

Summary

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 45% missing values.
- We also saw that the rows with the null value would cost us a lot of data and they were important columns. So, instead we replaced the NaN values with 'not provided'.
- Since India was the most common occurrence among the non-missing values, we imputed all not provided values with India.
- Then we saw the Number of Values for India were quite high (nearly 97% of the Data), so this column was dropped.
- We also worked on numerical variable, outliers and dummy variables.

2. Train-Test split & Scaling :

- The split was done at 70% and 30% for train and test data respectively.
- We did 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.
- Then RFE was done to attain the top 15 relevant variables.
- Later the rest of the variables were removed manually depending on the VIF values and p-value.
- A confusion matrix was created, and overall accuracy was checked which came out to be 92.16%.

4. Model Evaluation

▪ On Training Data

- The optimum cut off value was found using ROC curve. The area under ROC curve was 0.97.
- After Plotting we found that optimum cutoff was **0.3** which gave
 - Accuracy: 92.16%
 - Sensitivity: 91.99%
 - Specificity: 92.27%
 - Precision: 88.79%
 - Recall: 91.69%

- Prediction on **Test Data**

- Accuracy: 92.59%
- Sensitivity: 91.38%
- Specificity: 93.31%
- Precision: 89.75%
- Recall: 91.08%

CONCLUSION

TOP VARIABLE CONTRIBUTING TO CONVERSION:

- Total Time Spent on Website
- Lead Origin_Lead Add Form
- Last Activity_SMS Sent
- What is your current occupation_Working Professional
- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Tags_Will revert after reading the email
- Lead Source_Welingak Website

The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.