

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:

1. Cleaning data:

- Data was not very clean there were many fields with more than 35% null values
- I have dropped the columns with more than 35% null values
- Also for the columns having less null values imputed them with either mean() , mode() or median() based on outliers present in data.

2. EDA:

- I have performed EDA to check and not take columns having skewed towards one side of the population.
- Also checked outliers present on some of the fields.

3. Dummy Variables:

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

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.
- 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.
- Model evaluation was done with below metrics :
- Accuracy : 81%
- Sensitivity=85%
- Specificity = 78%

7. Prediction on Test Set:

- Finally I performed predictions on the test set and checked evaluation metrics which came as below.
- Accuracy : 80%
- Sensitivity : 82%
- Specificity : 78%

8. Conclusions :

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

1. The total time spent on the Website.

2. Total number of visits.
3. Page Views Per Visit
4. When the lead origin was:
 - a. Lead Add Form
 - b. Lead Import
 - c. Quick Add Form
5. When their current occupation is as a working professional.