

Lead Scoring Case Study Summary

Problem Statement:

Create a machine learning model for an education company, having online platform for their education courses which predicts and assigns a lead score to each lead based on different variables available from historical leads. This is a logistic regression problem due to predicting the classification of the leads as converted or not, with probability of conversion to be predicted.

There are a lot of leads in the initial stage but only a few of them are converted into paying customers.

The most numbers of leads are from INDIA and in terms of city highest number are from Mumbai.

There are a few columns in which there is a level called 'Select' which basically means that the student had not selected the option for that particular column which is why it shows 'Select'. To get some useful data we have to make compulsory selection. Likewise, Customer occupation, Specialization, etc.

The high number of total visits & Total time spent on platform may increasing the chances of lead to be converted.

The leads are joined course for Better Career Prospects, most of having Specialization from Finance Management. Leads from HR, Finance & marketing.

This method was also used to recheck and a cut off of 0.37 was found with precision around 68 % and recall around 81% on the test data frame .It was found that the variables that are most significant:

1. The total time spends on the website.
2. Total number of visits.
3. When the lead source was:
 - a) Google
 - b) Direct traffic
 - c) Organic research
 - d) Welingak website
4. When the last activity was:
 - a) SMS
 - b) Olark chat conversation
5. When the lead origin is lead add format.
6. When their current occupation is as a working professional.

With this in mind the X education company can grow as they have a very high chances to get almost all the hot leads to change their mind and buy their courses.