# Human Resource Analytics

**Chandrasekar Rajasekar - crajase**
**Dinesh Prasanth M K - dmolugu**
**Vishal Murugan - vmuruga**
**Vivek Mani - vmani2**

## 1 Data Set

Human Resources Analytics (Available at Kaggle) - contains data of about 15,000 employees.

## 2 Description

Every company wants to hire the best candidates and retain them over a long period. They provide attractive compensations in order to retain them. However, many such best employees switch over to other companies for various reasons; poor performance due to heavy workload, less promotions, low job satisfaction levels, salary. This will incur a huge loss for the company; the company has to hire a new resource and train him - takes several days to months to gain the knowledge/experience as the ex-employee.

This project aims to analyze the mindset of 15,000 employees (including people who have left the company) and predict the employees that the company is about to lose. Each employee is tested against the various features that include: Satisfaction level, Last rating, Number of projects completed, Number of years served, Work accidents, Promotions, Department, Salary

## 3 Softwares Required

R, Python - [Pandas, NumPy, Matplotlib, scikit-learn/tensorflow]

## 4 Papers to Read

[1] Kotsiantis,S. B.(2007) *Supervised Machine Learning: A Review of Classification Techniques.*

[2] Tin Kam Ho. & M. Basu.(2002) *Complexity measures of supervised classification problems, IEEE.*

## 5 Teammate and Work Allocation

All team members will initially familiarize with the data set and split records into training/test data:
**Chandrasekar Rajasekar**: Data Transformation, Logistic Regression Model
**Dinesh Prasanth M K**: Naive Bayes model, Tuning final model based on metrics of other models
**Vishal Murugan**: Cost Matrix Estimation, Support Vector Machine model
**Vivek Mani**: Feature Selection, Decision Tree

## 6 Midterm Milestone

Midterm goal is to identify the type of the attributes and apply standardization on the kaggle dataset suitable for feature selection. Next, we apply the decision tree and Naive Bayes classifier algorithm on the dataset to predict employee satisfaction level.