

# Lead Score Case Study



Submitted by:  
Pureesh Reddy  
Rijul Gaur  
Hardik Nigam

### **Problem Statement :**

- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

### **Business Goal:**

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with 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%.

# Strategy Followed

- Source the data for analysis
- Clean and prepare the data
- Model Building
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics -Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

# Methodology followed

## **Data Sourcing , Cleaning and Preparation**

- Data Sourcing , Cleaning and Preparation
- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis

## **Feature Standardization.**

- Feature Scaling and Splitting Train and Test Sets
- Feature Scaling of Numeric data
- Splitting data in train and test set.

## **Model Building**

- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculation of various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

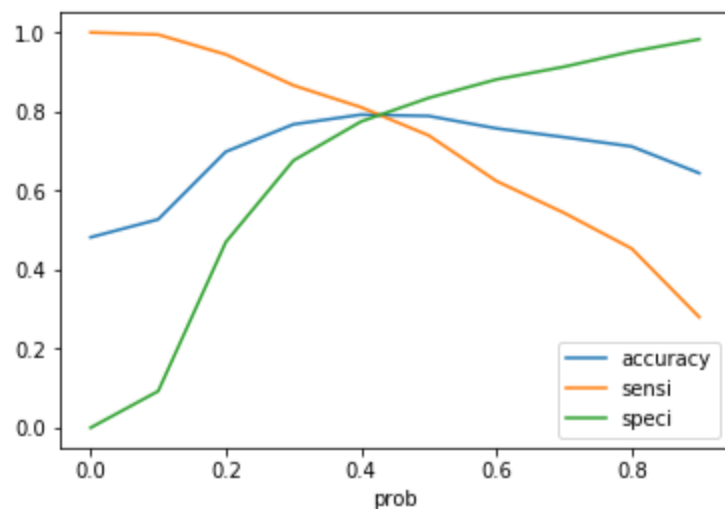
## **Result**

- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

# Variables Impacting the Conversion Rate

- Total Visits
- Total Time Spent On Website
- Lead Origin –Lead Page Submission
- Lead Origin –Lead Add Form
- Lead Source -Olark Chat
- Last Source –Welingak Website
- Last Activity –Email Bounced
- Last Activity –Not Sure
- Last Activity –Olark Chat Conversation
- Last Activity –SMS Sent
- Current Occupation –No Information
- Current Occupation –Working Professional
- Last Notable Activity –Had a Phone Conversation
- Last Notable Activity -Unreachable

# Model Evaluation -Sensitivity and Specificity on Train Data Set



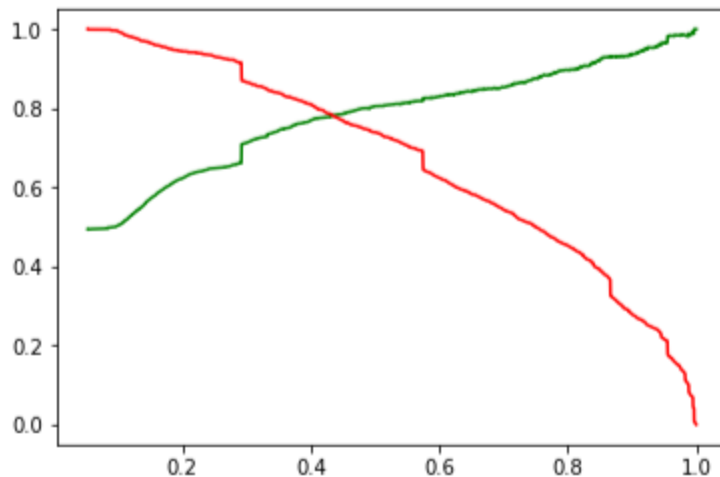
Confusion Matrix:  $\begin{bmatrix} 1823 & 489 \\ 444 & 1705 \end{bmatrix}$

Accuracy -79%

Sensitivity -79.33%

Specificity -78.84%

# Model Evaluation-Precision and Recall on Train Dataset



Confusion Matrix:  $\begin{bmatrix} 1852 & 460 \\ 479 & 1670 \end{bmatrix}$

Precision-78.4%

Recall-77.7%