# EMPLOYEE ATTRITION PREDICTION MODEL

**Shikhar Jamuar, Anish Pahwa, Rahul Raj, Vinitha Ravindran, Mengying Sun**

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

[sjamuar@purdue.edu](mailto:sjamuar@purdue.edu); [pahwaa@purdue.edu](mailto:pahwaa@purdue.edu); [rajr@purdue.edu](mailto:rajr@purdue.edu); [ravindr1@purdue.edu](mailto:ravindr1@purdue.edu); [sun907@purdue.edu](mailto:sun907@purdue.edu)

## **Abstract**

Employee attrition is one of the critical factors that hamstring a company’s growth. Understanding whether an employee would attrite or not based on a set of factors could help companies to observe patterns and take preventive actions to correct this. Therefore, we aim to develop a model that can be used by HR personnel to observe various factors of an employee and determine whether he would attrite or not.

High employee attrition is expensive, results in the loss of experienced employees, affects overall productivity and ultimately brings down the profit of an organization.1

Using an employee dataset from a company, we have used packages such as Caret and H2O to run descriptive analyses to determine various traits. Subsequently, we developed a model that predicts whether an employee would attrite or not and created a Shiny UI interface where employee factors could be entered to get the result.

## **Business Problem**

Employee Attrition is the reduction of staff by voluntary or involuntary reasons. These can be through natural means like retirement, or it can be through resignation, termination of contract, or when a company decides to make a position redundant. In 2017, the Bureau of Labor Statistics (BLS) found that 15.1% of the total U.S. workforce voluntarily quit a position, retired, was laid off or discharged.2

In order to function profitably and remain relevant in their industry, companies need to keep a track of their employee attrition rate and identify reasons for high attrition. An interface where the concerned stakeholders, i.e., HR personnel could gain information on employee attrition by inputting some relevant parameters would help companies track their attrition rate and possibly take steps to address this and subsequently improve its operating model.

## **Analytics Problem**

This could be regarded as a classification problem, where we assess whether an employee will attrite (yes) or not (no). This could be determined by considering factors such as employee age, job position, number of years and salary. These factors could also be used to conduct some descriptive analytics, where various trends can be observed.

While we can develop a predictive model, we need to be cognizant of the fact that there are several other intangible factors such as work location and demographics of the area and relationships with superiors at work which could affect an employee’s decision to attrite. It would be difficult to quantify these factors and incorporate into any model we make.

The factors mentioned above could serve as indicators as to how happy an employee is at her company, which reduces the chance of her switching.

## **Data**

The dataset is one hosted by IBM on Kaggle. The dataset contains 1,470 observations, with 34 potential predictors.

One of the main reasons why we took this dataset is because of the diverse set of observations available. The monthly income varies from $ 1,000 to $ 20,000, and age varies from 18 to 60. There is a good distribution of employees across Job Roles, and there is a good mix of males and females. Overall, the dataset looked to be one that represented a wide variety of employees, which would certainly improve model reliability.

## **Methodology Selection**

For descriptive analysis, we considered

As for predictive analytics - since it’s a classification problem, we considered Random Forest, Gradient Boosting Machine and Logistic Regression.

As for the data, since we wanted to create an interactive UI for HR personnel to update values, we filtered out variables based on Information Value (IV) and prepared the model with 9 variables. These 9 variables seemed to be in line with what one could expect to be critical factors in deciding employee attrition.

## **Model Building**

On running the aforementioned models, we found out that the best model was the Logistic regression mode, with a misclassification rate of below 0.15.

One of the main limitations of our model would be the fact that we only considered nine variables for simplicity of the interface. In doing so, we might have lost out on valuable information that could have improved the model accuracy.

## **Functionality**

Initially, we plotted few graphs to understand various employee trends.

Our DSS has two components: one where the HR personnel could upload a file containing employee details and the model would calculate attrition for each of them.

The other component enables the user to enter values for a specific employee and determine whether he would attrite or not.

Possible enhancements to the interface include improving the model reliability by incorporating datasets from companies in different fields of space and various locations.

## **GUI Design and Functionality**

The first component of the Shiny app has a file import option, where the user could enter a .csv or excel file with employee details and get predictions generated for them. The second component has input options for 9 variables for a particular employee, which is used to generate prediction on whether he will attrite or not.

**Conclusions**

Successful early detection of attrition could help companies take measures to ensure they can prevent this from happening, and henceforth help retain valuable talent and cut out on unnecessary costs. Hence, this model and interface would be of great use to HR departments.

**References**

[1] When Numbers Fall: 4 Negative Effects of Employee Turnover, gethhpy.com,

*https://gethppy.com/employee-turnover/numbers-fall-4-negative-effects-employee-turnover*

# 2 What is the difference between employee turnover and employee attrition, August 22, 2018 *https://business.dailypay.com/blog/employee-turnover-vs-attrition*