



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 3 - V7I3-1431)

Available online at: <https://www.ijariit.com>

## Employee attrition prediction

Varshit Jain

[jainvarshit99@gmail.com](mailto:jainvarshit99@gmail.com)

St. Francis Institute of Technology,  
Mumbai, Maharashtra

Ansh Aya

[anshaya92@gmail.com](mailto:anshaya92@gmail.com)

St. Francis Institute of Technology,  
Mumbai, Maharashtra

Dhairya Desai

[desaidhairya3@gmail.com](mailto:desaidhairya3@gmail.com)

St. Francis Institute of Technology,  
Mumbai, Maharashtra

### ABSTRACT

*The success of major corporations/organizations depend on the employees working in it. These employees leave the jobs due a variety of reason ranging from personal problems, unsatisfactory working conditions to being fired for not meeting the requirements of the job. This leaving of employees is affecting the organizations in terms of cost as well as lost productivity. In this paper, we attempt to develop a system to predict the employee attrition dependent on the data within the organization.*

**Keywords**— Employee Attrition, Classifications, Management, Random Forrest

### 1. INTRODUCTION

Employees are expensive to the organization. The organizations have to spend a lot of money, time and resources to train them enough to be able work on the projects within the organization. The other way for organization to get employees meeting their requirements is to get them from other organizations which requires offering them huge and generous packages to them to motivate them to join their organization. As we have established that the immense capital required by the organization for employees, it can be said that the leaving of an employee results in a loss to the organization. Any organization's goal is to reduce the employee attrition by creating a more employee friendly environment so as to persuade employees to not leave. Predicting the employees at the risk of attrition can help the organization identify the factors that can cause a employee to be unsatisfied and at risk to leaving the job. If identified correctly the internal policies or environment factors causing the employees to leave can be rectified. The at-risk employees can be identified and a retention strategy can be created to encourage the employees to stay with the organization. These strategies can include offering increased pay, improving infrastructure used by them or providing more leadership or development opportunities. The motivation of this project is to create a system that will help organization in their efforts to reduce employee turnover with the of machine learning models. This paper suggests the use of random forest machine learning algorithm, which is an ensemble learning method for classification and regression.

### 2. LITERATURE REVIEW

Employee attrition refers to the gradual loss of employees over time. Most literature on employee attrition categorizes it as either voluntary or involuntary. Involuntary attrition is thought of as the mistake of the employee, and refers to the organization firing the employee for various reasons. Voluntary attrition is when the employee leaves the organization by his own will. This paper focuses on voluntary attrition. A meta-analytic review of voluntary attrition [1] found that the strongest predictors of voluntary attrition included age, pay, and job satisfaction. Other studies showed that several other features, such as working conditions, job satisfaction, and growth potential also contributed to voluntary attrition [2][3]. Organizations try to prevent employee attrition by using machine learning algorithms to predict the risk of an employee leaving, and then take pro-active steps for preventing such an incident.

### 3. METHODOLOGY

#### 3.1 Dataset Description

In this paper, we have used a HR dataset. The dataset contains about 15000 employees across 10 attributes. The dataset contained 8 numerical attributes and 2 non numerical attributes. The non-numerical attributes were converted to numerical attributes before processing of the data. The salary level had 3 distinct values: Low, Medium and High which were respectively converted to 1 2 and 3. The department attribute consisted of 9 distinct values which were converted to numerical values as follows: IT=0, R&D=1, HR=3, Management=4, Marketing=5, Product Management=6, Sales=7, Support=8, Technical=9.

**Table 1: Dataset Features and type of Data**

Features	Type
Satisfaction Score	Numeric
Last Evaluation	Numeric
Average Monthly Hours	Numeric
Number of Projects	Numeric
Time Spent with Company	Numeric
Work Accidents	Numeric
Promotions in 5 years	Numeric
Department	Non-Numeric
Salary Level	Non-Numeric

### 3.2 Random Forrest

The proposed solution includes the use of random forest algorithm which is an ensemble classifier that uses decision trees to make predictions. Using a single decision tree can be very inefficient hence using random forest enables us to create multiple decision trees and merge them together to get a more accurate and stable prediction. Random forest also makes it extremely easy to measure the relative importance of each feature in the prediction.

