

# Summary

## Lead Scoring case Study

The analysis conducted for X Education aimed to identify strategies for attracting more industry professionals to enroll in their courses.

The following steps were undertaken:

### 1. Data Cleaning:

- Null values were replaced with "not provided" to retain data integrity.
- Categorical variables were refined, with "India", "Outside India", and "not provided" being the main categories for location.

### 2. Exploratory Data Analysis (EDA):

- A quick EDA was performed to assess data quality.
- Categorical variables were scrutinized for relevance, while numeric values were checked for outliers (none found).

### 3. Dummy Variables:

- Dummy variables were created for categorical features, and "not provided" categories were subsequently removed.
- Numeric features were scaled using MinMaxScaler.

### 4. Train-Test Split:

- The data was split into 70% for training and 30% for testing.

### 5. Model Building:

- Recursive Feature Elimination (RFE) was employed to identify the top 15 relevant variables.
- Variables were further pruned based on VIF values and p-values (keeping those with  $VIF < 5$  and  $p\text{-value} < 0.05$ ).

### 6. Model Evaluation:

- A confusion matrix was generated, and the optimum cut-off value (using ROC curve) was determined.
- Accuracy, sensitivity, and specificity were calculated, yielding approximately 80% for each.

### 7. Prediction:

- Predictions were made on the test data frame using an optimal cut-off of 0.35, resulting in 80% accuracy, sensitivity, and specificity.

### 8. Precision-Recall:

- Precision-recall analysis was conducted, determining a cut-off of 0.41 with precision around 73% and recall around 75% on the test data frame.

The key findings are as follows:

## **Key Findings:**

The most influential variables for potential buyers were identified, ranked as follows:

1. Total time spent on the website.
2. Total number of visits.
3. Lead sources, including Google, Direct traffic, Organic search, and Welingak website.
4. Last activity, particularly SMS and Olark chat conversation.
5. Lead origin as Lead add format.
6. Current occupation as a working professional.

### **Implications:**

- X Education stands to benefit significantly by targeting potential buyers based on these key variables, thereby increasing the likelihood of course enrolment among industry professionals.