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; pahwaa@purdue.edu; rajr@purdue.edu; ravindr1@purdue.edu; 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 see 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 datasets can be uploaded to obtain the attrition percentage.

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 take a position redundant. In 2017, the Bureau of Labor Statistics (BLS) found that 15.1% of the total U.S. workforce voluntarily quit a job, retired, was laid off or discharged.[2] Some studies predict that every time a business replaces a salaried employee, it costs 6 to 9 months' salary on average.[3] To function profitably and remain relevant in their industry, companies need to keep 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
The problem could be regarded as a classification one, where we assess whether an employee will attrite (yes) or not (no). It 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 we can observe various trends.
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 them into any model we make.
Factors mentioned above could serve as indicators to indicate how happy an employee is at one's company, and a high level of satisfaction reduces the chance of one's 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 ranges from 18 to 60. There is a proper distribution of employees across Job Roles, and there is a right mix of males

[1] When Numbers Fall: 4 Negative Effects of Employee Turnover. (n.d.). Retrieved from https://gethppy.com/employee-turnover/numbers-fall-4-negative-effects-employee-turnover
[2] Pawlewicz, K. (2018, August 22). What is the difference between employee turnover and employee attrition? Retrieved from https://business.dailypay.com/blog/employee-turnover-vs-attrition
[3] Merhar, C. (2016, February 4). Employee Retention - The Real Cost of Losing an Employee: 2019. Retrieved from https://www.peoplekeep.com/blog/bid/312123/employee-retention-the-real-cost-of-losing-an-employee

and females. Overall, the dataset looked to be one that represented a wide variety of employees, which would undoubtedly improve model reliability.

Methodology Selection

For descriptive analysis, we considered plots where we compare attrition with job roles, departments and the overall attrition rate in the company. Furthermore, we consider attrition in employees varying in age, education level and so on. This analysis helps us understand how these factors affect employee attrition.

As for predictive analytics - since it's a classification problem, we have considered the Random Forest model, and the Gradient Boosting Machine model.

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 nine variables. These nine variables seemed to be in line with what one could expect to be critical factors in deciding employee attrition.

Model Building

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

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

Functionality

Our DSS has two components: one where various trends affecting employee attrition can be identified. The other component can be used by the HR personnel to upload a file containing employee details, and the model would calculate attrition percentage for each of them.

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

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.