

Lead Score Case study

The problem statement was to increase the conversion rate of leads into conversion at X Education. Currently the conversion rate at X Education is 30%. The CE wants to increase the conversion rate up to 80%.

The goal of case study is to build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that lead is cold and will mostly not get converted.

We imported the data set into a pandas data frame and treated null values and 'select' values in the data set. And We performed Exploratory data analysis (EDA) to understand the data.

We used RFE to extract best 16 variables from the list of all variables in our data set.

Then we built a logistic regression model on the data and calculated the accuracy and other metrics. Some of the variables had higher p-value in magnitude indicating a higher impact on conversion rate. Based on p-value we removed some of the variables and calculated final model.

The final model had an overall accuracy of 84%, sensitivity of 75%, specificity of 90% and False positive rate of 10%.

We then built a ROC curve to find optimal cutoff point to predict where the lead will convert or not. We found 0.35 to be the optimal point to set as cutoff for final prediction. We then evaluated the final model on train and test data.

We found that variables like lead_origin_lead_add_form, lead_source_welingak_website and lead_source_olark_chat have significant positive impact on lead conversion possibility.

Also, variables like lead_quality_worst, lead_quality_not_sure and lead_quality_might_be have high negative impact on lead conversion possibility.

In order to increase conversion rate to 80%, X Education must focus on getting more leads from lead_add_form, welingak and olark_chat. And also, aggressively contact leads from these sources to convert them.

Similarly, X Education must try to get better quality of leads. Lead quality is significantly impacting the conversion rate. Lead quality of worst, not sure or might be have negative impact on conversion rate. So, it is better to avoid leads from with these lead qualities.

The sales team should be given leads from sources which positively impacts conversion rate and avoid leads which are negatively impacting the conversion rate.