

Lead Scoring Case Study: Final Summary

Team: Suresh Kumar Singh, Shoyeb Patwekar & Sourabh Bhatt

This analysis is done for X Education and to find to get more industry professional to join their courses. The basic data provided, gave us a lot of information about how the potential customers visit the site, the time they spent there, how they reached the site and conversion rate.

The following steps are used

- 1) **Cleaning Data:** The data was partially clean except for a few null values and the option selected had to be replaced with a null value since it did not give us much information. Few of the null values were changes to "Not Provided" so as to not lose much data. However, they were later removed while making dummies. Since, there were from India and few from overseas, the elements were changed to "India", "Outside India" and "Not Provided"
- 2) **EDA:** It was performed to check integrity of data. It was found that a lot of elements in the categorical variables were inappropriate. The numerical values seems good as compare to categorical and very minimal outliers were found
- 3) **Dummy Variables :** The dummy variables were created
- 4) **Standardization :** For numeric vales, we have used the MinMaxScalar
- 5) **Train – Test Split :** The Train – Test spilt data was done in the ratio of 70:30
- 6) **Model Building:** Firstly, RFE was done to attain the top 15 relvant variables. Later the rest of variables were removed manually depending on the VIF values and P Values with common standard $VIF < 5$ and $P - Value < 0.05$
- 7) **Model Evaluation:** A confused matrix was made. Later on the optimum cut off vale using ROC curve was used to find the accuracy, sensitivity and specificity which came around 90+ % each
- 8) **Prediction:** Prediction was done on the test data frame and with an optimum cut off with accuracy, sensitivity and specificity 90%
- 9) **Precision – Recall :** precision approx 89% and recall approx 91 % on the test data frame

It was found that the below listed variables that matters the most in the potential buyers

1. The total Time spent on website
2. Total number of Visits
3. When the lead source were Olark chat and Website
4. When the last activities were SMS and Olark Chat conversation
5. When the lead origin is Lead add Form
6. When The current occupation were – Working Professionals, Student, Unemployed and other

Keep the above mentioned recommendations in mind the X Education can increase all the potential buyers to make their mind and buy their courses to achieve their target to convert at least 80 % leads into subscription