

## SUMMARY

This analysis was conducted for X Education to explore strategies for attracting more industry professionals to their courses. The initial data provided valuable insights into potential customers' website visits, the duration of their sessions, their entry sources, and the conversion rate.

The following steps were taken:

### 1. Data Cleaning

The data was mostly clean, with a few null values. We replaced a less informative option with a null value. Some nulls were labeled as 'not provided' to retain data integrity, although they were later excluded during dummy variable creation. To categorize users effectively, entries were classified as 'India,' 'Outside India,' or 'not provided.'

### 2. Exploratory Data Analysis (EDA)

A preliminary EDA revealed many irrelevant elements within the categorical variables. The numeric data appeared robust, with no outliers detected.

### 3. Creation of Dummy Variables

Dummy variables were generated, and those marked as 'not provided' were subsequently removed. For numeric values, we applied the MinMaxScaler.

### 4. Train-Test Split:

The dataset was divided into training (70%) and testing (30%) subsets.

### 5. Model Building:

Recursive Feature Elimination (RFE) was employed to identify the top 15 relevant variables. Additional variables were removed manually based on Variance Inflation Factor (VIF) and p-values, retaining those with  $VIF < 5$  and  $p\text{-value} < 0.05$ .

### 6. Model Evaluation

A confusion matrix was created, and the optimal cutoff value was determined using the ROC curve. The resulting accuracy, sensitivity, and specificity were all approximately 80%.

### 7. Prediction

Predictions were made using the test dataset, with an optimal cutoff set at 0.35, achieving accuracy, sensitivity, and specificity around 80%.

### 8. Precision-Recall Analysis

This method was applied for verification, yielding a cutoff of 0.42, with precision at 73% and recall at 75% on the test data.

**The most significant factors influencing potential buyers, in descending order, were identified as:**

- Lead Origin\_Lead Add Form
- What is your current occupation\_Working Professional
- Lead Source\_Welingak
- Last Activity\_SMS Sent
- Total Time Spent on Website
- Last Activity\_Email
- TotalVisits

By focusing on these key factors, X Education can significantly enhance its chances of converting potential buyers into course participants.