Summary

This analysis is done for X education and go to find ways to get more industry professionals 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 spend there, how they reached the site and the conversion rate.

The following steps are used:

1.Data Cleaning:

The data was partially clean except for a few null values and option select had to be replaced with a null values were changed to "not provided" so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed t '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 categorical variables were irrelevant. The numeric values seems good but few upper range outliers were present in the data so we treated these outliers.

3.Dummy Variables:

The dummy variables were created and later on the dummies with 'not provided' elements were removed. we used 'standardScaler' for scaling.

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 the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-values (The variables with VIF<5 and p-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, specificity which came to be around 80%.

7.Prediction:

Prediction was done on the test data frame with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.

8.Precision -Recall:

This method was also used to recheck and a cut off of 0.42 was found with precision around 72% and recall around 80% on the test data frame.

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

1. Total time spend on the website ..
2. Total no of visits.
3. When the lead source was Welingak website.
4. When the last activity was SMS.
5. When the Lead origin is Lead add format.
6. When their current occupation is 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 buyers to change their mind and buy their courses.

X-----X-----X------X