

SUMMARY

- X Education , An education company named sells online courses to industry professionals. Many interested professionals land on their website.
- The company markets its courses on several websites 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
- 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 but its lead conversion rate is very poor . 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. We will help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- We are required to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.
- To solve this problem we need to use Logistic Regression Model to probability of lead to join the course
- The provided test data has lots of missing values so we needed to remove first columns with more than 45% null values then null values treatment and combining low frequency values in categorical column and then after doing Univariate and Bivariate analysis we found that
 1. Total Time Spent on Website
 2. Last Activity_SMS Sent
 3. Lead Source_referenceThese are the top three variables in my model, that contribute towards lead conversion are
- The top three variables in my model, that should be focused are:
 1. Last Activity_SMS Sent
 2. Last Source_Direct Traffic
 3. Lead Source_Olark Chat
- Our Logistic Regression Model is decent and accurate enough, with 92.78% Accuracy on Test Set, 91.98 % Sensitivity and 93.26 % Specificity.
- We can vary these parameters by varying the cut-off value and thus predict Hot leads based on scenarios like availability of extra resources and vice-versa.

- Conclusions from these model are

X Education Company needs to focus on following key aspects to improve the overall conversion rate

1. Increase user engagement on their website since this helps in higher conversion
2. Increase on sending SMS notifications since this helps in higher conversion
3. Get Total visits increased by advertising etc. since this helps in higher conversion
4. Improve the Olark Chat service since this is affecting the conversion negatively