Lead Scoring Case Study

Problem Statement-

X Education sells online courses to industry professionals and they get a lot of leads, but their lead conversion rate is inferior. For example, for 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot leads' If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Summary-

Data Cleaning:

- First step to clean the dataset .

- Identify data shape , info and null values.

- Dropped the columns with null values >= 30%

- Treated the missing values by imputing the favorable Aggregate function like (Mean, Median and Mode).

- Detecting the outliers.

Exploratory Data Analysis:

- Performed EDA to check the condition of our data. The numeric value seems good but there are outliers.

- Performed Univariate Analysis for both continuous and categorical variables.

- Performed Bivariate Analysis with respect to Target variable.

Creating dummies:

- The dummy variables are created for all categorical columns.

Train-test split:

- The Train-Test Split was done at 70% and 30% for the data respectively.

Feature Scaling:

> • Used Min Max scalar to scale the data .

Model Building:

> • Using RFE we attained the top 15 relevant variables. Later the irrelevant features were removed manually depending on the VIF values and p-value (The variables with VIF > 3 and p value > .05 were dropped)

Model Evaluation:

> • A confusion matrix was made. Later on the optimum cut-off value by using ROC curve was used to find the accuracy, sensitivity and specificity which came to be around 86%.

Prediction:

> • Prediction was done on the test data frame an optimum cut-off as .44 with accuracy - 78%, sensitivity - 77% and specificity - 78%.

Precision - Recall:

> • The method was also used to recheck and a cut-off of 0.44.

Conclusion:

- Visitors who spend more on the website tend to enroll for the courses most of the time.
- Visitors who visit multiple times to the website and making queries are interested in course enrollment.
- Most of the converted leads are from Google, direct traffic and organic search. As recommended the firm can invest on digital marketing/google ads.
- Most of the persons enrolled themselves after getting SMS and through Olark chat conversation.
- Mostly the working professionals and people who are unemployed are making the queries so that they can get a job and upscale themselves. So the company can effectively pitch to working professionals and unemployed people to increase the lead conversion rate.