

LEAD SCORING – SUMMARY

This research is being conducted for X Education in order to discover new strategies to attract more industry experts to their courses. The basic data provided us with a wealth of insight about how the system worked. Visits to the site by potential customers, the amount of time they spend there, how they got there, and the products they buy rate of conversion

The following are the steps that were taken:

1. Data cleaning: Except for a few null values, the data was mostly clean, and the option choose had to be replaced with a null value because it didn't provide much information. To avoid losing too much data, a few of the null values were changed to 'not provided.' They were later removed while manufacturing dummies, though. The elements were altered to 'India,' 'Outside India,' and 'not provided' because there were many Indians and few outsiders.

2. EDA: A short EDA was performed to assess the state of our data. Many factors in the categorical variables were found to be irrelevant. The numerical figures appear to be accurate, and no outliers have been discovered.

3. Dummy Variables: Dummy variables were created, and those with 'not given' items were later eliminated. We utilised the MinMaxScaler for numeric values.

4. Train-Test Split: The train and test data were split 70 percent and 30 percent, respectively.

5. Model Development: To begin, RFE was used to identify the top 15 relevant variables. The remaining variables were then manually deleted based on their VIF values and p-values (the variables with VIF 5 and p-value 0.05 were maintained).

6. Model Evaluation: A confusion matrix was created to evaluate the model. Later, the ideal cut off value (as determined by the ROC curve) was utilised to determine the accuracy, sensitivity, and specificity, all of which were found to be around 80%.

7. Prediction: Prediction was performed on the test data frame using an ideal cutoff of 0.35 and an accuracy, sensitivity, and specificity of 80%.

The variables that mattered the most to potential purchasers were discovered to be (in descending order):

1. Total time spent on the Website.
2. The total number of visits
3. If the lead came from:
 - a. Google,
 - b. direct traffic,
 - c. organic search, or
 - d. the Welingak website
4. When was the most recent activity:
 - a. Olark chat conversation
 - b. SMS
5. When the source of the lead is Lead add format.
6. If they are currently employed as a working professional.

With these factors in mind, X Education may thrive since they have a very good probability of convincing practically all potential buyers to alter their minds and purchase their courses