Employee attrition is the departure of an employee(s) from employment, for whatever reason, be it voluntary or involuntary. Some of the major causes of attrition includes resignation due to dissatisfaction, lack of opportunities for growth and discrimination, followed by terminations, lack of skills and retirement. Presently as shown, attrition stands for 16.8% in UK while the accepted maximum threshold is 20%. Once it touches the maximum threshold, it would see a gradual decrease in economy and overall development of the country. We have chosen Logistic Regression as our model, as an employee of a company resigning his workplace would be a binomial answer.

At the conclusion of our analysis, we realized that we have achieved 85% accuracy after a optimisation based on Recursive Feature Elimination (RFE), which is based on the idea to repeatedly construct a model and choose either the best or worst performing feature, setting the feature aside and then repeating the process with the rest of the features. This process is applied until all features in the dataset are exhausted. Some of further analysis includes:

1. Binning data – Although used to reduce errors, it may get complicated based on the dataset we have got
2. Feature Selection – This would help us in known irrelevant features and picks the best set of features for better dimensionality
3. Closer look at pre-processing
4. Combining methods

Whether an employee is going to stay or leave a company, his or her answer is just binomial i.e. it can be “YES” or “NO”. So, we can see our dependent variable Employee Attrition is just a categorical variable. In the case of a dependent categorical variable, we can not use linear regression, in that case, we have to use “LOGISTIC REGRESSION“.