

# IBM HR Analytics: Employee Attrition & Performance

---

## Table of Contents

1. Introduction
2. Problem Statement / Objective
3. Dataset Description
4. Tools and Technologies Used
5. Data Cleaning and Preprocessing
6. Exploratory Data Analysis (EDA)
7. Feature Engineering
8. Modeling / Analysis
9. Results and Interpretation
10. Limitations and Future Scope
11. Conclusion

## Introduction

This project focuses on analyzing employee attrition using HR analytics. The objective is to uncover patterns that influence employees' decisions to leave the company and predict potential attrition using machine learning models.

## Problem Statement

The primary goal is to identify key factors that lead to employee attrition and develop predictive models to assist HR in making data-driven decisions to retain talent.

## Dataset Description

The dataset used is the 'IBM HR Analytics Employee Attrition & Performance' dataset. It includes various features such as employee demographics, job role, work environment, and performance ratings. The target variable is 'Attrition'.

## Tools and Technologies Used

- Python
- Jupyter Notebook
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn

## Data Cleaning and Preprocessing

Missing values were checked and handled, data types were verified and adjusted as needed. Categorical variables were encoded using label encoding and one-hot encoding. The dataset was then split into training and testing sets.

## Exploratory Data Analysis (EDA)

EDA was conducted to understand the distribution and relationship of features with the target variable. Key visualizations included bar plots, histograms, and correlation heatmaps. Overtime, job satisfaction, and monthly income showed strong correlation with attrition.

## Feature Engineering

New features were derived where necessary and irrelevant or redundant features were removed to improve model performance. Categorical variables were encoded and the dataset was scaled.

## Modeling / Analysis

Multiple machine learning models were trained including Logistic Regression, Random Forest, and Decision Trees. Hyperparameter tuning was performed to improve model accuracy.

## Results and Interpretation

Random Forest performed the best with the highest accuracy. Key predictors of attrition included overtime, job satisfaction, and income. Model metrics such as accuracy, precision, and recall were used for evaluation.

## Limitations and Future Scope

The dataset is relatively small and may not generalize well. Future work could include deep learning models, integration with real-time HR systems, and more granular features like sentiment analysis from employee feedback.

## Conclusion

The project successfully demonstrated how HR analytics can be leveraged to predict employee attrition. The insights derived can be used by HR professionals to enhance employee retention strategies.