

Summary

The objective of this case study is to identify variables or attributes that the X Education company can target in future in order to get more leads and conversions. The basic data shows information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The following are the steps used:

1. Cleaning data: The data was partially clean except for a few null values and we tried to impute missing values based on high frequency variables. Particularly in specialization column null values are assigned as Student in order to retain information.
2. EDA: We have done both numerical as well categorical variable analysis. The numeric values analysis, we found few outliers and these were removed using percentiles method.
3. Dummy Variables: The dummy variables were created for columns like 'Lead Origin', 'What is your current occupation', 'City' etc. For scaling the data, we used the StandardScaler.
4. Train-Test split: The split was done at 70% and 30% for train and test data respectively.
5. Model Building: Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with $VIF < 5$ and $p\text{-value} < 0.05$ were kept).
6. Model Evaluation: A confusion matrix was made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity.
7. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.39.
8. Precision – Recall: Precision was around 90% and recall around 86% on the test data frame. It was found that the variables that mattered the most in the potential buyers are 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.

These indicates that X Education can focus on these attributes to have a very high chance to get almost all the potential leads to get converted