

Summary – Lead Scoring Case Study

Problem Statement:

X Education sells online courses to industry professionals. X Education needs help selecting the most promising leads, i.e., the leads most likely to convert into paying customers. The company needs a model wherein a lead score is assigned to each lead such that the customers with higher lead scores have a

higher conversion chance and the customers with lower lead scores have a lower conversion chance. The CEO has given a ballpark of the target lead conversion rate to be around 80%.

Our Goals of the Case Study:

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.

To adjust if the company's requirement changes in the future, you will need to handle these as well.

To solve the problem:

we take the following major steps:

Step1: Data understanding, Data Cleaning and Visualization

Step2: Model Pre-processing

Step3: Machine Learning Model building - Logistic Regression

Step4: Machine Learning Model Analysis with Training Data

Step5: Adding Lead Score to the Training Data

Step6: Making a Prediction using the Final Model with Test Data

Step7: Machine Learning Model Analysis along with Test Data

Step8: Conclusion/Inference from Machine Learning Model

Summary:

- As we have checked Sensitivity, Specificity, Precision, and Recall as Metrics, we have considered the optimal cutoff limit as 0.36 for calculating the final prediction.
- Accuracy, Sensitivity, and Specificity values of the test set are approximately closer to the respective values calculated using the trained set.

For Train data	For Test data
<ul style="list-style-type: none">• Accuracy: 76.40%• Sensitivity: 72.83%• Specificity: 78.61	<ul style="list-style-type: none">• Accuracy: 76.11%• Sensitivity: 72.83%• Specificity: 78.61%

- Also, the lead score calculated in the trained set of data shows the conversion rate on the final predicted model.
- Hence overall this model seems to be good.
- Firstly, need to sort out the best prospects from the leads you have generated. 'TotalVisits', 'Total Time Spent on Website', and 'Page Views Per Visit' contribute most towards the probability of a lead getting converted.
- Focus on converted leads.
- Hold question-answer sessions with leads to extract the right information you need about them.
- Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.
- API and Landing Page Submission have around 35-40% conversion rate but the count of lead originating from them is considerable.
- Lead Add Form has a more than 90% conversion rate, but the count of leads is not very high.
- Lead Import and Quick Add From are very low in the count.
- Direct Traffic and Google are generating a high number of leads; should be focused on their conversion.
- Welingak and Reference have higher conversion rates and hence should be focused on generating more leads.

- leads from Olark Chat and Organic Search should be focused on conversion.
- Most of the leads have emails opened and SMS Sent. More focus should be on leads opening emails.
- The conversion rate for SMS sent is very high and hence they should be targeted for lead conversion.
- Working professionals have high conversion. focus should be on them for more leads.
- Most of the leads are from the Unemployed section, the focus should be increasing conversion from them.
- More leads are from those who do not ask for a free copy of Mastering Interviews. Can be focused on conversion.
- More leads are from those who have modified their account or opened emails. Thus, the focus should be on their conversion.
- SMS sent have high conversion.

Conclusion :

As seen from our final model, the following are the Parameters that help us to predict the probability of Lead conversion and hence increase the Chances of Lead Conversion from 30% to higher than 70 %:

- Lead Origin_Lead Add Form 3.306135
- Last Activity_Had a Phone Conversation 1.519508
- Total Time Spent on Website 1.076119
- Lead Source_Olark Chat 0.340387
- Lead Source_Direct Traffic -0.322981
- Lead Origin_Landing Page Submission -0.629927
- Last Activity_Form Submitted on Website -1.055841
- Last Activity_Email Link Clicked -1.060514
- Last Activity_Converted to Lead -1.820172
- Last Activity_Olark Chat Conversation -2.120415
- Last Activity_Email Bounced -2.582260