Lead Scoring Case Study Report:

Summary:

Problem Statement:

The company requires to build a model wherein a lead score is assigned 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.

This case study focuses on building a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

Identification of such leads which can possible be converted is the focus of the case study

Below Steps are performed in the case study for the outcome :

Data Loading & Cleaning
Data Quality & Missing Values Check
Handling Outliers
Exploratory Data Analysis
Data Preparation for Modelling
Train-Test Data Split
Scaling
Feature Selection
Recursive Model Building to find the optimal model
Model Evaluation using Performance Metrics & Building ROC Curve Finding Optimal Cut-Off point
Predictions on Test data using Final Model
Final Evaluation using Performance Metrics on Test Data
Calculating Lead Score

Conclusion:

While we have checked both sensitivity and specificity as we as precision and recall metric techniques. We have concluded sensitivity and specificity technique is more efficient for calculating.

The accuracy, sensitivity, specificity values of test set are around 80% that means conversion rate for final predicted model is around 80%

Below three factors contribute towards the probability of lead getting converted

- Total time spent on Website

- Lead Source

- Current occupation

X Education company needs to focus on the following key aspects:

- Total time spent on website

- Lead source: With lead source as direct traffic, Google, Olark Chat and Organic Search are not potential for taking the education program

- Last notable activity

- What is your current occupation

- Specialization

From final model we find the lead score for each lead. Observed that increase or decrease in probability threshold value will in turn decrease or increase the sensitivity and increase or decrease the specificity of the model.

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

Logistic Regression Model:

The model shows high close to 81% accuracy

The threshold has been selected from Accuracy, Sensitivity, specificity measures and precision, recall curves.

The model shows 76% sensitivity and 83% specificity

The model finds correct promising leads and leads that have less chances of getting converted

Overall this model proves to be accurate