

## Summary

This analysis is done for X Education and 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 are the steps used:

Cleaning data:

- The data was partially clean except for a few null values, we treated the null values by deleting columns which have more than 40% of null values and rows having null values less than 2%. We also replaced the "Select" values to NaN. We also treated the null values by imputing Mean/ Mode values

EDA:

- EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. We also checked and treated the outliers present in our data set. We also did Univariate and Bivariate Analysis by deleting the columns which had same value.

Train-Test Split:

- The split between the train and test data was done at 70% and 30%

Scaling the features:

- The dummy variables were created and later on the dummies with 'not provided' elements were removed. For numeric values we used the Min-MaxScaler.

Model building:

- Firstly, RFE was done to attain the top 25 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).

### 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 85% each.

### Prediction:

- Prediction was calculated on the test data frame and with an optimum cut off as 0.37 with accuracy, sensitivity and specificity of 80%.

### Precision:

- This method was also used to recheck and a cut off of 0.37 was found with Precision around 86%

It was found that the variables that mattered the most in the potential customers in enrolling our course are:

- The total time spend on the Website.
- Page Views Per Visit
- When the lead source was:
  - Google
  - Direct traffic
  - Organic search
  - Welingak website
  -
- When the last activity was:
  - SMS
  - Olark chat conversation
- When the lead origin is Lead add format.
- 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 buyers to change their mind and buy their courses

