

Employee Attrition and Salary Loss Prediction: Final Project Report

Employee Attrition and Salary Loss Prediction: Final Project Report

1. Objective

The goal of this project was to:

- Predict employee attrition (whether an employee is likely to leave).
- Predict future salaries for employees based on profile features.
- Calculate expected salary loss considering the attrition risk.

2. Data Cleaning and Preprocessing

- Loaded the employee dataset containing features like Performance Rating, Total Working Years, Job Level, OverTime, Marital Status, etc.
- Handled missing values appropriately (if any were present).
- Encoded categorical features using LabelEncoder.
- Standardized numerical features using StandardScaler wherever necessary (especially for Ridge Regression and Logistic Regression).

3. Attrition Classification Models

Steps Taken:

- Built multiple classification models to predict whether an employee is likely to stay or leave.
- Explored different combinations:
 - Without scaling and without sampling.
 - With scaling but without sampling.
 - With scaling and RandomOverSampler (ROS) sampling.
 - With scaling and SMOTE sampling.
- Also trained Random Forest classifier with SMOTE sampling.

Employee Attrition and Salary Loss Prediction: Final Project Report

Results:

- Without scaling, no sampling: 87.75% accuracy (poor recall for attrition)
- With scaling, no sampling: 89.79% accuracy (better recall)
- With scaling and ROS sampling: 70.65% accuracy (balanced but lower)
- With scaling and SMOTE sampling: 75.91% accuracy (improved balance)
- Random Forest + SMOTE: 93.11% accuracy (best overall model)

Conclusion:

Random Forest with SMOTE sampling was selected as the best attrition model.

4. Future Salary Regression Models

Steps Taken:

- Built Ridge Regression (linear) and Random Forest Regressor (nonlinear ensemble)

Training Results:

- Ridge Regression: $R^2=0.9997$, RMSE=83.33, MAPE=1.01%
- Random Forest Regressor: $R^2=0.9999$, RMSE=44.11, MAPE=0.27%

Testing Results:

- Ridge Regression: $R^2=0.9994$, RMSE=119.33, MAPE=0.74%
- Random Forest Regressor: $R^2=0.9994$, RMSE=121.79, MAPE=0.75%

Conclusion:

Both models predicted future salaries with extremely high accuracy.

Employee Attrition and Salary Loss Prediction: Final Project Report

5. Combining Attrition Prediction and Salary Estimation

Steps Taken:

- Predicted attrition probabilities.
- Defined threshold ($P_{\text{stay}} > 0.6$).
- Predicted future salary only for "Likely to Stay" employees.

6. Expected Salary Loss Calculation

Formula:

$$\text{Expected Loss}_i = P(\text{leave})_i \times \text{FutureSalary}_i$$

$$\text{Total Expected Loss} = \text{Sum}(\text{Expected Loss}_i)$$

Final Result:

Total Expected Loss (for "Likely to Stay" employees): INR 860,099.00

Interpretation:

Even among employees classified as "likely to stay", there is hidden attrition risk amounting to INR 860,099.

7. Key Insights

- Random Forest + SMOTE provides robust attrition prediction.
- Ridge and Random Forest Regressor models excel at salary prediction.
- Combining attrition probabilities with salary projections enables realistic financial risk analysis.

8. Final Conclusion

- Data preprocessing was handled carefully.
- Multiple models were built and evaluated systematically.

Employee Attrition and Salary Loss Prediction: Final Project Report

- Financial risk analysis based on ML outputs provides valuable insights.

End of Report