

Summary

Analysis is done for X Education and to find ways to get more industry professional to joins their courses. The data provided us information about how the potential professional visit the site, time spent, how they landed the site and the conversion rate.

Following steps are used:

1. **Cleaning Data:** The data was partially clean except for few null values and the option select had to be replaced with a null value as it didn't fetch required information. Few null values were changed to 'not provided' so as not to lose data. Although they were later removed whilst making dummies. There were from India as well as from outside, the elements were changed to 'India', 'Outside India' and 'not provided'
2. **EDA:** A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irreverent. The numeric value seems good and no outliers were found.
3. **Dummy Variables:** The dummy variables were created and later on the dummies with 'not provided' elements were removed and for numeric value we used the MinMaScaler.
4. **Train-Test split:** The split was done at 70% and 30% for train and test data respectively.
5. **Model Building:** RFE was done to attain 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 which came to be around 81%.
7. **Prediction:** Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.
8. **Precision - Recall:** This was use to re-check and a cut of 0.41 was found with Precision around 73% and recall around 76% on the test data frame.

It was found that the variables that mattered the most in the potential enroller are (In descending order):

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.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential enroller to change their mind and enroll their courses.