

Employee Attrition Analysis and Prediction

Abstract

This project focuses on analyzing employee data to identify attrition patterns and build a machine learning model to predict employees at risk of leaving. The solution helps HR teams take proactive retention decisions.

Problem Statement

Employee attrition increases recruitment cost and impacts productivity. Predicting attrition in advance enables organizations to reduce turnover and improve employee satisfaction.

Dataset Description

The dataset consists of 35 employee-related features including demographics, job role, salary, satisfaction, performance, and tenure details. Attrition is the target variable.

EDA – Key Observations

- Employees with low job satisfaction show higher attrition
- Overtime employees have increased turnover
- Lower income groups exhibit higher attrition
- Early-career employees are more likely to leave

EDA Visualizations

Charts Used:

1. Attrition vs Job Satisfaction (Bar Chart)
2. Attrition vs Monthly Income (Box Plot)
3. Attrition by Department (Count Plot)
4. Attrition vs Work-Life Balance

(These charts visually represent key attrition drivers.)

Feature Engineering

Created tenure buckets, encoded categorical variables, removed constant features, and selected high-impact attributes based on correlation analysis.

Model Development

Logistic Regression and Random Forest models were trained. Random Forest performed better in terms of F1-score and AUC-ROC.

Model Evaluation

Accuracy: ~86%

Precision, Recall, F1-Score used to handle class imbalance

AUC-ROC curve validated strong classification performance

Streamlit Dashboard

The Streamlit application provides interactive dashboards with attrition insights, employee-level predictions, and probability-based risk classification.

Business Impact

- Reduced attrition risk
- Cost savings in hiring & training
- Improved workforce planning

Conclusion

The project successfully predicts employee attrition and provides actionable HR insights.