

HR Analytics Report on Employee Attrition

Problem Statement:

Employee attrition is a critical challenge faced by organizations worldwide. High attrition rates can lead to increased recruitment costs, loss of organizational knowledge, and disruptions in workflow. Identifying the key factors contributing to employee attrition is essential for developing strategies to retain valuable talent.

Objective:

This report aims to:

1. Analyse employee attrition trends using key organizational metrics such as job satisfaction, work-life balance, and departmental affiliation.
2. Utilize machine learning (Logistic Regression) to determine the most influential factors contributing to employee attrition.
3. Provide actionable insights to help the organization improve employee retention.

Dashboard Insights

The visualized data reveals the following trends:

1. Attrition Rate by Job Involvement:

- Employees with the lowest job involvement (1) show the highest attrition rate (34%), indicating a strong correlation between engagement and retention.

2. Attrition Rate by Job Satisfaction:

- Employees with job satisfaction level 1 have the highest attrition rate (23%), highlighting dissatisfaction as a significant risk factor.

3. Attrition Rate by Department:

- Sales department employees experience the highest attrition rate (21%), followed by Human Resources (19%). Research & Development shows a comparatively lower rate (14%).

4. Attrition Rate by Work-Life Balance:

- Employees with a poor work-life balance (1) show the highest attrition rate (31%), emphasizing the importance of flexibility and personal well-being.

5. Attrition Rate by Over Time:

- Employees working overtime are at a higher risk of attrition (31%) compared to those who do not (10%).

6. Attrition Rate by Education Field:

- Human Resources and Technical Degree holders exhibit the highest attrition rates (26% and 24%, respectively), potentially signalling misalignment between expectations and job roles.

7. Attrition Rate by Environment Satisfaction:

- Lower environment satisfaction (1) correlates with a high attrition rate (25%), underlining the significance of a positive work environment.

8. Attrition Rate by Job Role:

- Sales Representatives show the highest attrition rate (40%), suggesting a need for focused retention strategies in this role.

9. Attrition Rate by Marital Status:

- Single employees have the highest attrition rate (26%), while divorced employees have the lowest (12%), indicating potential lifestyle influences on retention.

Machine Learning Analysis on Employee Attrition

Summary:

We developed a logistic regression model to predict employee attrition probability. After evaluating multiple thresholds, we selected 0.575 as the operating point for classification because it balances precision (48.28%) and recall (53.16%) with an overall accuracy of 83.27% and F1-score of 50.60%.

This balance ensures HR is not overwhelmed by false positives while still identifying a meaningful share of at-risk employees for proactive retention actions.

Context & Objective:

1. Objective: Identify employees at risk of attrition to enable timely, data-driven retention interventions.
2. Business Impact: Reduces the cost of replacement hiring and the loss of institutional knowledge by enabling targeted outreach.

Data & Feature Selection:

1. Initial variable screening via correlation ($|r| \geq 0.10$) against Attrition, excluding weakly correlated features.
2. Reduced multicollinearity by dropping highly correlated pairs (e.g., JobLevel \leftrightarrow MonthlyIncome; YearsAtCompany, YearsInCurrentRole, YearsWithCurrentManager).
3. Post-fit significance pruning: Dropped TotalWorkingYears and StockOptionLevel (p-value > 0.05).

Modeling Approach:

1. Algorithm: Logistic Regression.
2. Evaluation: Confusion-matrix–derived metrics across thresholds (Accuracy, Precision, Recall, Specificity, F1).
3. Threshold Selection: 0.575 chosen for a balanced trade-off between Precision and Recall.

Performance Across Thresholds:

Threshold	Accuracy	Precision	Recall	Specificity	F1_Score
0.50 (initial)	0.2789	0.1773	0.9536	0.1492	0.2989
0.60 (initial)	0.8626	0.6768	0.2827	0.9740	0.3988
0.60 (after drop: Total Working Years)	0.8619	0.6700	0.2827	0.9732	0.3976
0.60 (after drop: StockOptionLevel)	0.8605	0.66	0.2785	0.9724	0.3917
0.55	0.6952	0.3171	0.7721	0.6804	0.4496
0.57	0.8177	0.4485	0.5696	0.8654	0.5019
0.575 (selected)	0.8327	0.4828	0.5316	0.8905	0.5060
0.58	0.8469	0.5283	0.4726	0.9189	0.4989

Selected Threshold: 0.575

At threshold 0.575:

- Accuracy: 83.27%
- Precision: 48.28%
- Recall: 53.16%
- Specificity: 89.05%
- F1- Score: 50.60%

Precision and Recall are closely matched, minimizing bias toward only one class and keeping both false positives and false negatives at manageable levels.

