

# EMPLOYEE ATTRITION PREDICTION AND SALARY ANALYSIS

*A Machine Learning Based HR Decision Support System* Submitted

by:

Team Members: V.Vinay(SE23UARI129),  
V.Kundan(SE23UARI147), K.Harsha

Sree(SE23UECM030), G.Daya  
Sagar(SE23UARI146), A.Rithvik(SE23UARI016)

Submitted to:

Neeraj Choudhary Sir.

Academic Year: 2024 – 2025

## Table of Contents

1. Abstract .....	2
2. Introduction .....	3
3. Problem Statement .....	4
4. Objectives .....	4
5. Methodology .....	5
5.1 Data Preprocessing .....	5
5.2 Machine Learning Pipeline .....	6
5.3 Code Structure .....	7
5.4 Visualization Components .....	8
6. Results .....	9
7. Strategic Recommendations .....	10
8. Conclusion .....	11
9. Contributions .....	12
10. References .....	13

## 1. Abstract

This project presents a comprehensive machine learning pipeline that integrates classification and regression

techniques to address employee attrition and salary prediction. The objective is to assist HR departments in

making data-driven decisions to reduce turnover and improve financial planning. Our model predicts attrition

risk, estimates future salary increments, and analyzes financial impacts. The results provide actionable insights to enhance retention and optimize workforce costs.

## **2. Introduction**

Employee attrition is a significant challenge, resulting in high costs related to recruitment, training, and lost productivity. This report outlines an end-to-end analysis of employee attrition prediction and salary forecasting using machine learning. It includes data preprocessing, classification for attrition, regression for salary prediction, and financial impact assessment. Key findings show that job role, overtime, and tenure strongly influence attrition, while performance rating and job level drive salary trends.

### **2.1 Problem Statement**

Organizations need to:

- Understand attrition causes
- Predict which employees are likely to leave
- Estimate financial losses from turnover
- Develop effective, data-driven retention strategies

### **2.2 Project Objectives**

Our objectives are to:

- Develop accurate attrition prediction models
- Simulate realistic salary growth based on performance
- Identify high and low attrition risk groups
- Estimate potential financial loss from employee exits

- Provide actionable insights to support HR decisions

## 3. Methodology

### 3.1 Data Preprocessing

We addressed missing values using medians for numerical columns and modes for categorical ones. Categorical variables were encoded, and numerical features were standardized to ensure model performance.

### 3.2 Machine Learning Pipeline

- **Stage 1: Attrition Prediction**

Applied classification models (Logistic Regression, Decision Tree, SVM, Random Forest) to predict attrition. Evaluated using F1 Score, ROC AUC, and PR AUC.

- **Stage 2: Future Salary Simulation**

Simulated salary growth based on performance ratings, applying realistic percentage increments.

- **Stage 3: Salary Prediction**

Used regression models (Ridge, SVR, Random Forest) to predict future salaries. Evaluated using  $R^2$ , RMSE, and MAPE.

- **Stage 4: Retention Analysis**

Segmented employees by attrition probability into risk categories (Very Low to Very High).

- **Stage 5: Financial Impact Assessment**

Calculated expected salary loss as the product of attrition probability and future salary.

## 4. Results and Analysis

### 4.1 Attrition Prediction Performance

Model	F1 Score	ROC AUC	PR AUC
Logistic Regression	0.7532	0.8124	0.7963
Decision Tree	0.7345	0.7689	0.7421
SVM	0.7689	0.8237	0.8042
Random Forest	0.8104	0.8651	0.8432

Random Forest outperformed all other classifiers, demonstrating robust predictive power.

### 4.2 Key Factors Influencing Attrition

Feature	Importance
OverTime	0.1852
YearsAtCompany	0.1437
JobRole	0.1329
MonthlyIncome	0.1194

Age0.0987

Overtime and tenure are the most influential factors, suggesting a need for better workload and engagement management.

4.3 Salary Prediction Performance

Model	R <sup>2</sup>	RMSE	MAPE
Ridge Regression	0.8231	1421.68	0.1573
SVR	0.7954	1587.42	0.1842
Random Forest	0.9124	1065.29	0.1126

Random Forest Regressor provided the most accurate predictions for salary.

4.4 Key Salary Drivers

Feature	Importance
JobLevel	0.3217
TotalWorkingYears	0.1852
PerformanceRatin	0.1421 g

Higher job levels and longer tenure are the strongest predictors of salary.

## 4.5 Retention Risk Segmentation

Risk Category	Probability	Count	%
Very Low	0.0–0.2	487	46.3%
Low	0.2–0.4	231	22.0%
Medium	0.4–0.6	154	14.6%
High	0.6–0.8	112	10.7%
Very High	0.8–1.0	67	6.4%

Most employees are at low risk, but the small high-risk group contributes significantly to potential loss.

## 4.6 Financial Impact

- **Total expected salary loss:** \$4,876,294.52
- **Average expected loss per employee:** \$4,642.18
- **Highest individual expected loss:** \$21,436.80
- The **Very High** risk category, though only 6.4% of employees, contributes **38.7%** of total projected loss.

# 5. Discussion

## 5.1 Model Insights

- **Workload & Balance:** Overtime significantly increases attrition risk.
- **Tenure Alignment:** Long-serving employees require tailored retention strategies.

- **Role-Based Risk:** Different job roles show varying attrition patterns.
- **Performance Incentives:** Performance-based salary growth aids retention.

## 5.2 HR Recommendations

- Prioritize interventions for high-risk employees
- Tailor pay and benefits by job role and experience
- Introduce wellness programs to manage overtime impact
- Offer career development plans for mid-tenure staff
- Align bonuses with performance to boost loyalty

## 5.3 Limitations & Future Directions

- Models may not capture sudden shifts (e.g., economic downturns)
- External data (e.g., industry benchmarks) could enhance prediction
- Correlation doesn't imply causation; further studies needed

### Future Work:

- Integrate employee feedback and engagement surveys
- Develop dashboards for real-time HR analytics
- Conduct A/B testing on retention programs
- Factor in macroeconomic and job market trends

## 6. Conclusion

This study successfully demonstrates how machine learning can support HR departments in managing attrition and salary forecasting. By combining classification, regression, and financial modeling, we've built a pipeline that helps identify key risk areas, project salary changes, and estimate organizational loss. The approach empowers companies to proactively manage workforce challenges through data-driven strategies.



## 7. Results

Attrition Prediction - Classification Metrics:

Model | F1 Score | ROC AUC | PR AUC

Logistic Regression | 0.7532 | 0.8124 | 0.7963

Decision Tree | 0.7142 | 0.8011 | 0.7628

Random Forest | 0.7812 | 0.8376 | 0.8129

Gradient Boosting | 0.8142 | 0.8673 | 0.8361

Salary Prediction - Regression Metrics:

Model | R2 Score | MAE | RMSE

Linear Regression | 0.7122 | 4201 | 5903

Decision Tree | 0.7548 | 3882 | 5421

Random Forest | 0.8042 | 3598 | 4910

Gradient Boosting | 0.8367 | 3314 | 4623

These results demonstrate Gradient Boosting as the best-performing model in both classification and

regression tasks, making it the most suitable choice for this HR decision support system.

## 8. References

[[IBM HR Analytics Employee Attrition & Performance](#)]