**HR Analytics- Analysis**

The dataset hr\_data.csv contains sample of candidates that we part of a recruitment process of client of ScaleneWorks. ScaleneWorks supports several information technology(IT) companies in India with their talent acquisition. One of the challenge they face is about 30% of the candidates who accept the jobs offers, do not join the company. This lead to huge loss of revenue and time as the companies initiate the recruitment process again to fill the workforce demand. ScaleneWorks want to find out if a model can be build to predict the likelihood of a candidate joining the company. If the likelihood is high, then the company can go ahead and offer the jobs to the candidates.

The data set contains several attributes about candidate along with a column (or variable)that indicates if the candidate finally joined to the company or not.

Here is the description of the candidates attributes

1 Candidate reference number Unique number to identify the candidate

2 **DOJ extended Binary variable identifying whether candidate asked for date of joining extension (Yes/No)**

3 Duration to accept the offer Number of days taken by the candidate to accept the offer (continuous variable)

4 Notice period Notice period to be served in the parting company before candidate can join this company (continuous variable)

5 Offered band **Band offered to the candidate based on experience and performance in interview rounds (categorical variable labelled C0/C1/C2/C3/C4/C5/C6)**

6 Percentage hike (CTC) expected Percentage hike expected by the candidate (continuous variable)

7 Percentage hike offered (CTC) Percentage hike offered by the company (continuous variable)

8 Joining bonus **Binary variable indicating if joining bonus was given or not (Yes/No)**

9 Gender **Gender of the candidate (Male/Female)**

10 Candidate source Source from which resume of the candidate was obtained (categorical variables with categories: Employee referral/Agency/Direct)

11 REX (in years) Relevant years of experience of the candidate for the position offered (continuous variable)

12 LOB Line of business for which offer was rolled out (categorical variable)

13 DOB Date of birth of the candidate

14 Joining location Company location for which offer was rolled out for candidate to join (categorical variable)

15 Candidate relocation status Binary variable indicating whether candidate has to relocate from one city to another city for joining (Yes/No)

16 **HR status Final joining status of candidate (Joined/Not-Joined)**

Questions…………

* **Build a logistic regression model to predict the probability of candidate joining the company: Assume not joined as positive case and joined as negative cases**
* **Find the significant features from the above model and build another logistic regression model with only significant features**
* Assume the following costs to find the optimal cut off probability to determine if a candidate will join or not.
  + COST OF PREDICTING “not joining” as “joining” (FPs) case is 3 times more than predicting “joining” as “ not joining” (FNs)
* **Build a confusion matrix based on the cut off probability found in previous question and report the precision and recall of the model for joining cases**
* HR want to understand the key parameters effecting the joining of candidates. So, build a decision tree with optimal parameters and provide the some rules to HR for building strategies to ensure candidates offered job most likely will join the company in future.