# HR Analytics Report: Predicting Employee Attrition

## Introduction

Employee attrition affects organizational stability, productivity, and morale. This project uses data-driven analysis and machine learning to identify the primary drivers of attrition and predict which employees are at risk of leaving, enabling targeted retention strategies.

---

## Abstract

This analysis investigates HR data to uncover attrition trends and build predictive models. The process includes classification modeling using Decision Tree and Random Forest classifiers, followed by SHAP (SHapley Additive exPlanations) to interpret the impact of each feature. A Power BI dashboard supports dynamic visual analysis to assist decision-makers.

---

## Tools Used

- **Python Libraries**: Pandas, Seaborn, Scikit-learn, SHAP
- **Visualization Tools**: Power BI
- **Techniques**: Classification (Decision Tree, Random Forest), Model Explanation using SHAP

## Steps Involved in the Project

### 1. Exploratory Data Analysis (EDA)

Power BI dashboard insights:

- **Attrition Rate**: 16.12% (237 out of 1470 employees).
- **Department-wise Attrition**: Most from **R&D (133)** and **Sales (92)**.
- **Age Group**: Highest attrition in the **21–40** age range.
- **Gender**: **Male employees** showed higher attrition.
- **Job Satisfaction**: Employees with lower satisfaction (levels 1 & 2) tended to leave more.
- **Years Since Last Promotion**: Higher attrition seen at 0 years (i.e., no recent promotion).
- **Marital Status**: Single employees were more likely to leave.

### 2. Data Preprocessing

- Categorical features (Gender, Marital Status, Department, etc.) were label encoded.
- Target variable ("Attrition") was encoded to binary.
- **Class imbalance** was handled using **class weights** in the model.

**3. Model Building**

- Models Used: **Decision Tree** and **Random Forest**.
- Class weights were applied to reduce bias toward the majority class.
- Evaluation metrics (accuracy, precision, recall) were calculated, with **Random Forest performing better** in predictive accuracy.

**4. SHAP Analysis**

- SHAP (SHapley Additive Explanations) was used to interpret model predictions.
- Key features influencing attrition:
    - **Age**
    - **YearsAtCompany**
    - **MaritalStatus**
    - **JobSatisfaction**
    - **Work-Life Balance**
- SHAP plots provided both global and local feature attributions.

**5. Power BI Visualization**

The dashboard visualizes:

- Filters by **Age Group**, **Job Role**, & **OverTime.**
- Charts for attrition distribution across **departments**, **genders**, **age**, & **satisfaction.**
- Interactive exploration for data-driven HR decisions

---

## Conclusion & Recommendations

**Key Insights:**

- High attrition among younger, single, male employees in R&D and Sales.
- Career stagnation (e.g., no recent promotions) and overtime significantly contribute to attrition.
- Low job satisfaction and poor work-life balance are key red flags.

**Recommendations:**

1. **Enhance Career Growth**: Create structured promotion pipelines and skill development programs.
2. **Boost Engagement**: Address job dissatisfaction through pulse surveys and feedback mechanisms.
3. **Regulate Overtime**: Promote healthy work-life balance and avoid burnout.
4. **Personalized Interventions**: Use model outputs to identify at-risk employees and implement tailored retention strategies.
5. **Leverage Predictive Monitoring**: Deploy the Random Forest model in HR systems for real-time attrition risk tracking.