Employee Churn Reduction

Shrikrishna Padalkar

November 2018

1 Introduction

From a firms perspective, it is always difficult to hire new employees than to retain the existing ones. So companies need to spend a substantial amount of resources in terms of both money and manpower for hiring new employees. Reducing the employee attrition will result in reducing the companies' expenditure. This report is all about understanding the reasons for Churn of Employees. We are going to elaborate on the methods used for predicting whether an Employee will churn out or remain with the company.

2 Exploratory Data Analysis

The data set comprises of two files training file and and test file each containing 21 columns. The training file consists of 3333 rows. whereas the testing file consists of 1667 rows.
Let's have a look at the top 5 rows of the training file

2.1 Plotting

3 Check Class Imbalance

Class Imbalance occurs when the number of observations of one or more classes are very less as compared to the number of observations of other classes. Here, we have two classes Churned Employees and Not Churned Employees. There exist 2850 records of employees who remained with the organization and 483 employees churned out. Thus the difference is substantial.

Following are a few techniques to deal with class imbalance:-

- Under Sampling:- In this technique we choose a subset of samples from the majority class in such a way that the number of samples are closer to the number of samples in the minority class.

Here we have used **Systematic Sampling** as a choice of sampling technique. Consider a two class classification problem. Suppose there are N observations from the majority class and n from the minority class. We choose every k^{th} observation from the majority class. Where $k = N/n$. This will result in the majority class having number of observations close to the minority class.

- Over Sampling:- In the case of Over Sampling we increase the size of the minority class by adding a dummy observations of minority class.