

Summary of Approach and learnings

Following Step are performed to Prepare Data for Model Building

Checked Statistical aspects of the dataframe

Missed Values are checked and Appropriate columns are removed

Data having "Select" values are handled

Dropped Columns that are not necessary for analysis

Continuous and Categorical Columns are detected

Dummy Variables are created for categories in categorical columns

Outlier Analysis is performed and values above 0.99 percentile are removed

Model Building and Evaluation

Data is split into Train and Test Data

Recursive Feature Elimination is done to select 15 relevant features

Logistic Regression Models are build with the relevant feature variables

Models are assessed and 6th Model is selected as best fit Model.

Optimum Probability is found to be 0.35

Predicted converted values are found out with the Optimum Probability value

ROC is applied and found the model to be 88% accurate

Confusion Matrix and other Model features like accuracy,specificity,sensitivity,false positive rate

And negative predictive values are calculated.

Model is applied and Test data set and found to be matching with train Dataset.

Conclusion:

Model is built with 88% accuracy

Hot Leads dataset is created with conversion rates greater than 80%

Below Features are found to be most important:

"Lead Origin_Lead Add Form"

"Lead Source_Welingak Website"

"What is your current occupation_Working Professional"