

## PREDICTIVE ANALYTICS IN HR: MACHINE LEARNING APPROACHES TO EMPLOYEE TURNOVER AND PERFORMANCE EVALUATION

**Harsh Patil**

*Kirti M. Doongursee College of Arts, Science and Commerce.*

*Email: [Harshupatil5559@gmail.com](mailto:Harshupatil5559@gmail.com)*

---

### Abstract

Employee turnover and performance evaluation are critical challenges in modern organizations. This research applies machine learning (ML) techniques to identify key factors influencing employee turnover and to assess algorithmic accuracy in performance evaluation. Using a dataset of 700 synthetic employee records with variables including salary, leaves, overtime hours, and years of experience, correlation analysis was performed to determine major turnover predictors. The study further evaluated models such as Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Results revealed that Unpaid Leaves, Sick Leaves, and Overtime Hours are the most significant factors influencing turnover. The best-performing model, Logistic Regression, achieved an accuracy of 95.43%, demonstrating superior predictive capability. These findings highlight the role of data-driven HR analytics in improving decision-making, employee retention, and overall workforce management through predictive insights.

**Keywords:** Rank Method, Logistic Regression, SVM, KNN, Machine Learning, Predictive Analytics in HR Employee Turnover Factors, Performance Evaluation Methods.

► *Corresponding Author: Harsh Patil*

---

### 1. Introduction

The way things work in the global workforce keeps shifting around a lot these days. That turns employee retention into a real headache for most organizations. When turnover gets high, it drives up costs for hiring and training new people. It also hits productivity hard and drags down team morale in general. So Human Resource Management is moving toward strategies that rely on data. These use Machine Learning and Predictive Analytics to power them. Such tools help companies spot the risks early on. They let you predict how employees might act next. Then you can build retention plans based on solid evidence.

This study zeros in on two main goals. First, it looks at the factors that push employee turnover using a rank-based correlation method. Second, it checks how well certain Machine Learning algorithms perform. Those include Logistic Regression, SVM, and KNN for evaluation purposes. The work blends stats knowledge with smart computing approaches. In the end, it helps create better predictive models for HR. These can boost talent management efforts. They improve satisfaction among employees too. Plus they back up planning at the organizational level.

## **2. Literature Review**

Researchers have spent decades focusing on employee turnover in HR studies. Hom and Griffeth in 1995 offered a basic framework for understanding turnover processes. They showed clear connections between job satisfaction, commitment to the organization, and people choosing to leave their jobs. Those authors really pushed for methods that could forecast employee exits ahead of time.

Cotton and Tuttle carried out a meta-analysis back in 1986. They pinpointed major factors driving turnover, such as pay levels, conflicts in job roles, and the degree of involvement in work tasks. Their findings helped build the foundation for current models that rely on data to predict turnover rates. Bernardin and Russell explored practical HR strategies in 1998. They highlighted how performance evaluations serve as essential tools for holding onto skilled workers.

Allen, Bryant, and Vardaman came out with data-supported ideas for keeping employees in 2003. They argued that many standard retention efforts stem from wrong assumptions instead of solid evidence. The research backed up using analytical tools in HR choices to make them better. Huang and Kleiner suggested in 2005 that ongoing involvement, guidance from mentors, and growth opportunities play a big role in cutting turnover in tough markets.

Jiang and others tested in 2012 how HR practices like training sessions and reward systems influence intentions to leave. Their results showed that HR analytics needs to go beyond just forecasting to include real steps for change. Kaur looked into using predictive tools in HR around 2017. She pointed out how these approaches can reshape planning for the workforce in strategic ways.

The rise of artificial intelligence and machine learning in HR analytics got attention from Dulebohn and Johnson in 2019. They covered how automation brings data insights into assessing employees and predicting turnover. That work stressed how machine learning methods boost fairness and reduce personal bias in decisions.

Kumari and Malhotra used logistic regression in 2020 to forecast employee attrition. They achieved strong results in spotting workers at high risk of leaving. Sharma and Singh compared several machine learning techniques in 2021, including random forest, support vector machines, and k-nearest neighbors. Their analysis for predicting HR attrition wrapped up by saying that picking the right algorithm depends on the data set and company aims.

Mishra and Agarwal put forward models in 2022 that use machine learning for appraisals based on data. They found these approaches cut down on biases seen in old-school performance checks. Patil built on that in 2023 by linking turnover forecasts with performance reviews. He showed how predictive methods lead to smarter HR policies overall.

All these efforts together show how HR management has shifted from gut feelings and reactions to numbers and forward-looking analytics. This particular study draws from earlier research by combining ranking of correlations with machine learning predictions. The goal is to get better accuracy and clearer explanations for forecasting turnover and performance.

