

Employee Attrition Prediction – Internship Task 1 Report

1. Objective

The goal of this project is to build a classification model using HR data to predict whether an employee is likely to leave the company and to extract meaningful insights for employee retention strategies.

2. Dataset Description

- **Source:** IBM HR Analytics Dataset
- **Records:** 1,470 employees
- **Target Variable:** Attrition (Yes/No)
- **Features:** Includes demographics, job satisfaction, travel, income, experience, etc.

3. Exploratory Data Analysis (EDA)

- 16% of employees in the dataset left the company.
- Overtime, lower income, low satisfaction ratings, and early-career employees showed higher attrition.
- Visualizations included distribution plots, boxplots, and correlation heatmaps.

4. Data Preprocessing

- Categorical features were converted using Label Encoding and One-Hot Encoding.
- Removed redundant columns such as EmployeeNumber, StandardHours, etc.
- Stratified train-test split used to preserve target class distribution.

5. Model Building

Two models were used:

- **Random Forest Classifier** – Performed best in terms of F1 score and interpretability.
- **Logistic Regression** – Served as a baseline model.

Evaluation metrics included:

- Accuracy
- Precision
- Recall

- F1 Score
- Confusion Matrix

6. Explainability with SHAP

SHAP (SHapley Additive exPlanations) was used to interpret model predictions. Key contributing features to employee attrition were:

- OverTime
- Monthly Income
- Job Role
- Environment Satisfaction
- Age
- YearsWithCurrentManager

SHAP bar plots and summary plots clearly highlighted the importance of these features.

7. Key Insights

- Reducing employee overtime could directly reduce attrition.
- Increasing employee satisfaction and providing fair compensation are crucial.
- Employee retention policies should be tailored based on department and job role.
- Young, new hires require stronger onboarding and early-stage engagement.

8. Challenges Faced

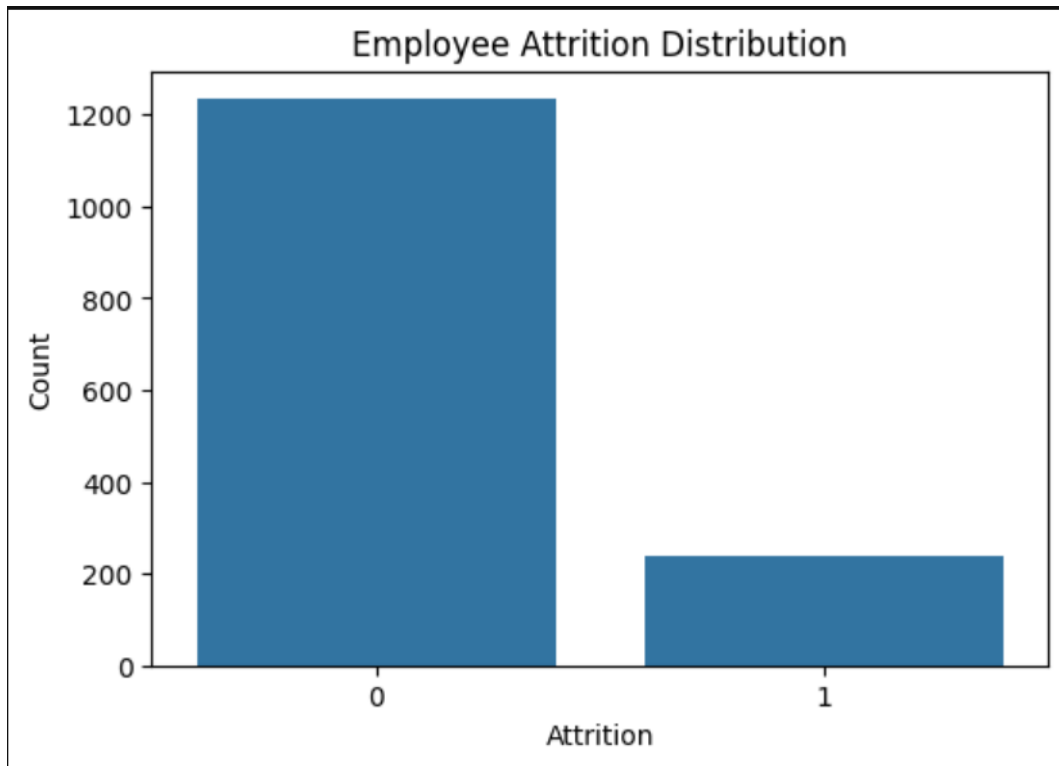
- Imbalanced target classes (16% attrition vs. 84% retention)
- Complex categorical features requiring careful preprocessing
- Interpretation of model outputs for non-technical stakeholders

