LEAD SCORING CASE STUDY

An education company named X sells online courses to industry professionals on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a Lead.

Although X Education gets a lot of leads, its lead conversion rate is very poor. To make the lead selection process more efficient, the company wishes to identify the most potential leads - 'Hot Leads'.

We have built a Logistic Regression prediction model to predict potential leads.

Following are the steps followed to prepare a goof-fit Logistic Regression model for Lead selection:
Step 1: Reading and understanding the data

Step 2: Preparing the data - below are some of the sub-steps

a. Converting binary variables (Yes/No) to 0/1

b. creating dummy variables for categorical values

c. handling null values etc.

Step 3: Test-Train Split - Split the records into train and test data set.

Train Data Set - used for training the model

Test Data Set - used to test the prepared model and check its accuracy and other metric values

Step 4: Feature Scaling and feature selection using RFE – for easy processing and better results

Step 5: Start prediction model on Train Data Set

a. Initially, check with cut-off=0.5 and validate using different metrics (accuracy, precision, recall etc.)

b. Plot accuracy, sensitivity and specificity curve to get better-cut-off point, in this case measured cut-off was 0.38

c. Again run the predictions and check the metric values Step 6: Predict on Test Data Set

Step 6: Predict on Test Data Set.

Below are some of the metric values calculated on Test Data Set:

Accuracy = 81%

Sensitivity = 79%

Specificity = 82%

Precision = 74%

Recall = 79%

It shows that the model is predicting potential Leads Selection efficiently