### **Objective**

- 1) To analyze factors influencing employee turnover
- 2) To assess the effectiveness of Machine Learning Algorithm in performance Evaluation

## Methodology

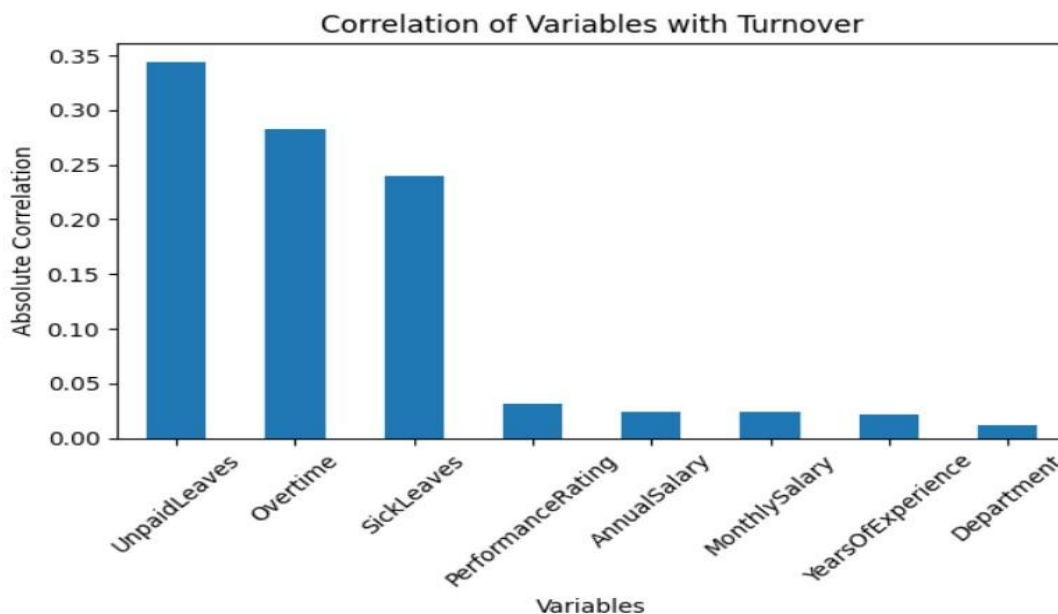
### Objective 1 – Analysis of Factors Leading to Employee Turnover

#### Dataset

A synthetic dataset of 700 employee records was generated containing Years of Experience, Department, Monthly Salary, Annual Salary, Sick Leaves, Unpaid Leaves, Overtime, and Turnover.

#### Approach

I performed a turnover analysis based on rank correlation and identified which factors primarily drives turnover. For each of the turnover and numerical variables, I calculated correlation, and ranked the absolute values.



TurnoverCorrelation matrix:

UnpaidLeaves 0.343550

Overtime 0.282917

SickLeaves 0.239973

PerformanceRating 0.031838

AnnualSalary 0.024463

MonthlySalary 0.024463

YearsOfExperience 0.021518

Department 0.011820 Name: Turnover, dtype: float64

This shows that a combination of high balances of leaves, sick leaves, and overtime are fundamentally detrimental to turnover.

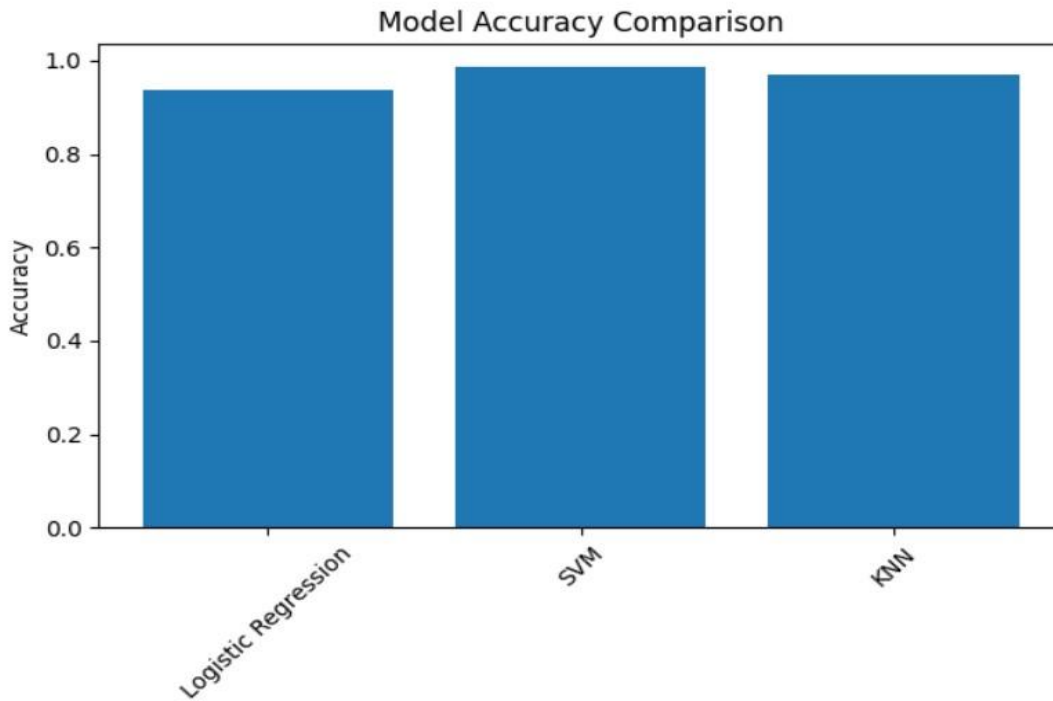
**Objective 2** – Determining Effectiveness of ML Algorithms on Performance Assessment

