

SUMMARY REPORT

The model building and prediction is being done for company X Education and to find ways to convert potential users. We will further understand and validate the data to reach a conclusion to target the correct group and increase conversion rate. Let us discuss steps followed:

1. EDA:

- Quick check was done on % of null value and we dropped columns with more than 35% missing values and dropped unique columns
- Which columns less than 30% null values, we'll examine each column. We can then add values, or delete columns if you find it inappropriate.

2. Train-Test split & Scaling:

- The split was done at 70% and 30% for train and test data respectively.
- We will do min-max scaling on the variables ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']

3. Model Building

- RFE was used for feature selection.
- Later the rest of the variables were removed manually depending on the VIF values and p-value.
- We built 24 models to find the optimal value.
- A confusion matrix was created, and overall accuracy was checked which came out to be 80.72%.

4. Model Evaluation

On Training Data, we get:

The optimum cut off value was found using ROC curve. The area under ROC curve was 0.89.

After Plotting we found that optimum cutoff was 0.36 which gave

- Accuracy 80.33%
- Sensitivity 81.55%
- Specificity 79.56%.

On Test Data, we get:

- Accuracy 80.86%
- Sensitivity 82.32%
- Specificity 80.00%

5. Conclusion

First of all, in the process of searching for the best model, we found that there are too many variables affecting the selection process, leading to too many models to run. In general, running many test models in

this case we find is not optimal. However, within the requirements we also found variables that influence potential buyers:

- The total time spend on the Website.
- Total number of visits.

Lead source (In descending order):

- Google
- Direct traffic
- Organic search
- Welingak website

Priority Active:

- SMS
- Chat conversation

Priority should be given to the following industries to have a high chance of achieving leads:

- Working professional