SUMMARY
Lead Scoring Case Study
by
Divya Sonali Minz

The analysis conducted for X Education aimed at devising effective strategies to attract more industry professionals to enrol in their courses. The initial dataset provided valuable insights into customer behaviour, visit patterns, and conversion rates. The following stepwise approach was employed in the analysis:

1. **Data Cleaning**:
   - Addressed null values, replacing 'select' with a null value.
   - Transformed some null values to 'not provided' to retain data.
   - Categorized locations into 'India,' 'Outside India,' and 'not provided.'

2. **Exploratory Data Analysis (EDA):**
   - Conducted a quick EDA to assess data conditions.
   - Identified and removed irrelevant elements in categorical variables.
   - Detected no outliers in numeric values.

3. **Dummy Variables:**
   - Created dummy variables and removed those with 'not provided' elements.
   - Applied MinMaxScaler for scaling numeric values.

4. **Train-Test Split:**
   - Split the dataset into training and test sets (70% and 30%, respectively).

5. **Model Building:**
   - Utilized Recursive Feature Elimination (RFE) to identify the top 15 relevant variables.
   - Manually removed other variables based on VIF values and p-values.

6. **Model Evaluation:**
   - Constructed a confusion matrix for evaluation.
   - Determined the optimum cutoff value using the ROC curve.
   - Computed accuracy, sensitivity, and specificity (each around 80%).
7. **Prediction:**
   - Executed predictions on the test dataset using an optimum cutoff of 0.35.
   - Achieved 80% accuracy, sensitivity, and specificity.
8. **Precision-Recall Analysis:**
   - Conducted precision-recall analysis with a cutoff of 0.41.
   - Validated the model's performance with precision around 73.24% and recall around 76.61%.
9. **Key Variables Influencing Potential Buyers:**
   - Total time spent on the website.
   - Total number of visits.
   - Lead source categories: Olark chat, Wellingak Website.
   - Last activity categories: SMS, Olark Chat Conversation.
   - Lead origin: Lead add Form.
   - Current occupation categories: Working Professionals, Student, Unemployed, Other.

In conclusion, the insights derived from this comprehensive analysis provide actionable information for X Education to tailor effective engagement strategies, ultimately leading to an increase in course enrollments.