

# STUDY VARIOUS FACTORS AFFECTING ATTRITION FACTOR IN A COMPANY AND PREDICT USING MACHINE LEARNING ALGORITHMS

Name: Tushar Singh

Email: [tushar.24.ts@gmail.com](mailto:tushar.24.ts@gmail.com)

Graduate,

Extc, Ramrao Adik Institute Of Technology,

Navi Mumbai

---

*ABSTRACT: Attrition rate is high in IT companies due to various reasons. This article is going to help us understand various factors which affect the attrition rate. With the help of statistical tools and data visualization will be able to draw various conclusions such as which factors affect the most. Employees tend to leave the company for better hikes, environmental protection. The role of the HR department is to try and contain the employees and avoid employees leaving the company, try and study various reasons which are prominent and important for employees to sustain in the company. High attrition rate signifies that the employees are usually not happy with the company and low attrition rate suggests that people are happy with the company.*

---

## I. INTRODUCTION

Attrition rate stands for the number of employees vacated by the total number of employees. It can also be said as Employee Turnover or Employee Defection". Mostly a highly-trained and talented employee leaves the organization, it creates a huge void in the company. So, the company loses key knowledge and business relationships which the vacating person had. Modern HR Managers are majorly focused and interested in decreasing the Attrition in the organization, in such a way that it will contribute to the maximum effectiveness, growth, and progress of the organization.

Employees are the most valuable and important assets of an organization. Therefore, it is important to maintain a permanent relationship with the entire workforce in order to maximise the company's profit. Over the years it has become a tough task for employers and thereby increased attrition in the organizations.

HR managers play a key role in studying various reasons and factors which contribute towards the attrition. Due to rise in competition between companies as well as employees there's a huge market which is available for the employees in order to satisfy their intentions and dreams.

## **II. REVIEW OF LITERATURE**

Michael Armstrong (2006) defines Employee Attrition as a normal flow of people out of an organization through retirement, career or job change, relocation, illness and so on. Jack, Philips and Adele define Employee Attrition as the percentage of employees leaving the organization for whatever reasons. Turnover rate can be briefly described as how fast the employers recruit and lose employees (Chikwe, 2009). It is used to measure the effectiveness of recruitment (Mondy, 2010) and is sometimes considered as one of the indicators of organizational performance (Cho, Woods, Jang, & Erdem, 2006). Mondy (2010) clearly defined turnover rate as how many new recruitments were hired to replace resigned employees. By these definitions, turnover ‘occurs’ only when a replacement is successfully hired. Wayne F Cascio and John W Boudreau (2008) introduced two popular ways of classifying employee attrition as voluntary attrition versus involuntary and functional attrition versus dysfunctional attrition. One of the earliest models of turnover was developed by March and Simon in 1958 (Hom & Griffeth, 1995). This model described individuals and organizations as being in a state of equilibrium, where the members contributed to the organization while the organization provided members with compensation in return. March and Simon posited that when the compensation provided by the organization is no longer balanced with the contribution of the organizational members, individuals quit the organization. Price (1977) identified five primary determinants of turnover – pay levels, integration (defined as the involvement one has in one’s relationship with a supervisor or coworker), instrumental communication (defined as how clearly the work role is communicated to the employee), formal communication (defined as how well the organizational communicates practices and policies), and centralization (defined as the distribution of power in the organization). He proposed that the first four determinants are positively related to turnover while centralization is negatively related to turnover. Mitchell and Lee (2001) suggested that there when individuals have multiple attachments to the organization, these attachments are likely to hold them back from leaving even if they think about leaving due to particular circumstances (e.g., getting another offer, company relocation to a non-preferred location). Thus, individuals who are high on job embeddedness might choose to stay with the organization even if circumstances are less than ideal. Job embeddedness is a multidimensional construct that describes the various attachments that an individual has with the organization and community.

## **III. SCOPE OF THE STUDY**

The data was downloaded from Kaggle which is an open source platform where various datasets are available for various studies. The dataset contains information about 1470 employees from age 18-60.