9. Conclusion

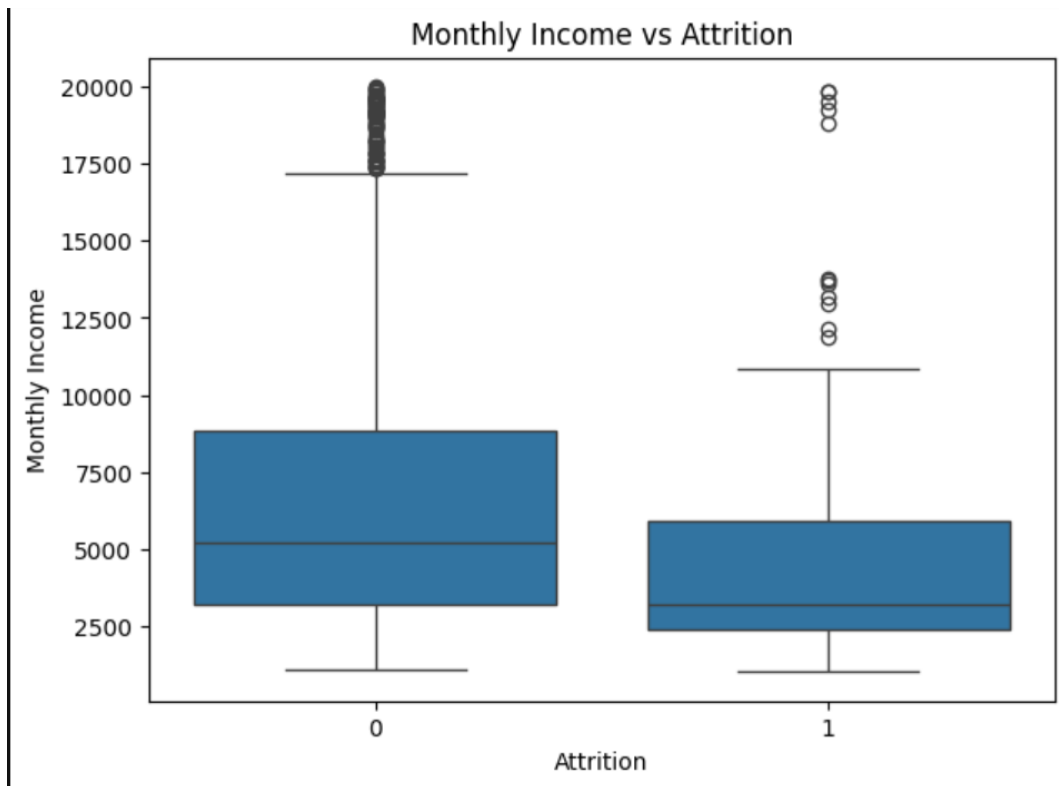
The project successfully predicted employee attrition and identified critical features influencing it. The model can help HR departments proactively identify at-risk employees and take necessary actions to improve retention.

Graphs

1. Attrition Countplot



2. Monthly Income vs Attrition (Boxplot)



3. Confusion Matrix(Heatmap)

Confusion Matrix - Random Forest

