

Executive Summary

Important outcomes from the analysis:

Test-Train Split

- The split was done at 0.7 and 0.3 of train and test data
- Min-Max scaling was found on the variables of ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']

Model Building

- REF was used for feature determination
- REF is done for assessing the variables

Creating Prediction

- P-value and VIF were used for predicting final set of data

Model Evaluation

- Confusion matrix was created
- Overall accuracy= 0.738857501569366
- Sensitivity= 0.5269121813031161
- Specificity= 0.8731094591130479

With the current cut off as 0.5 we have around 73% accuracy, sensitivity of around 52% and specificity of around 87%

Optimise Cut off (ROC Curve)

- ROC function was used

The area under ROC curve is 0.78 which is a decent value

- Columns created with different probability cutoffs
- The dataframe for cutoffs are ['prob', 'accuracy', 'sensi', 'speci']
- The cutoff was plotted based on ['accuracy', 'sensi', 'speci']

From the graph it is visible that the optimal cut off is at 0.35

- Overall accuracy= 0.7004080351537979
- Confusion matrix was created

- Sensitivity= 0.679481991096722
- Specificity= 0.7136631632914637

Prediction on Test set

- Scalling of numeric values was done on ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']
- Boolean colums were created for numeric
- Prediction of test set in variables was sorted with 'y_test_pred'
- Overall accuracy= 0.7056023434639326
- Calculating the sensitivity= 0.679481991096722
- Calculating the specificity= 0.7136631632914637

With the current cut off as 0.35 we have accuracy, sensitivity and specificity of around 70%.

Precision-Recall

- Precision= 0.7245409015025042
- Recall= 0.5269121813031161

With the current cut off as 0.35 we have Precision around 72% and Recall around 52%

Precision and recall tradeoff

- Accuracy= 0.724105461393597
- Precision= 0.6658911121451838
- Recall= 0.5791177660866046

Prediction on Test set

- Precision= 0.6674418604651163
- Recall= 0.5797979797979798

Conclusion:

Top variable for conversion

- *Lead Source*
 - Total Visits
 - Total Time Spent on Website
- *Lead Origin*
 - Lead Add Form
- *Lead source*
 - Direct traffic
 - Google
 - Welingak website
 - Organic search
 - Referral Sites

