

Lead Scoring Assignment Summary

Question-

Problem Statement

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. A typical lead conversion process can be represented using the following funnel:



Lead Conversion Process - Demonstrated as a funnel

As you can see, there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Summary-

About this Lead conversion Assignment. The threshold of converting lead into positive mark is about 30%, and the company wants us to convert this leads into 80%

So overall progress for building the model is as follows:-

First the work is done by us is importing the dataset through the use of Pandas, then looking at the insights of data is done by us so that we can look the type of data inside the dataset

Then we have performed EDA into it in which lots of noise which is already present in the dataset is handled by our team- like null value handling, imputation, Univariate, Bivariate Analysis in which so many graphs were made to look the relation between the variables and looking at the false note and outliers

Then comes the step of making dummy variables (one hot encoding) in which lots of categorical values are converted as dummies

Inside the categorical column data was present in so many categories, so grouping was done to lower the traffic

Then comes the job of splitting the data into Train and test we have splitted the data in the ratio 70 – 30%

Then scalling is done by the help of fit transform part, after this model building process is started and during the model building regular VIF selection also continuously proceeded final model is selected which have 13 columns

Confusion matrix of the model comes out to be as

[[3384 618]

[347 2119]]

Overall recall comes out to be 85% after taking optimum threshold of 0.3%

And prediction done on the test data which comes out to as 84% so overall conclusion of this assignment is we have found the variable which would help in increasing the lead.