

SUMMARY REPORT FOR LEAD SCORING CASE STUDY

As per the given problem statement, Analysis is done for X Education and find ways to get more professional to join online courses. As an input, we were given a dataset that how potential customer visit the site, their source and so on.

Below are the steps that we have taken:

1. Data Cleaning: We inspected Data and full null values. So, imputed the null value with different value not provided as we don't want to alter the data. Have dropped the columns where null value is more than 30%. We have also converted binary variable to 1 or 0 and so on other checks.
2. EDA: Have performed a quick check to visualize condition of Data through different plots. Dropped unnecessary columns.
3. Dummy Variable: have created dummy variables and after creation of that, dropped irrelevant columns.
4. Splitting Data: Now, Data is split in train dataset and test dataset.
5. Feature Scaling: Performed minmax scaling.
6. Checking Correlations: Checked correlations and dropped high correlated columns.
7. Model Building: Have used RFE for feature selection, selected top 15 relevant features. After that we have dropped some features on basis of p-value and vif.
8. Model Evaluation: A confusion Matrix was made. Later on, optimum cut off value used to check accuracy, sensitivity and specificity.
9. Prediction: Prediction was done on test data frame and with optimum cut off as 0.42.

Below are the features which have high potential:

- Total Time Spent on Website
- Total Visits
- Page views per visit
- Lead source (welingak website)
- Last Notable Activity_other_notable_activity
- Last Activity (Olark chat conversation, sms_sent)