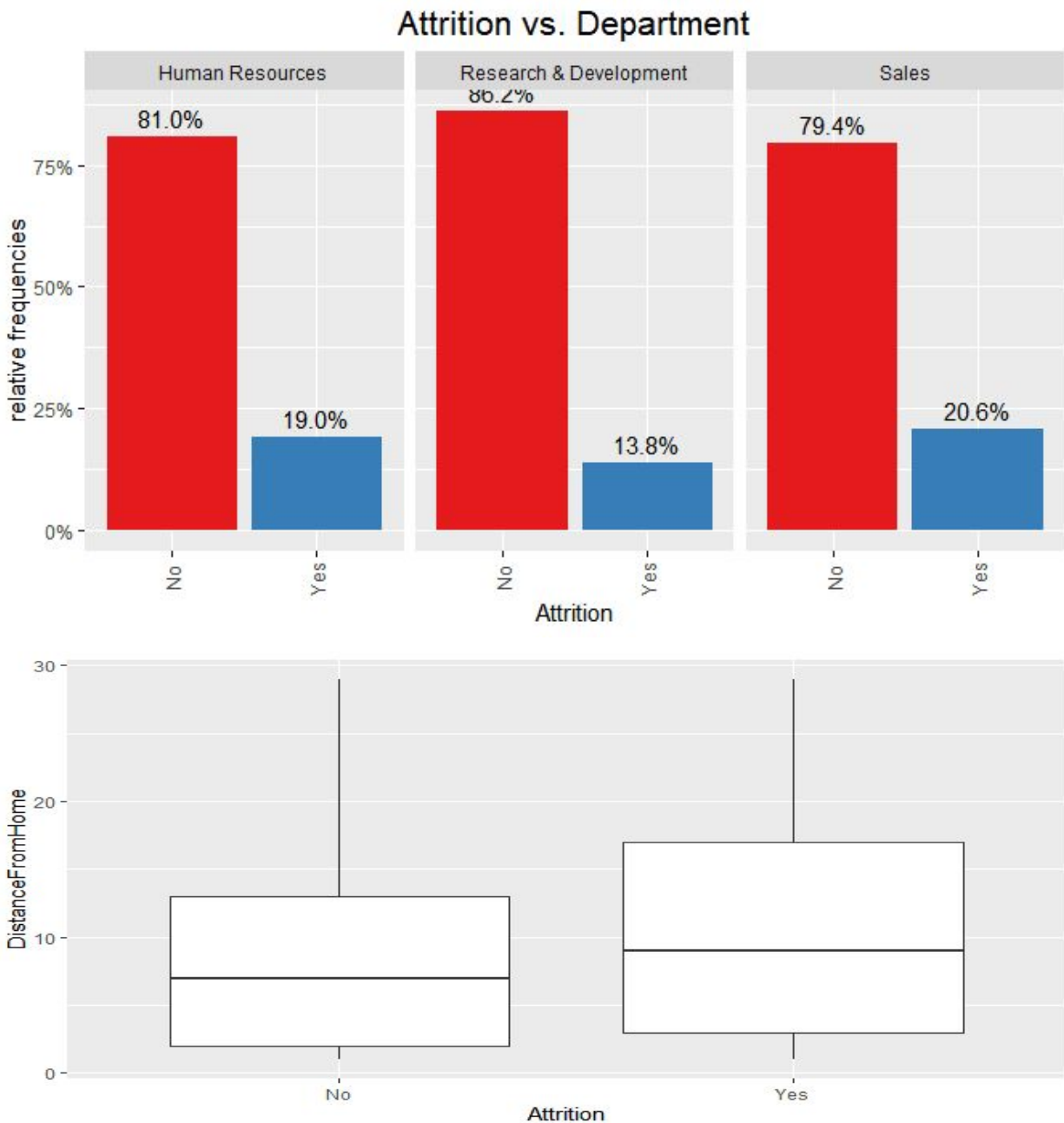
## **IV. OBJECTIVES OF THE STUDY:**

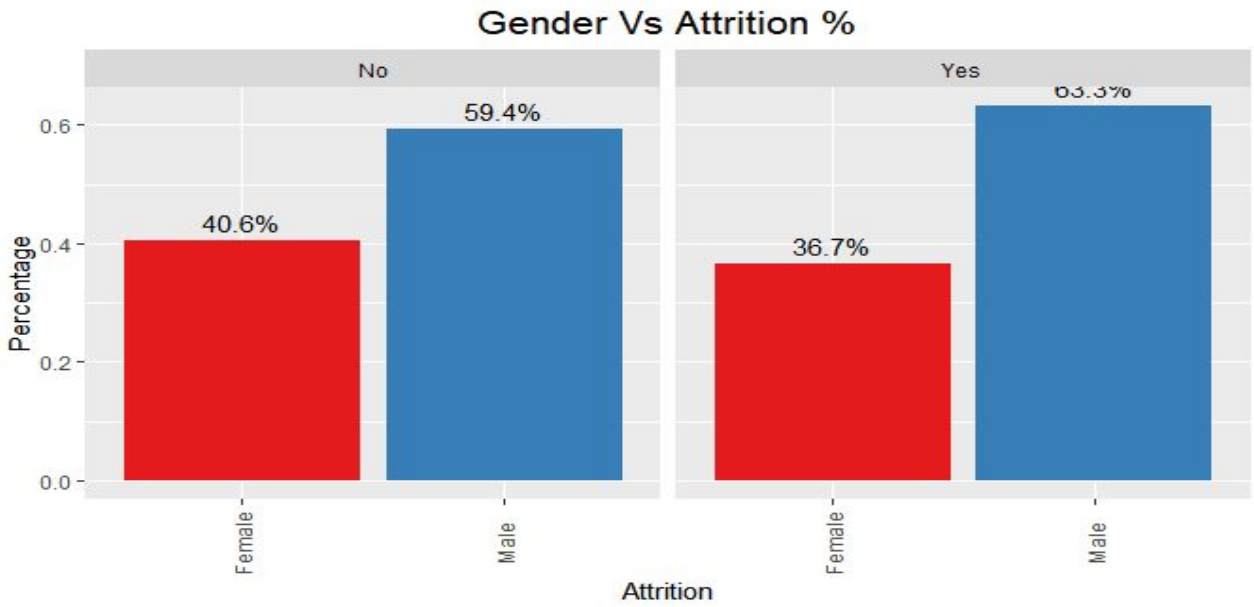
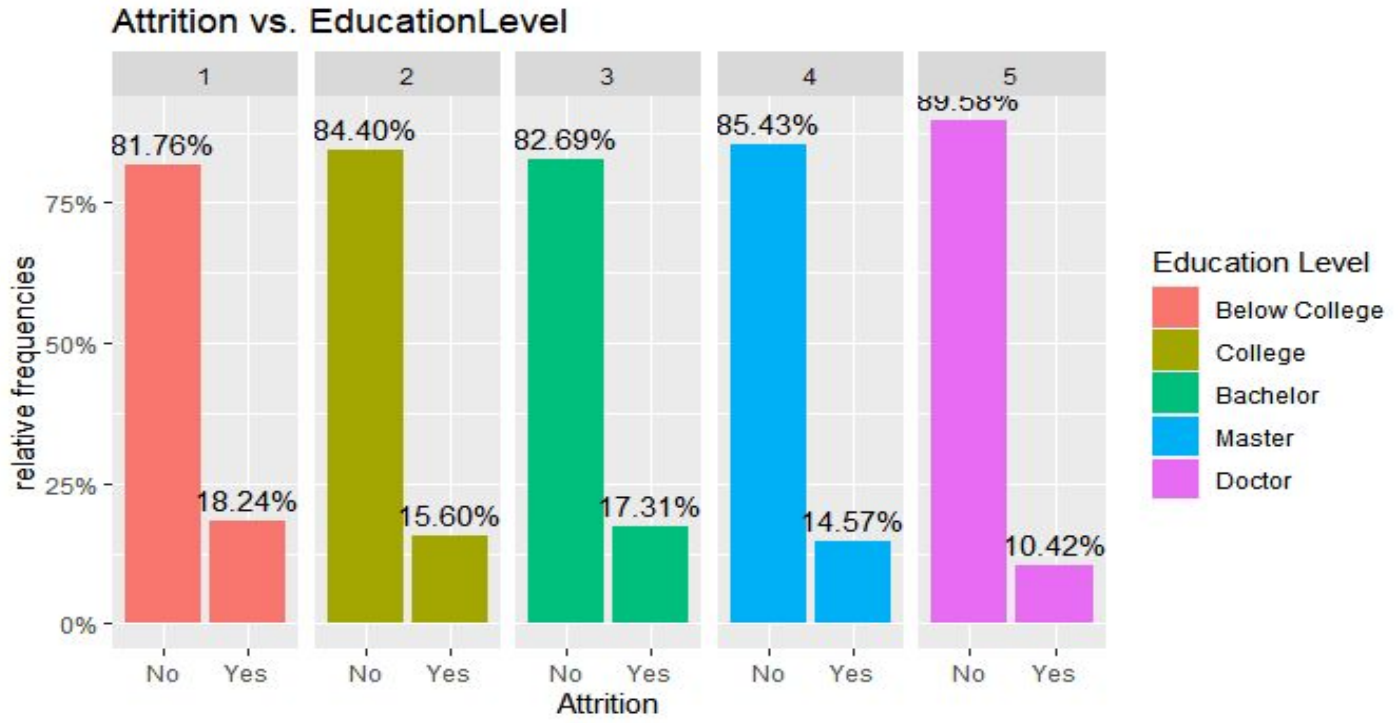
1. To study the factors that increase employee attrition in the IT sector.
2. To study the associated factors for attrition which can be adjusted to retain employees.

