**HR Analytics Project**

* **Problem Definition**

To find an ideal candidate is a crucial part of recruitment and one need to access some kind of insight for the number of people leaving the organization and the reason behind leaving. Attrition is inevitable part and poses a great challenge for any organization, hence with vast data available at hand we can analyze our human resource leaving and staying in organization and factors responsible for it.

* Data Analysis.

We have 34 different features available to predict our label(Attrition). Feature data include both numerical and object datatypes. Brief descriptions of important features are as follows.

Age*:* - We have employees between the age group of 18 to 60 years

Business Travel: - Employee’s frequency of business related travel

Daily Rate: - Per day pay employee is entitled to

Department: - list of different departments in a company

DistanceFromHome: - Distance in Kms from employee’s residence

Education: - Highest level of educational qualification

EducationField: - Field of qualification

EnvironmentSatisfaction: - How satisfactory is the environment for the employee as per the survey

JobInvolvement: - level of dedication/involvement in work

JobLevel: - Level of Job

JobSatisfaction: - How satisfied employee with job

MonthlyIncome: - Monthly salary

NumberOfCompanies: - Total number of companies employee has worked so far

OverTime: - Wage for working extra

PercentSalaryHike: - Salary increment percentage

PerformanceRating: - Rating of an employee as per the performance at work

RelationshipSatisfaction: - How satisfied is an employee with the relationship at work

StandardHours: - Usual hours employee work

StockOptionLevel: - Is employee provided a stocks options

TotalWorkingYears: - Total experience in years

TrainingTimesLastYear : - Number of times Training aligned in last year

WorkLifeBalance: - Degree of balance in work and normal life

YearsAtCompany: - Total experience in years in current company

YearInCurrentRole: - Total time in years in current role

YearsSinceLastPromotion: - Total time in years since an employee last promoted

YearsWithCurrentManager: - Total time in years working under same manager

Out of total 1470, 237 people left the company whereas 1233 people stayed with the organization

* EDA Concluding Remark.

On plotting a graph of Attrition with respect to age, we get to know that maximum numbers of employees leaving the company are from the age group of 29 and 30

* Building Machine Learning Models.

With Logistic regression model we got 78.9% accuracy for our training data and 77.30% accuracy for test data.

With K-Nearest neighbor machine learning model, we got 89.20% for our training data which is better than our previous model and got 82.65% accuracy for our test data.

With decision tree classifier, it is showing 100% accuracy for our training data and 91.86% for our test data which is better than Logistic regression and K-Nearest neighbor classifier.

With random Forest classifier model we are getting the accuracy of 100% for our training data and 97.64% accuracy score for our test data which is the best we have so far.

* Concluding Remarks.

We have tried to check this up with plotting of ROC, AUC curve and it is saying the same thing that Random forest classifier is the best model we have out of 4 machine learning models checked.

It means we only need to hyper parameter tune the random forest classifier model but our score didn’t improved much with tuning the parameters and hence the random forest classifier model is chosen the best model for this dataset.