

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:

- Renamed the columns to avoid ambiguity.
- Observed SELECT as null values and managed it properly by studying variable imbalance and impute proportionally.
- Then we saw the Number of Values for India were quite high (nearly 97% of the Data), so this column was dropped, Similar columns with high missing values are also 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 will do StandardScaler 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.
- Finally we arrived at the top 9 features

4. Model Evaluation

CONCLUSION

TOP THREE VARIABLE CONTRIBUTING TOWARDS THE PROBABILITY OF A LEAD GETTING CONVERTED:

- Total Time Spent on Website (Coefficient: 4.6833)
- Lead Origin_Lead Add Form (Coefficient: 3.0865)
- What is your current occupation_Working Professional (Coefficient: 2.7694)

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.