Machine Learning case study –approach 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.

- 1) Imported required libraries and the data .

EXPLORATORY DATA ANALYSIS-

- EDA:

- Quick check was done on % of null value and dropped columns with more than 45% missing values.

- 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, removed all not provided values with India.

- Then the Number of Values for India were quite high (nearly 97% of the Data), so this column was dropped.

- Also worked on numerical variable, outliers and dummy variables.

Train-Test split & Scaling :

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

Model Building

- RFE was used for feature selection.
- Then RFE was done to attain the 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 80.91%.

Model Evaluation

- Let us compare the values obtained for Train & Test:
- Train Data:
- Accuracy : 91.82%
- Sensitivity : 91.29%
- Specificity : 92.15%
- Test Data:
- Accuracy : 92.59%
- Sensitivity : 91.76%
- Specificity : 93.00%
- The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model