LEAD SCORING SUMMARY

PROBLEM DESCRIPTION

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%. X Education gets a lot of leads, its lead conversion rate is very poor.CEO has given a ballpark target of lead conversion rate to be around 80%.

APPROACH

We choose logistic regression to calculate the lead rate. Bellow steps are followed to solve this problem,

1.cleaning data: The data was partially clean except for a few null values and the option select had to replaced with null value. Although they were removed while doing dummys.

2.EDA: A quick EDA was done to check the condition of our data.

3.dummy variable: The dummy variables were created and later dummys with 'not approved' were removed.

4.train-test split: The split was done at 70% and 30% for train and test data.

5.model building: RFE was done and VIF also done.

6.model evaluation: A confusion matrix was made. ROC curve also done.

7.prediction: Prediction was done on the test data frame with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.

8.precision: This method we used to recheck.

The variables that mattered most in the potential buyers are

1. The total time spend on website.

2. Total number of visits

3. When the lead search was

   Google, direct traffic, organic search, welingak website

4. When the last activity was sms or olark chat.

5. When the lead origin is lead add format.

   Keeping these pointers in mind X education can flourish as they have high chance to get all potential buyers to change their mind and buy their courses.