
HR Analytics – Predict Employee Attrition

Student Name: Lalithya B C

College Name: ACU(BGSIT)

Internship Domain: Data Analytics

1. Introduction

The goal of this project is to analyze HR data and predict which employees are likely to leave the company (attrition). Understanding key factors behind attrition helps the organization improve retention strategies and employee satisfaction by using statistical and machine learning. Employee attrition affects productivity, morale, and overall business performance. Predicting attrition helps management make proactive decisions to retain skilled employees. Data analytics offers insights that traditional HR methods cannot easily identify. The study aims to bridge the gap between HR management and data-driven decision-making.

2. Abstract

This project uses a dataset from IBM HR Analytics, which includes employee details like job role, salary, and satisfaction. Data were cleaned, encoded, and analyzed using Python. Two models – Logistic Regression and Decision Tree – were trained to predict employee attrition. The Logistic Regression model achieved an accuracy of 87.8 %, while Decision Tree achieved 79.9 %. Insights were visualized using Seaborn and Matplotlib and SHAP analysis was used to understand important factors. Results showed that OverTime, Monthly Income, and Job Satisfaction were the strongest predictors of employee turnover.

3. Tools Used

Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, SHAP), Google Colab for execution.

4. Steps Involved

1. Data cleaning and pre-processing
 2. Encoding categorical variables
 3. Exploratory Data Analysis (EDA)
-

-
4. Model building (Logistic Regression and Decision Tree)
 5. Model evaluation and comparison
 6. Feature importance using SHAP

5. Results & Comparison

Model	Accuracy	Key Notes
Logistic Regression	87.8 %	Best performance; interpretable results
Decision Tree	79.9 %	Captures non-linear relations, slightly lower accuracy

Key Insights:

- Employees with low salary and low job satisfaction are more likely to leave.
- Frequent overtime work and low tenure increase attrition risk.
- Improving engagement and revising compensation can reduce attrition.

6. Conclusion

- Logistic Regression performed better with 87.8 % accuracy, making it the most reliable model for predicting employee attrition.
 - Improving salary, reducing overtime, and increasing job satisfaction can significantly lower employee turnover.
 - High overtime and low monthly income contribute significantly to employee attrition.
 - Data analytics techniques helped uncover actionable HR insights from raw data.
 - The study improved practical knowledge in Python, EDA, and machine learning model evaluation.
 - The project enhanced practical skills in data preprocessing, visualization, and model evaluation using Python.
 - The insights from this analysis can help HR teams make data-driven decisions to improve employee retention and workplace culture.
-