## V. DATA DESCRIPTION AND ANALYSIS

**1. Data Description:** There are 31 columns and 1470 rows for study. Columns such as age, distance from home, Department, Education Qualification, Business trips. Dependent variable is the column “attrition” which is dependent on various independent variables.

**2. Data Visualization:** In this section with the help of RStudio and available various graphs are being plotted in comparison with dependent and independent variables and then draw conclusions which variables are prominent. As there are approximate 31 columns so it’s not possible to draw out all conclusions so only few graphs are considered in this article.





## VI. MODEL BUILDING

As seen above there are various graphs plotted with respect to dependent variable. Now using different machine learning algorithms will try to build an accurate model which will predict take various factors as an input and determine whether the following employee will take an exit from the company or not.

**1. Decision Tree:** A decision tree is a flowchart-like structure in which each internal node represents a test on a feature (e.g. whether a coin flip comes up heads or tails) , each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules. Information Gain and Gini index are two essential properties in the decision of tree computation. Information Gain is defined as the amount change in entropy. Higher entropy indicates more effectiveness of the substance. Gini Index measures how regularly an arbitrarily picked component would be falsely recognized. It indicates that characteristics with a lower Gini index ought to be preferred first.

**2. Random Forest:** This is an algorithm which ensembles the less predictive model to produce better predictive models. It aggregates the base model to create a large model. The features are sampled and passed to trees without replacement to obtain the highly uncorrelated decision trees. To select the best split it is required to have less correlation between the trees. The main concept that makes random forest different from the decision tree is aggregated uncorrelated trees.

**3. Boosting:** Boosting is an important and most effective mechanism in machine learning. Unlike previous methods and other methods available in ML, boosting algorithms try to learn and fix their mistakes. Initially the same weight is assigned to all the rows and then after classification the rows which were misclassified then in the next iteration those misclassified rows are assigned with higher weights in order to correct themselves. A small learning rate is assigned and the steps are repeated for a few number of times till the error is below a threshold.

**4. Regularization:** This is again a prominent ML algorithm which penalizes the coefficients and reduces the value of coefficients. Regularization methods for e.g Ridge and Lasso Regularizations are used in situations where the problem of overfitting is arised. After penalizing the coefficients the problem of overfitting is reduced. Lasso regularization also penalizes some coefficients to zero which are less important hence these algorithms also play an important role in feature selection

## VII. EXPERIMENTAL RESULTS

<b>Model Name</b>	<b>Accuracy</b>
Decision Tree	84.7%
Random Forest	85.44%
Gradient Boosting	87.77%
Ada Boosting	85.5%
Ridge	89.93%
Lasso	87.3%
XGBoost	93.35%

## VIII. FINDINGS OF THE STUDY

As we study various factors, the factors which were responsible for high attrition rate are:

1. Among youngsters as they wanted new job roles as per their choice.
2. Employees which had homes far away from the company.
3. People who were working under the same manager for more than 5 years.
4. Employees who weren't offered more business trips.
5. Employees who had issues with the work environment as well as job role.

The results showed that lack of recognition for their work and lack of future growth is one of the reasons for attrition. Fair treatment and respect of employees are taken into stake, which thereby lead to no promising services from the side of employees. Gender inequality in the organisation also contributes to attrition. The policies that do not support the staff gradually lead to difficulties in getting the work done as expected and retention of employees.

## IX. CONCLUSION

After going through the entire article one can conclude that with the help of reliable data the HR can study various factors which lead to attrition in the organization and experts can build efficient graphs and models. With more data collection HR and other experts can hence determine whether the employee will take an exit or not.

## X. REFERENCES

- Dr K. MalarMathi, G.Malathi (2015). *Analyzing the causes of attrition in IT industries-opinion of employee and employer*. Asia Pacific Journal of Research, Vol1.
- Dr Sneha Manikar (2013). *Infant attrition in Indian IT sector: An indication*. International journal of scientific and research publications, Vol 3.
- Dr shine David (2015). *Attrition in "IT" sector*. International journal of core engineering and management, Vol 2.
- Savneet Kaur (2013). *Attrition in India- causes & remedies*, International Journal of Emerging Research in Management & Technology, Vol-2, Issue-6, pp. 39-42
- Caramollah Daneshfard & Kokab Elsadat Ekvaniyan, (2012). *Indisciplinatory journal of contemporary research in business*, vol.3 no.9