

# Lead Scoring Case Study

BY - Nikhil Kawale  
Pritam Shetty  
Dhathri S

# Problem Statement

- ▶ X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

# Approach

- ▶ Reading the data
- ▶ Inspecting the Data
- ▶ Data Cleaning
- ▶ Exploratory Data Analysis
- ▶ Creating Dummy Variables
- ▶ Splitting the data set to Test and Train
- ▶ Scaling of Data
- ▶ RFE
- ▶ Model Building
- ▶ Model Evaluation

# Technical Aspects of Case Study

- ▶ Under Categorical Variables, Select data is converted to NULL value.
- ▶ Column with missing value greater than 40% is dropped.
- ▶ For column with categorical data, NULL values are replaced by Mode value of the column.
- ▶ Outliers Treatment - Q1 - 5% and Q3 - 95%.
- ▶ `train_size = 0.7`, `test_size = 0.3` of the complete dataset.
- ▶ Feature scaling done using `StandardScaler`.
- ▶ ROC curve (area = 0.88)
- ▶ The optimum value selected is 0.35.

- ▶ Comparing the values obtained for Train & Test:

- ▶ Train Data:

- ▶ Accuracy : 81.24%

- ▶ Sensitivity : 80.7%

- ▶ Specificity : 81.58%

- ▶ Test Data:


- ▶ Accuracy : 80.20%

- ▶ Sensitivity : 81.09%

- ▶ Specificity : 79.70%

# Business Related Observation

- ▶ The majority of the leads' most recent action was Email Opened.
- ▶ About 70% of leads with an SMS sent as their most recent activity converted.
- ▶ Maximum leads are generated from Unemployed but maximum lead are converted from Working Professional.
- ▶ Following are the categorical variables the have high probability of lead getting converted
  - ▶ Lead Source
  - ▶ Last Activity
  - ▶ What is your current occupation

- 
- ▶ Following are the variables which contribute the high probability of lead getting converted
    - ▶ Lead Source\_Welingak Website.
    - ▶ Lead Source\_Reference.
    - ▶ Last Activity\_Had a Phone Conversation

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect. The shapes are concentrated on the right side of the image, with some extending towards the left.

Thank You