

Summary

This X Education tries to have more industry professionals to join its courses. Based on the given data of Prospect ID, occupation of the subjects, conversion rates of the subjects, etc and lots of dummy variables, the analysis done.

At very first the data is cleaned in terms of missing values and outliers. After that the dummy variables are created based on their corresponding categories. Now, based on explanatory data analysis it is obtained that there are several irrelevant variables and those variables were dropped. The train and test split has been done at 70% for train data and 30% for test data. Using Recursive Feature Elimination top 15 relevant variables are selected. And, for finally a model is selected where all the variables are significant (p value less than level of significance of 0.05) and variance inflation factor less than 5. Now, using ROC curve a optimal cut off value of 0.42 is obtained.

Now, the accuracy, sensitivity and specificity obtained about 80% on train data. The prediction is made on test data where accuracy came as about 77%, the sensitivity came as 76% and recall came as about 78%.

From the analysis, the top three variables in the model which contribute most towards the probability of a lead getting converted are, “Total Time Spent on Website”, “What is your current occupation_Student”, and “What is your current occupation_Unemployed”.

The top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are, “Last Activity_SMS Sent”, “Lead Source_Olark Chat” and “Lead Origin_Lead Add Form”.

It is obtained that the lead origin as well as lead source are significant dummy variables in order to increase the probability of lead conversion. In addition, the X-education team should target those people who spent higher time on the website. Among these people they need to focus mainly only those people who are not students and who are unemployed or employed, but want to have more business knowledge. The interns should reach them more competitively and explain them how good the X-education is compared to other institutions.

It was found that the variables that mattered the most in the potential buyers are (In descending order):

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - 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.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

X.....X.....X.....X