

SUMMARY REPORT

Problem Description

X Education, an online course provider, seeks to enhance lead conversion efficiency. After acquiring leads through website interactions, the sales team engages in calls and emails, with a typical conversion rate of 30%. To improve efficiency, the company aims to identify potential leads, or "Hot Leads," to focus sales efforts more effectively. Strategies include implementing lead scoring, analysing historical data for ideal customer profiles, using automation tools, personalizing communication so that we can boost overall lead conversion rates to be around 80%.

Approach-

We used Logistic regression modelling and followed below steps-

1. Data Reading and Understanding-
 - Importing necessary Libraries and warnings
 - Importing Data and Checking Data Types
2. Data Cleaning-
 - Checking Null values, dropping the columns with null values greater than 30%
 - Dropping columns which were not needed for the analysis
3. Data Vizualization-
 - We plotted graph for those columns which have "select" values and dropped them later on as they were equivalent to null values
 - Dropped those columns which have "No" values mostly present in the data set
 - Performed Analysis on Numerical columns by plotting pair plot and Heatmap.
 - Performed Analysis on the categorical columns by plotting count plot
4. Feature Scaling
 - We converted all the categorical columns to numerical by converting them to dummy variables.
 - We scaled all the Numerical variables using min- max scaler function
5. Model Building
 - We used RFE to select 15 columns to build the model.
 - We performed logistic regression and optimised variables based on p-values (<0.05) and VIF values
 - Repeated above step in order to optimise our model
6. Model evaluation on Train Set
 - We created confusion matrix and calculated the metrics- accuracy, sensitivity, specificity, precision and recall
 - Plotted graph for all the matrix and found optimum cutoff as 0.28
 - Plotted ROC curve to find the area under the curve

7. Model Predictions on Test Set

- After calculating optimum cut off by plotting graph which we got as 38%, then we calculated the metrics and predicted data on the test set.

8. Observation and Conclusions:

- When the Company has limited time and resources, it should approach Hot_leads i.e. those leads who have more than 80% of conversion chances to achieve maximum conversion & to avoid useless phone calls.
- Train Data:
 - Accuracy : 91.23%
 - Sensitivity : 93.6%
 - Specificity : 91.3%
 - Precision : 91%
 - Recall : 94%
- Test Data:
 - Accuracy : 94%
 - Sensitivity : 93.8%
 - Specificity : 94.7%
 - Precision : 94%
 - Recall : 94%
- Additional Observations-
 - Lead Source- Company can focus on the lead source from "Google", "Direct Traffic" and "Reference"
 - Lead Origin- Company can focus on the customers who have landed on the "Landing Page Submission"
 - What is your current occupation- Company can focus on customers who are unemployed.
 - Specialization- "Finance Management", "Marketing Management" and "Human Resource Management".
 - Last Activity- Company can focus on the customers who have – "SMS Sent" and "Email Opened"