

Project Report – HR Analytics: Employee Attrition Prediction

Project Objective:

The aim of this project is to analyze IBM HR Analytics Employee Attrition data and predict employee attrition using machine learning.

The insights will help HR teams identify key factors driving attrition and improve retention strategies.

Dataset:

- IBM HR Analytics Employee Attrition Dataset
- 35 features, ~1470 employee records

Tools & Technologies Used:

- Platform: Kaggle Notebook
- Language: Python
- Libraries: Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn

Workflow:

1. Data Import & Preparation

- Loaded dataset into Kaggle
- Removed ID/constant columns
- Created Attrition_Flag (Yes=1, No=0)

2. Exploratory Data Analysis (EDA)

- Visualized attrition distribution
- Examined attrition by Department, Age, Job Role, Overtime, Monthly Income
- Key insights:
 - * Higher attrition in Sales & Overtime employees
 - * Younger employees more likely to leave

3. Feature Engineering

- One-hot encoded categorical variables
- Prepared dataset for modeling

4. Model Building

- Logistic Regression with balanced class weights
- Train-test split (80-20)

5. Model Evaluation

- Accuracy: ~72%
- Precision: ~0.31
- Recall: ~0.62
- F1 Score: ~0.41

- ROC-AUC: ~0.78

6. Key Drivers of Attrition (Top Features)

- Overtime (Yes)
- Marital Status (Single)
- Environment Satisfaction
- Business Travel (Frequent)
- Stock Option Level
- Job Role (Sales Representative, Lab Technician)

7. Export & Results

- Prepared a scored dataset with predictions, probabilities, and actual attrition
- Results ready for visualization (Power BI/Tableau optional)

Final Outcome:

The project successfully built a predictive model for employee attrition, achieving strong recall and ROC-AUC performance.

Key business insights include the impact of overtime, job role, and work-life balance on attrition rates.

Status: ■ Project 1 Completed Successfully