

Summary report of lead score case study

The whole process is done in step by step

Step 1 – importing data and libraries

The data has been imported to jupyter notebook in a data frame

Step 2- reading and understanding data

We have checked the shape, size columns of the data

Step 3- data preparing

Some data has value 1 yes or no so I changed it into binary after that I have made dummy variable to categorical column and use drop first but in some categorical column there is a option 'select', which is as equivalent to null so I drop it in process of dummy making

After that I have checked outlier and two numerical column has outlier value so I removed outlier, after that I have seen that some column has all the values are zero so I dropped it as it will not show any variance, after that I have checked null value and in two column there is 46% null value so it's better to drop it so I dropped it

Step 4 – train test split

Now I have done train test split on data and in train set I have kept 70 % and in test set 30% data

Step 5- scaling the feature

Now I have used standard scaling of numerical feature

We have checked that there is 37% conversion rate so it is not a case of class imbalanced

Step 5 – checking correlation

Ahead checked correlated variable and found some so dropped from both of data set

Step 6 – model building

Now I have built GLM model of data with all variable after that uses rfe for feature selection and chooses 20 variable from rfe column, after that I have dropped two variable which is insignificant to model.

Step 6 – prediction

Now created a column named lead score by multiplying converted probability and predicted the outcome from the train set model I have got 92% accuracy from the model. and also a good ROC curve and precision of 86% and recall of 93% also checked vif which has no high value

Now checking outcome from test set data we have got accuracy of 92% and also 91% specificity and 93% sensitivity from the model.

Final result

We have got 18 feature from final model which will help in making a better business decision and should focus on that variable for getting target.

Summary report of lead score case study