

Lead scoring case study

Summary:

An education company named X Education sells online courses to industry professionals on a social marketing campaign. The company markets its courses on several websites and search engines like Google. As the people surf through the internet and come across the course via any means such as videos or any other form of add, those people might end up filling the form. As they start to respond to the adds the employees from the X education starts to give them a call back or write email etc.

Let there be N number of people filling out the form, out of which only few accepts to actually join the course later. Those people will be considered as leads.

The people joining the course which be the referred as “converted Leads” as they have successfully claimed for the course.

The people who have not joined the course are the “Unconverted Leads”.

Required Model



Pictorial representation

If the score is more than the Probability of converting the lead into converted lead is more .

So we have to identify the most promising and potential leads.

In the case study we drop the unnecessary values

The typical lead conversion rate at X education is around 30%.

To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads' as shown in the fig.

The steps we follow are as follows:

1. Data gathering : Loading and observing the data provided by the company
2. Data cleaning: Duplicates removal null value treatment, unnecessary column elimination etc.
3. Performing EDA: Univariate and bivariate analysis and heatmap for numerical and categorical data.
4. Data preparation: Outlier treatment and feature- standardisation
5. Model building : Performing pre-requisites for RFE and logistic regression.

The lead data set is around 9000 points , consisting of various attributes such as lead sources , Total Time Spent on Website, Total Visits, Last Activity, etc. which can helpful in deciding if the lead is going to be converted or not

After running the model on the test data we obtain following figures:

- Accuracy: 95.53%
- Sensitivity: 96.67%
- Specificity: 90.60%

We can see that it is a good model .

-Submitted by

Anjali Singh

Ajinkya Shakkarwar

