

Summary Report

The analysis and our findings are on education company named X Education on ways in which more and more customers finally take their online courses. The dataset provided gave us many useful information on the potential customers total time spent on website, total visits, their occupation, last activity, specialization, occupation etc. For data cleaning purpose there were many categorical variables where only one value was majorly present for all the data points. These include Do Not Call, Search, Magazine, Newspaper Article, X Education Forums etc. Since practically all of the values for these variables are No, it's best that we drop these columns as they won't help with our analysis. There were also columns like City, Country that were removed variable City won't be of any use in our analysis as it doesn't matter which city, country the lead getting converted belongs to.

So once Data Cleaning was done, we use this cleaned data to perform analysis and build Machine Learning Model in Logistic Regression. The Logistic Regression built on the cleaned data gave us a good accuracy score of 79%. Based on the model built we found from the dataset provided which variables are more important in predicting whether the potential customer finally get converted and buy course online are as follows:

- a) Total Time spend on the website: More the time is spent on the website, shows that he/she is interested in buying the courses offered. It also shows that they are also doing investigation from their end and looking at the various curriculum, modules available in the courses, potential career prospects etc.
- b) Total Number of Visits: This variable also influences whether or not the customer will potentially be interested in taking the online course or not.
- c) When the Lead Source was (Google, Direct Traffic, Organic Search, Welingak Website)
- d) When the last activity was SMS , Olark Chat Conversation
- e) When the lead origin was from Lead Add form
- f) When the current occupation is working professional

The marketing team should focus on targeting the above customers if they want to select the most promising leads, i.e. the leads that are most likely to convert into paying customers for their courses offered.