSHOT:

Conclusion:

This project demonstrates that HR analytics integrated with machine learning provides an effective method for understanding and predicting employee attrition. By analyzing both individual employee features and overall workforce trends, HR departments can identify high-risk groups and take corrective action before resignation occurs. Insights gained from overtime patterns, promotion delays, salary disparities, and role-specific stress indicate that proactive retention such as salary revisions, workload management, career growth plans, and employee well-being initiative can significantly reduce attrition. The predictive dashboard empowers HR teams with data-driven decision support, improving organizational stability and long-term workforce planning.