Lead Scoring Case Study

SUMMARY REPORT

Prepared by:
Anish Lakhotiya, Kumar Shubham and Ankita Karekar

Contents

Problem Statement

- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once people land on the website and fill up a form for the course, they are classified to be a lead. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. to covert these leads to become students. The typical lead conversion rate is around 30%.

- To improve the lead conversion rate, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the sales team will now be focusing on more on communicating with the potential leads.
- The company is looking forward to build a model for assigning a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance.
- The target lead conversion rate to be around 80%.

Goal of the Case Study

- Building 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 the lead is cold and will mostly not get converted.

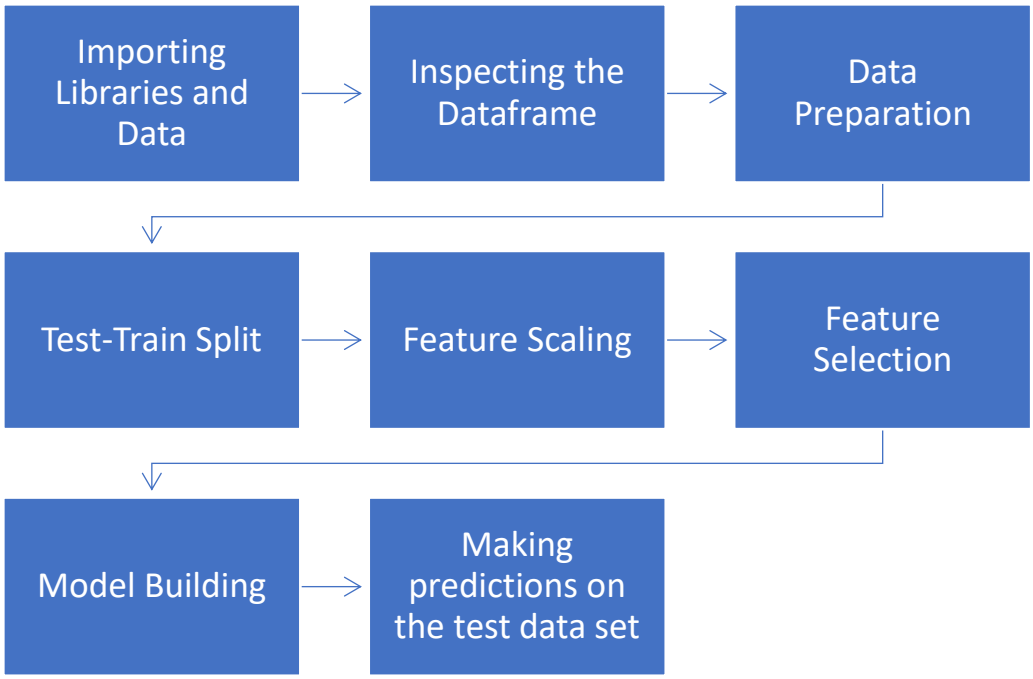
    a) The following business problems to be answered: a) Top 3 variables in the model which contribute most towards the probability of a lead getting converted.

    b) Top 3 categorical variables in the model which should be focused the most on in order to increase the probability of lead conversion.

    c) Formulating a good strategy to address almost all of the potential leads those have been predicted by the model, to be converted.

d) Suggesting a strategy to minimize the rate of useless phone calls (in conversion process), unless it's extremely necessary.

- ## Model building workflow



Outcome

Following are the predictor variables shortlisted for building the model after considering the p value and VIF. **Feature**	**Beta Value**
Do Not Email	-1.256546
TotalVisits	0.169341
Total Time Spent on Website	1.079498
Lead_Origin_API	-3.444542
Lead_Origin_Landing Page Submission	-3.460546
Lead_Source_Google	0.284136
Lead_Source_Olark Chat	1.046174
Lead_Source_Reference	0.688108
CO_Working Professional	2.855265

It is evident that leads originating through API or Landing Page Submission, as well as individuals with a current occupation as Working Professionals, exert the most significant influence on the overall score. Notably, Lead Originating through API or Landing Page

Submission negatively impacts the score, whereas Current Occupation Working Professionals have a positive impact on the score.

Model Performance

The optimal probability cut-off for predicting lead conversion was found to be 0.25.

The final model demonstrated an accuracy of 77.23% on the test dataset, with a sensitivity of 77.8%.

In light of these results, it is recommended that the sales team prioritize their calls based on the Score.

The focus should be on contacting prospects with the highest scores initially, followed by working through the list in descending order of the score.

This strategy aims to maximize the likelihood of successful lead conversion.