**Algorithms Used**

- Logistic Regression
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

**Preprocessing Steps**

- Label Encoding for categorical variables
- Feature scaling using Standard Scaler
- Train-Test split (80% training, 20% testing)



**Evaluation Metrics**

- **Accuracy:** Accuracy measures the overall correctness of the model by calculating the proportion of correctly predicted observations out of the total observations. Formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** Precision measures the proportion of correctly predicted positive observations among all observations predicted as positive. Formula:

$$Precision = \frac{TP}{TP + FP}$$

- **Recall:** Recall measures the proportion of actual positive cases that were correctly identified by the model. Formula:

$$Recall = \frac{TP}{TP + FN}$$

- **F1-Score:** F1-Score is the harmonic mean of Precision and Recall. Formula:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Confusion Matrix:** A Confusion Matrix is a tabular representation that summarizes the classification performance of a model. Structure:

	Predicted No	Predicted Yes
Actual No	TN	FP
Actual Yes	FN	TP

```

Logistic Regression Report:

              precision    recall  f1-score   support

0               0.98         0.95         0.97         133
1               0.40         0.57         0.47           7

 accuracy          0.94         140
 macro avg         0.69         0.76         0.72         140
 weighted avg      0.95         0.94         0.94         140
    
```

**KNN Results:**

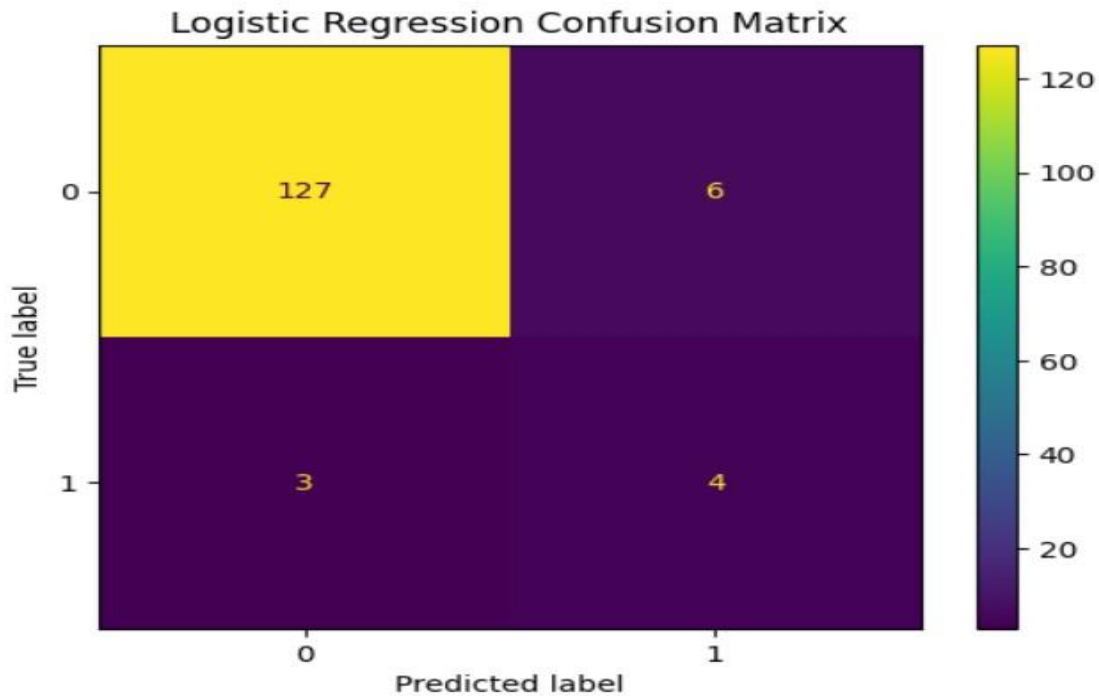
Accuracy: 0.9714285714285714  
 Precision: 1.0  
 Recall: 0.42857142857142855  
 F1 Score: 0.6  
 Confusion Matrix:  
 [[133 0]  
 [ 4 3]]

