# Employee Attrition Prediction & Prevention Suggestions


## Abstract

Employee attrition remains a pressing concern for modern organizations, directly impacting workforce morale, operational stability, and financial performance. This project leverages machine learning techniques to predict employee attrition and identify contributing factors using HR data. A Decision Tree Classifier model was built and evaluated, followed by model explainability using SHAP to gain transparent insights. The results were visualized using Power BI, and based on the analysis, targeted recommendations for preventing attrition were developed to assist HR decision-making.


## Introduction

In the competitive landscape of talent management, employee retention is key to sustaining organizational growth. Attrition leads to increased costs in recruitment, training, and productivity loss. Predictive analytics can play a pivotal role in understanding and addressing the root causes of attrition. This project develops a data-driven approach using machine learning to identify employees at risk of leaving and provides actionable insights to support strategic HR initiatives.

## Tools Used

- Python: Programming language used for data processing, model training, and evaluation.
- Pandas & NumPy: Used for efficient data wrangling, manipulation, and statistical analysis.
- Seaborn & Matplotlib: Used to create insightful visualizations during EDA.
- Scikit-learn: Used to build and evaluate the Decision Tree Classifier.
- SHAP (SHapley Additive exPlanations): Used to interpret model predictions and assess feature contributions.
- Power BI: Designed interactive dashboards to allow dynamic visualization of attrition patterns.
- Jupyter Notebook: Served as the integrated development environment for all project stages.

## Steps Involved in Building the Project

1. Data Collection and Preprocessing

- - The HR dataset included features like employee ID, department, job role, monthly income, satisfaction level, and attrition status.
  - Categorical features such as Department and Job Role were encoded for model compatibility.
  - Derived features like 'Salary Band' and 'Tenure Group' to enhance model interpretability.
  - Missing data was handled appropriately to ensure clean and robust datasets for training.

2. Exploratory Data Analysis (EDA)

- - Conducted visual analysis to detect trends, such as higher attrition among certain departments and roles.
  - Discovered that lower income groups and employees with fewer years at the company had higher turnover.
  - Created visual plots to communicate these patterns effectively.

3. Model Training and Evaluation

- - The dataset was split into training (80%) and testing (20%) subsets.
  - A Decision Tree Classifier was chosen for its interpretability and ability to capture non-linear relationships.
  - The model was evaluated using accuracy, precision, recall, and confusion matrix metrics.
  - Fine-tuned hyperparameters like max_depth and min_samples_split to optimize performance.

4. Model Explainability with SHAP

- - SHAP values were used to explain how each feature influenced model predictions.
  - Found that 'Job Role', 'Monthly Income', 'Work-Life Balance', and 'OverTime' were the most impactful features.
  - Visualized feature importance using SHAP summary and dependency plots.

5. Power BI Dashboard

- - An interactive dashboard was built to allow HR professionals to filter and explore attrition trends by department, job role, and other criteria.
  - Enabled drill-down capabilities and dynamic filtering for real-time insights.

## Attrition Prevention Suggestions

- **Competitive Compensation:** Regular salary benchmarking and targeted adjustments for underpaid roles.
- **Career Development:** Establish mentorship programs and transparent promotion policies.
- **Work Environment:** Conduct team satisfaction surveys and improve leadership communication.
- **Employee Engagement:**Offer recognition programs, flexible hours, and mental health support.
- **Work-Life Balance:** Provide hybrid or remote work options, manage workloads, and encourage time off.

## Conclusion

The project successfully developed an explainable machine learning pipeline to predict employee attrition. It offered detailed insights into the underlying causes of turnover, supported by data visualization and interpretable model output. These findings can help HR departments tailor proactive policies to retain valuable employees. Future enhancements could involve integrating real-time data and testing advanced models like XGBoost or Random Forest for higher predictive accuracy.

Prepared by: **MOHAMMAD AAS KHAN**

Date: 19/05/2025