Executive summary

Mentioned model is focus towards Selecting the most promising leads,

i.e. the leads that are most likely to convert into paying customers.

We built a model wherein assigned 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 target lead conversion rate to be around 80%.

Following steps helps us to achieve the same

- Dropped all the columns which is having high missing value – 35%
- Most of the data having named select converted to null/nan values
- Missing value imputations of numerical and categorical values
- All categorical variables dummies value creating
- Numerical values standardisation/scaling using MinMaxScaling
- Features selections using RFE module
- Manual features selections using Stats module checking P value and VIF score
- Evaluation of model using Confusion matrix
- Checking Accuracy , Recall and Precession scores
- Checking ROC curve and finding optimal cutoff
- Making predictions of test data and comparing results with train data