

HR Analytics for Employee Attrition Prediction

Akash M A (Student)

MBA – Rathinam Technical Campus, Coimbatore


Mr. Thanush N

(Assistant Professor)



<https://doi.org/10.55041/ijstmt.v2i5.407>

Cite this Article: A, A. M. (2026). HR Analytics for Employee Attrition Prediction. International Journal of Science, Strategic Management and Technology, 02(05). <https://doi.org/10.55041/ijstmt.v2i5.407>

License:  This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

Abstract

Employee attrition has become one of the major challenges faced by organizations because it affects productivity, increases recruitment cost, and reduces organizational performance. The present study focuses on identifying the major factors responsible for employee attrition and predicting employee turnover using HR Analytics and machine learning techniques. Both primary and secondary data were used for analysis. Primary data were collected from 100 employees through a structured questionnaire, while secondary data were obtained from the IBM HR Analytics Employee Attrition Dataset containing 1,470 employee records. Variables such as age, education, department, salary, overtime, work-life balance, and job satisfaction were analysed.

Statistical tools including percentage analysis and correlation analysis were used for interpretation. Machine learning techniques such as Logistic Regression and Decision Tree were applied to predict employee attrition. The findings revealed that younger employees and employees working in sales departments showed higher attrition tendency. Overtime, low salary, work pressure, and poor work-life balance were identified as major factors influencing employee turnover. Among the prediction models, Logistic Regression produced higher accuracy compared to Decision Tree. The study concludes that organizations should improve employee engagement, career growth opportunities, salary structure, and work-life balance strategies to reduce employee attrition and improve retention.

Keywords: HR Analytics, Employee Attrition, Employee Retention, Logistic Regression, Decision Tree, Machine Learning.

1. Introduction

Employee attrition refers to the process in which employees leave an organization due to resignation, retirement, termination, or personal reasons. In the modern business environment, employee attrition has become a major concern because it directly affects organizational productivity, employee morale, and profitability. High employee turnover results in recruitment and training costs and also causes loss of experienced employees.

Organizations are increasingly using HR Analytics to understand employee behaviour and improve employee retention strategies. HR Analytics helps organizations analyse employee-related data and identify employees who are more likely to leave the organization. It supports management in making better human resource decisions.

The present study focuses on analysing the major factors influencing employee attrition and predicting employee turnover using machine learning techniques. The study considers factors such as salary, overtime, department, work-life balance, and job satisfaction. Logistic Regression and Decision Tree models were used to predict employee attrition.

The study is important because retaining skilled employees has become a major competitive advantage for organizations. Effective employee retention strategies can improve organizational stability and long-term success.

2. Review of Literature

Several researchers have conducted studies on employee attrition and HR Analytics. Previous studies indicate that salary, overtime, stress, job dissatisfaction, and poor work-life balance are major reasons for employee turnover.

Research studies also reveal that younger employees are more likely to leave organizations because they seek better career growth and salary opportunities. Employees working in sales and target-oriented departments often experience high work pressure, leading to higher attrition rates.

Many earlier studies mainly focused on descriptive and statistical analysis. Only a limited number of studies applied machine learning techniques such as Logistic Regression and Decision Tree for employee attrition prediction.

However, research gaps still exist because many studies considered only limited variables and did not combine both primary and secondary data. Therefore, the present study attempts to analyse employee attrition using survey data along with IBM HR Analytics Dataset and machine learning techniques.

3. Objectives of the Study

The major objectives of the study are:

1. To study the concept of employee attrition and HR Analytics.
2. To identify the major factors responsible for employee attrition.
3. To analyse the relationship between salary, overtime, and employee attrition.
4. To study the effect of work-life balance and job satisfaction on employee retention.
5. To apply machine learning techniques for employee attrition prediction.
6. To compare Logistic Regression and Decision Tree models.
7. To provide suitable suggestions for reducing employee turnover.

4. Research Methodology

The study adopted descriptive, analytical, and predictive research design. Both primary and secondary data were used for analysis.

4.1 Primary Data

Primary data were collected from 100 employees using a structured questionnaire. The questionnaire included questions related to salary, overtime, work pressure, promotion opportunities, work-life balance, and job satisfaction. Convenience sampling method was used for selecting respondents.

