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 lot of information about how the potential customer visit the site, the time they spent, how they reached the site.

The following are steps used:

1.Exploratory Data Analysis (EDA):

Explored the dataset, handled missing values, and addressed 'Select' levels in categorical variables.

Identified key features and their correlation with the target variable, 'Converted.'

2.Data Preprocessing:

Handled missing values and outliers.

Encoded categorical variables and dropped unnecessary columns.

3.Dummy Variables:

The dummy variables were created and later on the categorical features.

4.Train test split:

The split was done at 70% and 30% for train and test data.

5.Model Building:

RFE was done to attain top 15 relevant variables, later according to p-value variables are removed.

6.Model Evaluation:

In this we used to find the accuracy, sensitivity, specificity which came to be around 90%each.

7.prediction:

We get optimum cutoff as 0.35 with accuracy and sensitivity, specificity of 80%

8.precision and recall:

This method is used to recheck and cut off of 0.41was found and precision is 70% and recall is around is 83%.

Note:

1.Lead scoring case study has been done using logistic regression model to meet constraints as per business requirements.

2.The high number of total visits & total time spent on platform may increasing the chances of lead to be converted.

3.The leads are joined course for better career prospects, most of having specialization from finance management, lead from hr high probability to convert.

4.when the last activity was SMS, phone conversation, unsubscribed, Olark chat will help to convert leads.

5.it was found that the variables that mattered the most in the potential buyer are when their current occupation is working professional.