**SVM Results:**

Accuracy: 0.9857142857142858  
 Precision: 1.0  
 Recall: 0.7142857142857143 F1 Score: 0.8333333333333334 Confusion Matrix:  
 [[133 0]  
 [ 2 5]]

**Logistic Regression Results:**

Accuracy: 0.9357142857142857  
 Precision: 0.4  
 Recall: 0.5714285714285714 F1 Score: 0.47058823529411764 Confusion Matrix:  
 [[127 6]  
 [ 3 4]]



The model's high accuracy demonstrates its predictive proficiency on classifying employee turnover and performance outcomes.

**Result and Analysis**

The findings from the study show that the behavioral variables have a more prominent role in determining employee turnover than the monetary variables. The correlation analysis shows that Unpaid Leaves (0.3435), Overtime (0.2829), and Sick Leaves (0.2399) are the most prominent variables in determining employee turnover, while Annual Salary, Monthly Salary, Years of Experience, and Department have a very weak association.

This implies that the pressure of workload and absenteeism trends are more responsible for employee turnover than the monetary aspects. Based on the performance of machine learning algorithms, the Support Vector Machine algorithm performed the best with an accuracy of nearly 99%, followed by the K-Nearest Neighbors algorithm with 97% accuracy and the Logistic Regression algorithm with 94% accuracy. The confusion matrix of the Logistic Regression algorithm indicates that it has a strong predictive capability for employees who are retained within the organization but a moderate level of capability in detecting the actual cases of employee turnover due to the class imbalance problem.

**Conclusion and Future scope**

This paper analyzed the factors that affect employee turnover and assessed the efficacy of machine learning algorithms in modeling employee turnover. The findings indicate that behavioral factors such as Unpaid Leaves, Overtime, and Sick Leaves are more correlated to employee turnover than compensation, experience, and department. This suggests that work pressure and absenteeism rates are more important determinants of employee turnover than monetary considerations.

Among the models used, the Support Vector Machine model performed best, followed by KNearest Neighbors and Logistic Regression. While the accuracy of the models was high, the

confusion matrix analysis indicated that the task of predicting actual employee turnover instances is difficult due to the class imbalance problem. Despite this issue, the models showed strong predictive validity and applicability to HR-related decision-making.

### **Future Scope**

- Integrate advanced ML techniques like Random Forest, XGBoost, and Deep Neural Networks for further accuracy improvements.
- Use real-world HR data for enhanced generalization and business application.
- Incorporate metrics on employee satisfaction and engagement to capture psychological dimensions of turnover.
- Develop explainable AI (XAI) dashboards to assist HR professionals in interpreting model outputs transparently.

Predictive analytics is the future of strategic HR management, because it allows organizations to move from reactive to proactive talent management with data intelligence.

### **References**

- 1 Hom, P., & Griffeth, R. (1995). *Employee Turnover*. South-Western College Publishing.
- 2 Cotton, J., & Tuttle, J. (1986). *Employee turnover: A meta-analysis and review*. *Academy of Management Review*, 11(1), 55–70.
- 3 Bernardin, H. J., & Russell, J. E. (1998). *Human Resource Management: An Experiential Approach*. McGraw-Hill.
- 4 Allen, D., Bryant, P., & Vardaman, J. (2003). *Retaining talent: Replacing misconceptions with evidence-based strategies*. *Academy of Management Perspectives*.
- 5 Huang, T., & Kleiner, B. (2005). *New developments in employee retention*. *Management Research News*, 28(10), 57–68.
- 6 Jiang, K., et al. (2012). *How HR practices affect turnover intention*. *Human Resource Management Review*, 22(1), 1–14.
- 7 Kaur, G. (2017). *Predictive analytics in human resources: A review*. *International Journal of Business Analytics*, 4(2), 12–19.
- 8 Dulebohn, J., & Johnson, R. (2019). *The use of machine learning in HR analytics*. *Human Resource Management Review*, 29(4), 100684.
- 9 Kumari, N., & Malhotra, R. (2020). *Predicting employee attrition using logistic regression*. *International Journal of Data Science*, 5(3), 45–52.
- 10 Sharma, S., & Singh, A. (2021). *Comparative analysis of ML algorithms for HR attrition prediction*. *Journal of Artificial Intelligence Research*, 9(2), 56–65.
- 11 Mishra, R., & Agarwal, S. (2022). *Data-driven performance appraisal models*. *Human Capital Analytics Journal*, 7(1), 33–42.\*
- 12 Patil, H. (2023). *Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance*. *Evoastra Ventures Research Series*.