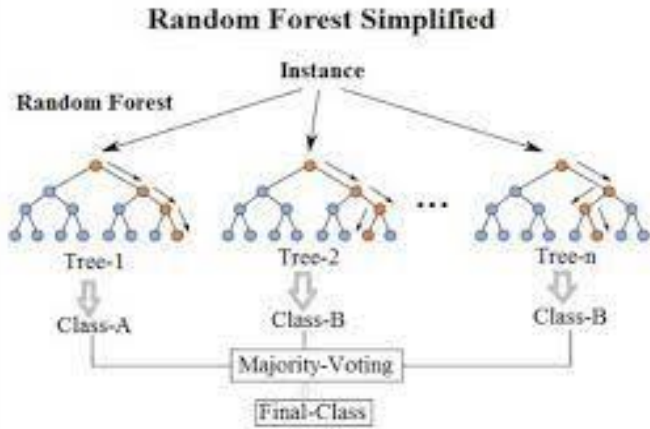


Fig. 1: Random Forest Visualization

1. Read the dataset
2. Clean the data, remove NULL values and fill in the missing values.
3. Select features important for prediction of the attrition.
4. Split the dataset into training and testing dataset for the model.
5. Create a prediction model based on random forest algorithm.
6. Use the model created on the current data to find employees at risk.
7. Create a strategic retention strategy for the employees.

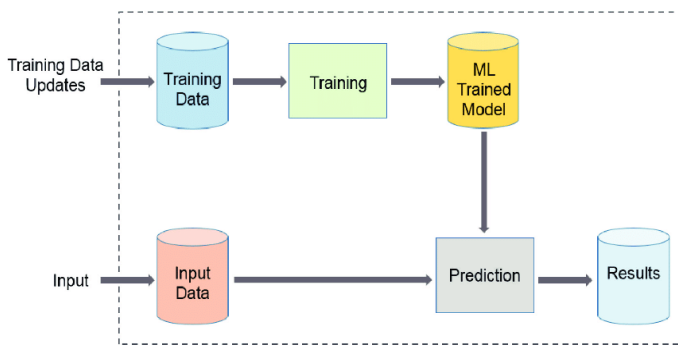


Fig. 2: Model Creation Workflow

### 3.3 Result

Table 2: Model Results

Accuracy	Precision	Recall	AUC
0.98	0.99	0.95	0.99

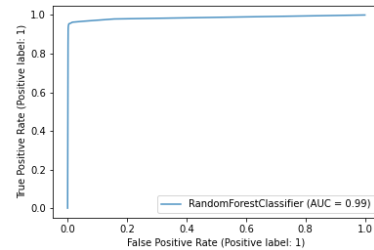


Fig 3: Area Under Curve

The proposed simple random forest model had an accuracy of 98% with the precision of 99% and Recall of 95%. When compared to other classification algorithm it performed better than others. The ROC Curve shows the general ‘predictiveness’ of the classifier. It measures the probability that the classifier ranks a randomly chosen positive instance higher than a randomly chosen negative instance. Closer the curve to the top left corner, better the classifier.

With the use of feature importance, we found out the satisfaction score is the most important feature in determining employee attrition followed by last evaluation score and number of projects with average monthly working hours. Work Accidents and Promotions have a lot less effect on the employee’s decision to leave the company.

### 4. CONCLUSION

In this paper we present a simple random forest model to determine employee attrition in organizations. Our model had a accuracy rate of 97%. The main reason for attrition is employee satisfaction. The organizations should work to improve the employee satisfaction via various strategies like increased pay, improved work-life balance. The organizations should try to determine the at risk the employees to better understand the major reasons of unsatisfactory conditions and improve them. These plans help both the employees with a better work life and the organization with decreased spending and improved productivity among employees. As seen in the result the highest turnover rate is in the Sales department and better plans should be recommended to reduce the higher turnover rate.

### 5. REFERENCES

- [1] Cotton, J.L. and Tuttle, J.M., 1986. “Employee turnover: A meta-analysis and review with implications for research” Academy of management review, pp.55-70.
- [2] Liu, D., Mitchell, T.R., Lee, T.W., Holtom, B.C. and Hinkin, T.R., 2012. “When employees are out of step with co-workers: How job satisfaction trajectory and dispersion influence individual-and unit-level voluntary turnover”. Academy of Management Journal, pp.1360-1380.
- [3] Heckert, T.M. and Farabee, A.M., 2006. “Turnover intentions of the faculty at a teaching-focused university”. Psychological reports, pp.39-45.
- [4] R. S. Shankar, J. Rajanikanth, V. V. Sivaramaraju and K. V. S. S. R. Murthy, "Prediction of employee attrition using datamining," 2018 IEEE international conference on system, computation, automation and networking (ICSCAN), 2018, pp. 1-8, DOI: 10.1109/ICSCAN.2018.8541242