

Lead Scoring Case Study Report

By Sagir Mehmood and Vaibhav Sah

Problem Statement

The X Education company is an Online Education Company that wants a model using which the Lead conversion rate can shoot up to 80% from current 30%.

Solution Approach

As the data collected by the company is highly categorical, and we require a predictive analysis model for determining the most relevant variables for identifying the most promising leads, so the Logistic Regression Model seems to be the perfect approach to come to any conclusion.

Exploratory Data Analysis

- Removed highly skewed Parameters
- Removed highly correlated Factors

Modelling

Once the data got cleaned, we splitted the data into Train(70%) and Test Data(30%). Then we created a model on the basis of the training data. We filter down the number of parameters to 8 as below:

Variables	Coefficient
const	-1.4575
Total Time Spent on Website	1.0201
API	0.4476
Lead Add Form	4.0795
Converted to Lead	-1.0033
Page Visited on Website	-0.362
SMS Sent	1.3736
Marketing Management	0.4415
Operations Management	0.4317

Here the variables with positive coefficient are those that are good for identifying the leads. We shall avoid the leads showing high values for parameters with negative coefficients.

Results:

We created a model using above variables with

- **Accuracy** as 78%
- **Sensitivity**, i.e, the measure of correctly identifying leads as ~80%
- **Specificity**, i.e. the proportion of negatives which are correctly identified as ~76%

Our Learning

We learnt that what actually sometimes comes in visualization is not much relevant. On looking at graphs initially it seems that factors such as **Landing Page Submission**, **Google Search** etc are high conversion factors. But after calculating their correlation with other factors, we found out that they were just the highly correlated factors and are not contributing much to lead conversion. Instead the sources from API or those filling the Lead Add Forms are the real Hot Leads.

If the people are visiting more pages on your website, it does not always mean they are probable customers. In our model, more the pages visited by customers, less is their probability of conversion.

Company has been calling again the leads that are already converted. This shall be strictly avoided.

So if the company really wants to cut down some cost and want to reduce their Sales team efforts, stop focusing on Converted leads and the people who are just randomly browsing the site pages. Instead focus on the time spent rather than pages visited on the site.

Also check for those who have sent the SMS as their recent activity. They are the ones to focus as the target leads.