Lead scoring case study

Approach



Objective:

Recommend X education on how to decide on which leads to target

Understanding Variables:

There are 35+ variables in this data set and data cleaning needs to be performed before proceeding

<u>Choosing</u> <u>variables:</u> This is

variables: This is done based on RFE

Model building:

Model was built using logistic regression

Model assessment:

Based on precision recall curve the probability cut off was decided

Recommendation:

Using probability cutoff the recommendation to go after a lead or not was provided.

Data cleaning and assessment

Strategy:

- Any column with more than 30% null values was removed
- The text 'Select' was imputed first with null values

Model building and assessment

Strategy:

- Logistic regression was used (since it's a case of binary prediction)
- Precision and accuracy curve was checked while deciding on cut off probability
- Confusion matrix was used to assess model

Recommendation

Strategy:

- The lead conversion accuracy has gone up to 80% based on the model and hence it is of a high business value
- The recommendation is to use this model for assessing whether to go after a a lead or not