

Summary of the Lead scoring case study

We are given a case study to analyse the low conversion rate for the organisation X Education offering online courses for industry professional. It was observed that when a buyer or prospect visit the website, they either browse the course offered by the website or fill up the form or watch some videos. Any customer filling up the form is considered as a lead along with past referrals.

Based on the leads, the sales team of X Education start making calls or reach out to the customers through various other channels. Through these various modes of conversations, some of the leads get converted and some may not.

As the conversion rate are around 30%, so it pretty worrying for the senior management and looking for option to increase this.

As per our analysis we did following steps

- 1) **Data Cleaning :** We dropped all the columns having more than 40% null vales, having high number of values as 'Select' or 'No'. We also dropped rows having more 30% null values and columns like Country having mostly having values as 'India'
- 2) **EDA:** As part of exploratory data analysis, we found lots of columns having categorical values and we couldn't infer much from it. The numerical values looked good and relevant and having not much outlier.
- 3) **Dummy Variable creation:** After creating dummies values for numerical features, we scaled the data using MinMax Scaler.
- 4) **Train Test Split the data:** The data was split in the ratio of 70% and 30% respectively for Train and test data.
- 5) **Model Building:** Using RFE, we got 15 relevant features through auto selection. Then, we eliminated features having VIF and P-values one by one.
- 6) **Model evaluation:** We created confusion matrix and derived optimum cut off value using ROC- AUC curve. With this optimum value, we derived Accuracy, Sensitivity and Specificity which came out to be around 80%.
- 7) **Prediction:** Prediction was done on test data with optimum cut off value as 0.42 and it was found to be have accuracy, sensitivity and specificity as 78%, 78% and 76% respectively.
- 8) **Precision and recall:** The Optimum cut off as 0.42, the precision and recall came out to be 80% and 73% respectively.

It was found that variables that matter most to the potential buyer or the prospects are

Total Visits
Total time spent on the website
Lead source
Lead Origin
Last activity
Current occupation

Keeping all these points in mind, X education can devise their strategy to convert the leads, prospects, potential buyers to their customers and increase their sales.

They should make calls to leads coming from references, working professionals, leads who have spent more time on the website, sources like Olark chat and to leads to whom SMS was sent.