Lead Scoring Case Study Summary

This analysis is done for X Education 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:

   The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Some null values were imputed with the mode, while some irrelevant columns were dropped altogether.

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 irrelevant. The numeric values had outliers and were capped at the 95 percentile.

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 StandardScaler.

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 depending on the VIF values and p-value (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 the ROC curve) was used to find the accuracy, sensitivity, and specificity which came to be above 80% each.

7. The prediction was done on the test data f and with an optimum cut-off of 0.34 with accuracy, sensitivity, and specificity of 86.6%, 83.4%, and 88.7%.

8. Precision-Recall:

   This method was also used to recheck and a cut-off of 0.40 was found with a Precision of around 82% and recall of around 83% on the test data frame.

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

1. Tags - Lost to EINS
2. Lead Source_Welingak Website
3. Lead Source_Reference
4. What is your current occupation_Working Professional
5. Last Notable Activity_Others

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