PROJECT REPORT ON

**HR Analytics - Predict Employee Attrition**

**INTRODUCTION**

This project focuses on analyzing employee data to identify factors that lead to voluntary resignations. The objective is to build a predictive model that can help HR teams anticipate attrition and take proactive steps to retain valuable talent. Insights derived from this analysis can support strategic HR decision-making and improve employee retention.

**ABSTRACT**

The HR Analytics project aims to predict employee attrition by analyzing organizational workforce data. By applying machine learning techniques, the goal is to identify key factors that influence voluntary resignation and provide actionable insights to improve employee retention. The project involved comprehensive data preprocessing, including handling missing values, encoding categorical variables, and treating outliers. A Random Forest Classifier was developed and evaluated, followed by SHAP value analysis to interpret model predictions. The findings highlight critical drivers of attrition such as overtime, salary levels, job roles, and tenure, enabling HR departments to make data-driven decisions to reduce turnover.

**OBJECTIVE**

Use analytics to understand the main causes of employee resignation and predict future attrition.

**TOOLS:**

Python (Pandas, Seaborn), Power BI, Sklearn

**STEPS INVOLVED**

1. **Data Understanding**

Analyzed a structured HR dataset containing employee information such as age, salary, job role, department, and attrition status.

1. **Data Cleaning & Preprocessing**

Handled missing values, removed or capped outliers, and transformed categorical variables using one-hot encoding.

Converted the target variable (Attrition) into binary format for modeling.

1. **Exploratory Data Analysis (EDA)**

Performed univariate and bivariate analysis using visual tools (Seaborn, Matplotlib) to detect patterns in attrition across roles, departments, and other variables.

1. **Model Building**

Trained a **Random Forest Classifier** with class imbalance handling .

1. **Model Interpretability (SHAP Analysis)**

Used **SHAP values** to interpret model outputs and understand which features most influenced employee attrition.

1. **Dashboard & Visualization**

Created dashboards using **Power BI** and Python to visualize attrition trends, salary distribution, and key metrics.

**CONCLUSION**

* The analysis showed that some work conditions strongly affect whether an employee stays or leaves.
* Employees who work overtime, have lower monthly income, or have not been promoted or changed roles for a long time are more likely to resign.
* Job role and department also play a role — for example, some roles have higher stress or fewer growth opportunities, which leads to more resignations.
* By using these patterns, the model helps HR teams understand which employees might leave soon.

**DASHBOARD INSIGHTS**

The dashboard reveals several key insights into employee attrition. Out of 23.2K employees, 3,674 have left the company, resulting in a 15.83% attrition rate. Female employees account for a larger share of attrition (around 61%) compared to males. Most resignations occur between the ages of 25 and 35, indicating younger employees are more likely to leave. Roles such as Sales Representative and Laboratory Technician show the highest attrition counts, highlighting potential dissatisfaction or stress in these positions. Additionally, single employees tend to leave more often than married or divorced ones. Department-wise, Sales and R&D experience the most attrition, signaling a need for focused employee engagement strategies in those areas.

**RECOMMENDATION**

* Focus on improving job satisfaction in high-attrition roles like Sales and Lab Technicians.
* Offer career development and engagement programs for younger (25–35) and single employees.
* Investigate reasons for high attrition in the Sales and R&D departments through employee feedback.