

HR Analytics – Predicting Employee Attrition

Introduction

Employee attrition is one of the toughest challenges for companies. When employees leave, organizations face higher recruitment costs, a loss of knowledge, and reduced productivity. This project explores how HR analytics can be used to understand the main drivers of attrition and predict which employees are more likely to resign. By doing so, HR teams can take preventive actions and improve employee retention.

Abstract

We used the IBM HR Analytics Attrition dataset with information about 1,470 employees. Through exploratory data analysis (EDA), we studied how attrition is related to age, salary, department, and job roles. We then trained machine learning models – Logistic Regression, Decision Tree, Logistic Regression with SMOTE, and Random Forest – to predict whether an employee will leave. Finally, we built an interactive dashboard in Power BI to present key attrition insights in a clear and practical way.

Tools Used

• Python (Pandas, Seaborn, Scikit-learn, Imbalanced-learn) • Power BI (for dashboard and visualization)

Steps Involved

1. Data Preprocessing – Encoded categorical variables and split dataset into training and test sets. 2. Exploratory Data Analysis – Visualized attrition trends by department, job role, and age. 3. Model Building – Applied Logistic Regression, Decision Tree, SMOTE Logistic Regression, and Random Forest. 4. Model Evaluation – Compared accuracy, recall, and confusion matrices. 5. Dashboard Creation – Built interactive Power BI dashboard with filters and charts.

Results

Model	Accuracy	Recall (Attrition=Yes)	Comment
Logistic Regression	86%	13%	High accuracy, weak at predicting leavers
Decision Tree	76%	44%	Lower accuracy, better recall
Logistic Regression + SMOTE	74%	44%	Balanced, but accuracy dropped
Random Forest	87%	10%	Best accuracy, still biased towards 'stay'

Insights & Dashboard Findings

- Employees who frequently work overtime are more likely to leave.
- Younger employees, especially those under 30, show higher attrition rates.
- Sales representatives and low-income groups have the highest resignation levels.
- Job satisfaction and work-life balance are major factors influencing attrition.

Conclusion

Logistic Regression achieved the highest accuracy (86%), but failed to detect most employees who left. Decision Tree and Logistic Regression with SMOTE improved recall, making them more practical for HR decision-making. Random Forest reached 87% accuracy and highlighted important features such as Overtime, Age, JobSatisfaction, and MonthlyIncome. For HR, recall is more important than overall accuracy, so Decision Tree is recommended for identifying employees at risk of leaving. Combining predictive models with Power BI dashboards provides a strong decision-support system for HR teams.