



LEAD SCORING CASE STUDY – SUMMARY REPORT

BACKGROUND CONTEXT:

The case study aims to build a machine learning model and perform analysis for X-Education company to improve their lead conversion rate.

APPROACH FOLLOWED:

The analysis is performed on the past data (~9000 customers).

- Understanding of data attributes with help of data dictionary
- Data inspection, cleaning, NULL/Unknown treatment, outlier treatment and etc
- Exploratory data analysis to derive useful insights
- Data Preparation, dummy data creation from categorical variables
- Model creation using logistic regression
 - o Training 70% of data (fit and transform) and Test (transform) on 30% of data
 - o Model evaluation.
- Model evaluation, predictions and measure effectiveness

LEARNING / INSIGHTS GATHERED DURING THE PROCESS OF CASE STUDY:

- The final model obtained has 11 features in total from which following features are mostly impacting the conversion of leads.
 - o Total time spent on the website
 - o Lead origin by adding form.
 - o Additionally, while EDA found that the lead Origin by adding forms has high correlation with leads who has approached via reference.
 - o Working professional
 - o Lead received from Olark chat

Features/ attributes negatively impacting conversion:

- o Last notable activity as Olark chat conversation.
 - o Instructed as not to email
 - o Have not provided occupation information
- While EDA additionally found that, current house-wife represent very small set of leads who are most likely to get converted.

MODEL EVALUATION:

- Lead score assigned in the range of 0-100 for each lead
- Results for Model evaluation has 0.35 probability as cut-off for optimum value of accuracy, sensitivity and specificity. X-Education should contact the leads who have 35 or more lead score in order to obtain best conversion rate (close to 80%).
- Based on precision-recall trade off the plot gives ~0.4 as probability precision and recall remains as 74% and 77% for the model.
- Company can change the cut-off based on business strategy. When we need to get aggressive on business and get more conversions as much as possible, we can rely on high sensitivity. And when company is in conservative approach and only spend efforts in case to reduce non-conversion then they will rely on high precision.

RECOMMENDATION:

Positive Cases:

- Focus more on the leads who **spent more time on the website** and they have **added the form** which indicates that they have more interest and eager on taking up the course.
- Generally, it is seen more **working professionals** they need to upgrade themselves in order to professionally grown in their career.
- Generally, conversion rate is higher for the leads which they have received through **referrals**

It would be win-win situation for sales people and need less efforts to convert above category of leads to paying customers. Although small population good results of conversion rates for housewives. It would be worth reaching out to these segments of customers.

Negative cases:

It would not make sense to spent time on those leads who have instructed not to be contacted.

