

Summary

This analysis is done for X Education in order to identify ways to draw more industry professionals to their courses. The basic data provided offered us a lot of information about how potential clients visit the site, how long they stay there, how they got there, and the conversion rate.

The steps are as follows:

Data cleaning:

- Columns with >20% null values were dropped as the columns would have affected the analysis
- The columns with lesser null values, the rows were deleted as it would not affect quantity of sample significantly.

EDA

- Data imbalance was discovered: just 48% of leads converted.
- For categorical and numerical variables, univariate and bivariate analysis was performed.
- The amount of time spent on a website has a beneficial impact on lead conversion.

Data preparation:

- Dummy variables were created for categorical variables with more than 2 categories.
- Data was split as train and test set in the ratio of 70% and 30%.
- Features were scaled using Standardization.
- Few columns were dropped as they had high VIF implying multicollinearity.

Model Building:

- A confusion matrix was constructed. Later, the best cut off value (using the ROC curve) was determined to be around 75%, 90%, and 60% for accuracy, sensitivity, and specificity, respectively.

Recommendations:

- More money can be spent on advertising and other aspects of the Welingak website.
- Incentives/discounts for providing references that convert to leads, to encourage more references.
- Working professionals should be aggressively targeted because they have a high conversion rate and have the financial means to pay higher rates.