4.2 Secondary Data

Secondary data were obtained from the IBM HR Analytics Employee Attrition Dataset containing 1,470 employee records and various employee-related variables such as:

- Age
- Gender
- Education
- Department

- Monthly Income
- Overtime
- Job Satisfaction
- Work-Life Balance
- Attrition Status

4.3 Tools Used for Analysis

The following tools and techniques were used:

- Percentage Analysis
- Correlation Analysis
- Logistic Regression
- Decision Tree
- Microsoft Excel
- Python

4.4 Machine Learning Models Logistic Regression

Logistic Regression is a machine learning model used when the dependent variable contains binary outcomes such as Attrition = Yes or No.

Decision Tree

Decision Tree classifies employees into different categories based on employee characteristics and identifies factors strongly influencing employee attrition.

5. Data Analysis and Interpretation

5.1 Attrition Rate Analysis

The study revealed that 30% of employees from the primary survey expressed intention to leave the organization within the next year. In the IBM dataset, the actual attrition rate was 16.12%. This indicates that employee turnover intention is comparatively high and organizations should focus on retention strategies.

5.2 Age and Attrition

Employees between the age group of 25–30 years showed the highest attrition tendency. Younger employees often leave organizations in search of better salary, promotion opportunities, and career growth. Employees above 40 years showed lower attrition because of greater job stability.

Department and Attrition

The Sales department recorded the highest attrition rate compared to other departments. High work pressure, targets, and stress were identified as major reasons for employee turnover in sales-related jobs.

5.3 Correlation Analysis

Correlation analysis revealed that overtime and work pressure had a positive relationship with attrition, while salary, promotion opportunities, and work-life balance had a negative relationship with attrition.

This means employees experiencing high stress and overtime are more likely to leave the organization, whereas employees receiving better salary and career growth opportunities are more likely to stay.

6. Model Performance and Prediction

Machine learning models were developed using Python for employee attrition prediction. The performance of Logistic Regression and Decision Tree models was compared.

Model	Accuracy
Logistic Regression	87.07%
Decision Tree	73.47%

The results indicate that Logistic Regression produced better prediction accuracy compared to Decision Tree. Therefore, Logistic Regression was considered the most suitable model for employee attrition prediction in the present study.

7. Findings of the Study

The major findings of the study are:

- Younger employees showed higher attrition tendency.
- Employees in sales departments experienced greater turnover intention.
- Overtime and work pressure significantly influenced employee attrition.
- Poor work-life balance increased employee dissatisfaction.
- Employees with low salary and limited promotion opportunities were more likely to leave the organization.
- Logistic Regression model produced better prediction accuracy than Decision Tree.
- HR Analytics helps organizations identify employees who are at risk of leaving.

8. Suggestions

Based on the findings, the following suggestions are provided:

1. Organizations should improve salary and incentive structures.
2. Flexible working arrangements should be introduced to improve work-life balance.
3. Promotion and career growth opportunities should be increased.
4. Employee engagement activities should be conducted regularly.
5. Work pressure and overtime should be reduced.
6. HR Analytics should be implemented for early identification of dissatisfied employees.
7. Organizations should provide counselling and employee support programs.

9. Conclusion

Employee attrition is one of the major challenges faced by modern organizations because it affects productivity, employee morale, and organizational performance. The present study identified the major factors influencing employee attrition using both primary and secondary data.

The analysis revealed that younger employees, overtime, work pressure, poor salary, and lack of promotion opportunities significantly influence employee turnover. Machine learning techniques such as Logistic Regression and Decision Tree were used for prediction, and Logistic Regression produced higher accuracy.

The study concludes that HR Analytics plays an important role in improving employee retention and organizational

decision-making. Organizations that effectively use employee data and predictive analytics can reduce attrition, improve employee satisfaction, and achieve long-term organizational success.

References

1. IBM HR Analytics Employee Attrition Dataset.
2. Dessler, G. (2021). Human Resource Management. Pearson Education.
3. Kothari, C.R. (2019). Research Methodology: Methods and Techniques.
4. Robbins, S.P., & Judge, T.A. (2020). Organizational Behaviour.
5. Armstrong, M. (2021). Handbook of Human Resource Management Practice.
6. Pedregosa, F. et al. (2011). “Scikit-learn: Machine Learning in Python.”
7. Sharma, P., & Verma, R. (2020). “Impact of Work-Life Balance on Employee