HR Analytics Project Report

# Introduction

Employee attrition poses a significant challenge to organizations, affecting operational continuity, morale, and financial performance. Human Resource (HR) departments aim not only to understand the causes of attrition but also to develop predictive systems to identify employees at risk of leaving. This project is centered on building a machine learning-based solution that leverages historical HR data to analyze patterns, uncover key indicators of employee turnover, and offer proactive recommendations for retention.

# Abstract

This project focuses on designing a data-driven approach to predict employee attrition and understand its underlying drivers. Using a comprehensive dataset comprising employee demographics, job roles, income, satisfaction levels, and promotion history, we conducted exploratory data analysis (EDA) and built classification models to detect potential attrition cases. The Decision Tree Classifier emerged as the best-performing model, achieving an accuracy of 76.4%. Key findings indicate that frequent overtime, lower income levels, lack of recent promotions, and poor work-life balance significantly contribute to attrition. These insights empower HR leaders to take data-backed actions to reduce employee turnover and improve workforce stability.

# Tools Used

• Python Libraries: Pandas, NumPy, Scikit-learn for data preprocessing, model building and evaluation.  
• Matplotlib & Seaborn: Visualization and EDA.  
• Google Colab: Development environment.  
• SHAP (optional): Model explainability through feature importance.  
• Joblib: Model serialization.  
• Power BI: Dashboard integration using the processed dataset.

# Steps Involved in Building the Project

1. Data Preprocessing: Cleaned the dataset, mapped the target variable, applied one-hot encoding, and scaled numerical features.  
2. Exploratory Data Analysis (EDA): Visualized attrition trends across departments, salary bands, and promotion history, identifying high-risk areas.  
3. Model Development: Trained Logistic Regression and Decision Tree Classifiers, employing class balancing and grid search for tuning.  
4. Model Evaluation: Used accuracy, precision, recall, F1-score, and confusion matrix, with the Decision Tree achieving 76.42% accuracy.  
5. Deployment & Output Generation: Saved the best model, exported a processed dataset, and generated visual assets for reporting.

# Conclusion

This HR analytics project successfully demonstrates how machine learning can be leveraged to predict employee attrition and support data-informed decision-making in talent management. By identifying key risk factors such as overtime, compensation, promotion stagnation, and department-level trends, the model provides HR professionals with a valuable tool to anticipate employee exits. With a predictive accuracy of over 76%, the system can assist organizations in minimizing attrition by enabling timely and targeted interventions, contributing to strategic workforce planning, improved retention, and long-term organizational health.

Note: This report is limited to two pages for concise executive communication.