**Data Driven Solution to Employee’s Attrition**

# Rasha Atshan

# August 19, 2019

Every employee plays an important role in an organization. In the meantime, good employees have a high value that can’t be replaced for reasons such as the deep understanding of the organization and ways for improvement based on past experiences at the organization. They hold a client relations and trust that have been established for years. And good employees influence their colleague’s confidence in the workplace and workload. Employee Attrition means the voluntary or involuntary reduction of staff or employee in an organization. Attrition is due to many factors such as employee retirement, employee resignation, contract termination, and so on. Employees attrition has a negative impact on the corporate culture and organization reputation in front of clients and other employees. Employee attrition forces organization to hire and train new employees which requires time and very costly. So, we decided to uncover the factors behind employee’s attrition and the quality of employee’s performance at work.

In the research, we analyze the IBM HR Employee-Attrition and Performance fiction data using logistic regression to discover the reasons behind employee’s attrition. The data consists of 1470 subjects and 35 variables. The 35 variables focus on demographics characters of the employees and work environment related questions. Simply Logistic is predictive analysis that is used to describe data and to explain the relationship between one dependent binary variable and one or more independent variable. In our research, the dependent binary variable is Attrition with a “yes or no” respond. First, we used all possible 34 independent variables to observe their effect and significance on the independent variable Attrition. Then, we performed three types of model selection methods: best subset selection, forward stepwise selection, and backward elimination, to obtain the best model. Then, we fitted a new model that consists of the significant independent variables. Finally, we confirmed the significance of the selected model using cross-validation and diagnostic plots.

The analysis results confirm that nineteen variables effect valuable employee’s decision to leave an organization. Some of the demographic personal variables such as marital status, age, education level, and relationship satisfaction are out of the organization’s control. In the meantime, some other variables such as job role, number of years since last promotion, overtime, life-work balance, home-work commute, and many others can be easily fixed to maintain valuable employees. For example, organization can offer working remotely or relocating assistance for the valuable employees. Organization could improve their work environment, and provide a better life-work balance to maintain their employees and help them perform best at their jobs.

These variables are age, home-work commute, education field, gender, marital status, relationship satisfaction, work-life balance,

job stratification, job role, job involvement, number of companies worked at, over time, number of working year, training time, years at the company, and years at the current role, years since last promotion, years with current manage,

Age + DistanceFromHome + EducationField

+ JobSatisfaction + Gender + JobInvolvement + JobRole + MaritalStatus

+ NumCompaniesWorked + OverTime + RelationshipSatisfaction + TotalWorkingYears

+ WorkLifeBalance + TrainingTimesLastYear + WorkLifeBalance + YearsAtCompany + YearsInCurrentRole

+ YearsSinceLastPromotion + YearsWithCurrManager

the model we selected using cross-validation,

<https://recruitloop.com/blog/7-ways-reduce-employee-attrition/>

<https://www.investopedia.com/terms/a/attrition.asp>

<https://www.forbes.com/sites/amyanderson/2013/02/13/great-employees-are-not-replaceable/#1261da4f5403>

<https://www.statisticssolutions.com/what-is-logistic-regression/>

<https://scientistseessquirrel.wordpress.com/2018/10/02/presenting-statistical-results-pointers-in-scientific-writing/>

<https://stats.stackexchange.com/questions/234998/logistic-regression-diagnostic-plots-in-r>