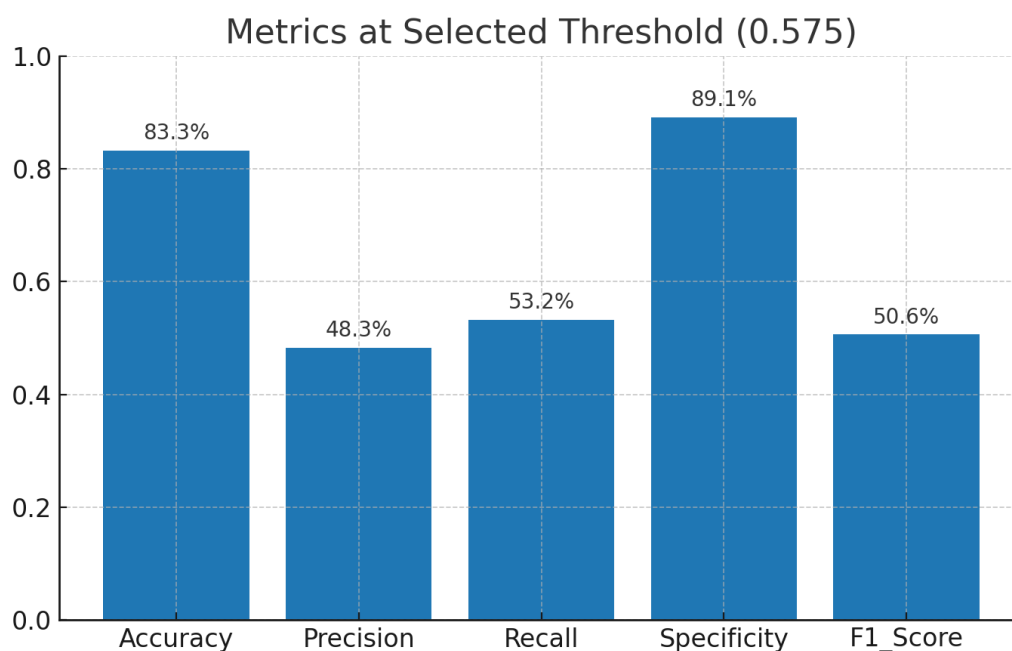


Figure: Metrics at selected threshold 0.575

Prediction of Employee Attrition Probability:

Each employee receives a probability score indicating their likelihood of leaving. Employees with a score above **57.5%** are flagged as “at risk,” based on the optimized threshold of 0.575. This ensures HR can proactively intervene with tailored engagement plans, career growth opportunities, and work-life balance improvements, while avoiding excessive false alarms.

Business Interpretation:

The logistic regression model assigns each employee a probability score between 0 and 1, representing their likelihood of resigning. By applying the optimized threshold of 0.575, employees with a predicted probability of 57.5% or higher are classified as “*at risk of attrition*.”

1. This threshold ensures a balanced trade-off between precision (48.28%) and recall (53.16%).
2. HR will correctly identify more than half of the employees who are likely to leave, while avoiding excessive false positives that could overwhelm retention resources.
3. The model’s specificity of 89.05% confirms that most employees predicted to stay are correctly classified, minimizing unnecessary interventions.

Practical Use for HR:

1. High-probability employees (≥ 0.575) can be prioritized for retention programs such as career growth discussions, workload adjustments, and flexible arrangements.
2. Medium-risk employees (0.40–0.575) can be placed on a watchlist, with periodic check-ins and engagement activities.
3. Low-risk employees (< 0.40) can continue under normal HR practices, with no additional intervention required.

This structured, threshold-driven approach allows HR leaders to take targeted, proactive action — reducing attrition costs, improving employee satisfaction, and safeguarding organizational knowledge.

Final Thoughts:

This predictive model equips HR with actionable intelligence to intervene early and retain valuable talent. The insights gained here should be integrated into ongoing retention strategies, ensuring long-term improvements in employee satisfaction and organizational stability.