Lead Scoring Case Study

Summary

Problem Statement:

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Steps used are:

- **Data set**: in the data set 9240 rows and 37 columns, data have missing values which can be handled by dropping the columns which have more than 40% missing data.

- **Outlier check**: did univariate analysis and then outlier treatment these were some potential outliers which I did capping.

- **Data Visualizing**: did bivariate analysis

  - Lead origins categories "Lead Add Form" category have highest conversion ratio

  - "SMS Sent" category in "Last Activity" column has highest conversion ratio

  - Following that "Email Opened" has the second highest

  - In "Lead Source" category "Reference" is performing better in conversion following with "Google"

  - Should target "Working Professional" more than "Unemployed".

➢ **<u>Scaling/Dummy variable:</u>** categorical variables are converted into dummy variable and scaling is done on both train and test data set .

➢ **<u>Train-Test Split</u>**: it was performed in the ratio of 70:30 and Logistic regression initiated.

✓ Edtech company should focus on :

1. Lead 'Source_Wllingak' Website.
2. Lead Origin Lead Add form.
3. Tags will revert after reading email.