

Summary Report

Problem Statement:

- Now, the overall lead conversion rate is 38%, which is very poor percentage for the company profit. So, the CEO has given a ballpark of the target lead conversion rate to be around 80%.
- To make this process more efficient, the company needs to identify the most potential leads, also known as 'Hot Leads'.
- If the company 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.
- This can be achieved by assigning 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.

Analysis:

- we need to find out the most relevant variables, which will be used for predicting the high converted score. So, We can select these relevant features by using "Converted" variable.
- Here, 'Converted' variable is the target variable, which indicates whether a lead has successfully converted or not.

- So, we need to do analysis of all the variables one by one.
- I followed below steps to analyze the variables:
 1. Importing Data
 2. Understanding Structure of Data
 3. Data Preparation
 - 3.1. Missing values treatment.
 - 3.2. Checking for outliers and treatment.
 4. Exploratory Data Analysis
 5. Feature Engineering
 - 5.1. Dummy variable Creation
 - 5.2. Dropping Repeated variables.
 - 5.3. Feature Standardization
 6. Splitting the data into Train and Test
 7. Model Building :
 - 7.1 Using Logistic regression model.
 - 7.2 Feature selection using RFE (Automated)
 - 7.3 Feature selection using p-Value and VIF
 - 7.4 Iterating model building till getting the perfect model
 8. Model Evaluation:
 - 8.1 Using Confusion Matrix
 - 8.2 Calculating overall accuracy score.
 - 8.3 Plotting ROC curve
 - 8.4 Calculating optimal cutoff point
 - 8.5 Using Precision and Recall
 9. Getting the predicted values (probability value for each lead)

Result:

- The lead scores are assigned to each lead and it will help X Education to identify the potential leads, called as 'Hot leads'. So that, the company can 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.
- The potential leads can be decided on the basis of lead scores now.
- If the lead score is higher i.e. greater than 80, then the lead is hot and is most likely to convert.
- If the lead score is low i.e. let's say less than 50, then lead is low and will mostly not get converted.

Recommendation to company:

- ***X Education can focus on below criteria's of the leads to identify the more potential leads.***
 - Lead Origin
 - Lead Source
 - Do Not Email
 - Do Not Call
 - Total Visits
 - Total Time Spent on Website
 - Page Views Per Visit

- Last Activity
- Specialization
- What is your current occupation