

Employee Attrition Prediction Using Machine Learning

Karthik Kumar Cheekati (2024511)



Problem Definition:

The project analyzes the use of machine learning to predict employee attrition to help organizations proactively retain employees. Based on an analysis of past HR information like age, education, job tenure, city, and bench time, predictive models were developed to classify at-risk employees. The study emphasizes the possibility of evidence-based insights to guide employee engagement and retention in modern organizations.

Introduction

Attrition is costly and disruptive. Traditional measures like exit interviews are narrow-gauge and reactive. Here, machine learning is applied to predict attrition based on demographic and professional features for proactive intervention. The goal: forecast attrition prior to happening and inform HR policy to improve organizational stability.

Methodology

A structured data science pipeline was followed:

Data Source: Public HR dataset

Preprocessing: Imputation, encoding, and scaling

EDA: Visualizations to explore trends in attrition

Modeling: Logistic Regression, Random forest, XGBoost **Evaluation:** Metrics including Accuracy, F1 Score, AUC

Tools: Python, Scikit-learn, XGBoost, Seaborn

Conclusions and Recommendations

Machine learning enables precise attrition forecasting and empowers HR teams to take preemptive measures.

Recommendations include:

- Personalized engagement strategies
- Minimizing bench periods
- Career development planning
- City-specific retention policies
- Gender-balanced retention initiatives



Results

Metric	Logistic Regression	Random Forest	XGBoost
Accuracy	0.63	0.78	0.81
Precision	0.60	0.85	0.84
Recall	0.24	0.59	0.63
F1 Score	0.34	0.69	0.71
ROC AUC	0.63	0.84	0.85

Top Features: Experience in current domain, Age, City, EverBenchd

XGBoost performed well

Insight: Employees who were benchd, from Pune, or with lower experience are at higher risk of attrition.

The XGBoost model proved to be most accurate when being most interpretable. The model's findings affirm that attrition is governed by a mix of organizational and personal issues. City trends reflect regional differences in labor market dynamics, whereas benching is linked to higher exit rates possibly due to job insecurity perceptions.

References:

IBM, 2020. HR Analytics for Workforce Retention. IBM Research Reports.

Microsoft, 2021. SMOTE for imbalanced classification. Available at: <https://docs.microsoft.com>