

SUMMARY

The model building and prediction is being done for X Education company to find ways to convert potential users.

1. EDA:

- Quick check was done on % of null value and we dropped columns with more than 40% missing values.
- Most of the rows were having null values which we replaced with 'Unknown'.
- The Number of Values for India were quite high (nearly 97% of the Data), so this column was dropped.
- We also worked on numerical variables, 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 standard scaling on the variables ['Total Visits', 'Page Views Per Visit', 'Total Time Spent on Website']

3. Model Building

- RFE was used to attain the top 15 relevant variables.
- Later the rest of the variables were removed manually depending on the VIF values and p-value.
- The overall accuracy was above 80%.

4. Model Evaluation

- 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 **0.35**
- **Precision – Recall:**
- Training data with the cutoff of 0.35 we get the Precision & Recall of 79.67% & 70.27% respectively.

CONCLUSION

TOP VARIABLE CONTRIBUTING TO CONVERSION:

- Lead Origin:
 - Lead Add Form
- LEAD SOURCE:
 - Total Visits
 - Total Time Spent on Website
- What is your current occupation
 - Working Professional

Last Activity:

- SMS sent
- Olark chat conversation