

# HR Analytics – Predict Employee Attrition

---

## Introduction

Employee attrition poses significant challenges to business continuity, increasing recruitment costs and disrupting team dynamics. This project uses data-driven techniques to understand the factors leading to employee resignation and builds predictive models to help HR departments proactively identify at-risk employees.

---

## Abstract

This project analyzes HR data to uncover trends and patterns associated with employee attrition. We performed thorough data cleaning, encoding, and exploratory data analysis (EDA) to identify influential features. Two machine learning models—**Logistic Regression** and **Decision Tree**—were implemented to predict employee attrition. The models were evaluated using accuracy metrics, classification reports, and confusion matrices.

We also created a Power BI dashboard that allows for interactive analysis of attrition based on various features such as gender, department, marital status, and overtime status. These insights help HR departments make informed and strategic decisions to reduce turnover.

---

## Tools Used

- **Python:** Pandas, Seaborn, Scikit-learn, Matplotlib
  - **Power BI**
  - **Jupyter Notebook**
  - **LabelEncoder, StandardScaler, LogisticRegression, DecisionTreeClassifier**
- 

## Steps Involved

1. **Data Cleaning**  
Removed redundant columns (EmployeeCount, StandardHours, Over18) and verified the absence of missing values.
2. **Exploratory Data Analysis (EDA)**  
Visualized attrition across features like Gender, Department, OverTime, JobRole, and MaritalStatus. Observed higher attrition among employees in Sales and R&D, particularly those working overtime.
3. **Preprocessing**  
Encoded categorical variables using LabelEncoder. Scaled numerical values using

StandardScaler. Split the data into 80% training and 20% testing sets with stratification.

#### 4. Model Building

- **Logistic Regression:** A linear model was trained as a baseline for binary classification. It helped in understanding feature coefficients and the linear relationship between predictors and attrition.
- **Decision Tree Classifier:** A non-linear model trained using the entropy criterion, ideal for interpreting complex decision paths and feature interactions.

#### 5. Model Evaluation

- Both models were evaluated using classification reports and confusion matrices.
- Logistic Regression provided good precision and recall for balanced classes.
- Decision Tree revealed clearer boundaries and feature importance rankings.

#### 6. Feature Importance & Interpretation

Features such as OverTime, JobRole, MonthlyIncome, and Age were identified as key drivers of attrition.

#### 7. Dashboard Development

Created a dynamic Power BI dashboard to visualize attrition trends. Users can filter by department, gender, education field, and more to gain actionable insights.

---

### Conclusion

This HR Analytics project successfully combined exploratory data analysis, visualization, and machine learning to address the challenge of employee attrition. Using both **Logistic Regression** and **Decision Tree**, we gained diverse perspectives on the attrition problem—linear interpretability and non-linear decision-making.

#### Key Findings:

- Employees working overtime or in specific job roles showed higher attrition risk.
- Logistic Regression provided a baseline understanding of feature influence.
- Decision Trees offered deeper interpretability and feature importance metrics.
- Power BI dashboards facilitated interactive analysis for stakeholders.

Moving forward, incorporating behavioral and performance data, along with advanced models like Random Forest or Gradient Boosting, could further improve prediction accuracy and strategic HR planning.

