

# Lead Scoring Case Study Summary

## Problem Statement:

- X-Education is an education company sells online Education courses to professionals and does its marketing through many online advertisements. Company gets information through different channels and if candidates enquiring with certain education level it calls lead. Typically lead conversion is 30% of certain education. Company identifying Hot Leads on certain criteria also. Lead conversion ratio is lesser than number of enrollments. company given Target to achieve 80% of total enrollment.

## Solution Summary:

1. Reading and Analyzing Data
2. Data Cleaning
  - Includes Null value treatment and Outliers treatment.
3. Exploratory data Analysis
  - To understand the distribution of target variable and the relation of the independent features with dependent feature(Target Variable).After plotting the graphs realized that few columns have only one value and few columns are imbalance in data were dropped
4. Dummy Value creation
  - Created dummy values for categorical variable.
  - Converted few columns with yes/no to 1 and 0
5. Train Test data split
  - Split the entire data into train (70%) and test (30%)
6. Feature selection using RFE
  - Using RFE we have selected top 20 important features.
7. Feature scaling using Standard Scaling mechanism.
8. Building the model
  - Using the 20 features we have build an initial model and using the statistics generated we have finally build a model with to 14 important features where even VIF's were found good
  - We then created a data frame with the converted probability values initially assuming the probability value of more than 0.5 means 1(converted) else 0(Not converted)
  - With the above assumption calculated confusion matrix and calculated overall accuracy of the Model
9. Plotting ROC curve
  - We plotted the ROC curve for the features which looks decent with an area coverage of 89%
10. Finding Optimal cut off value.

- Calculated Accuracy, sensitivity and specificity with all probability values
- Then plotted the graph of all above metrics and found that all 3 graphs intersecting at 0.37 which has been considered as optimal cutoff value.
- Based on the new cutoff value evaluated all the metrics again (81.5%, 79% and 83%) and noticed that conversion rate is close to 80 % which means the model is decent enough

#### 11. Computing metrics

- Calculated Precision and Recall values which are close to 80 % and 70 % respectively.
- Based on Precision and Recall trade off we have got the cutoff value close to 0.41

#### 12. Evaluating the model with test set

- Implemented the learnings from the train set on test set and found that, Accuracy, sensitivity and specificity are 81%, 77% and 84% which are almost close to the train set values.