

# **Lead Scoring Case Study Summary Report**

## **By Yash Sahindrakar and Sagar Patil**

This analysis is carried out for X Education in an effort to attract more business professionals to their courses. We learned a lot from the fundamental statistics on how potential customers use the site, how long they stay there, how they got there, and the conversion rate.

The following steps used for overall model building :

### **1. Data Cleaning**

Except for a few null values, the data was mostly clean. However, the option choose had to be changed to a null value because it provided little useful information.

To avoid losing too much data, a few of the null values were converted to "unknown."

Nevertheless, they were eventually taken out while manufacturing dummies.

The components were altered to "India," "Outside India," and "not given" because there were a lot of people from India and a small number from elsewhere.

### **2. EDA**

To quickly assess the state of our data, an EDA was performed.

It was discovered that several of the category variables' components were unnecessary.

The numerical figures are accurate, and no anomalies were discovered.

### **3. Dummy Variables**

After the dummy variables were made, those that had the phrase "not given" were later deleted. We utilized the MinMaxScaler to scale numerical numbers.

### **4. Train Test Split**

A split of 70:30 is used for the model building process in which 70 goes for training and 30 goes for testing.

### **5. Model Building**

RFE was first done to select the 15 relevant variables for analysis afterward some of those variables were removed by manual elimination determined by the factor of p-value and VIF. If  $p\text{-value} > 0.05$  and  $VIF < 5$  then that variable was eliminated from the process.

## **6. Model Evaluation**

Model evaluation was done on the basis of a confusion matrix and all the parameters were checked such as accuracy, precision, recall, sensitivity and specificity on both train and test sets. And the results came out to be satisfactory.

## **7. Precision and Recall**

For this case study accuracy was pretty much misleading and our main focus was on recall value. Recall value came as 69% and 75% on train and test set respectively.

## **8. Conclusion**

The factors that affected potential purchasers the most were discovered to be: Total time spent on the Website. The total quantity of visitors. When Google, direct traffic, organic search, or the Welingak website was the top source when the last activity occurred, it was either an SMS or an Olark chat. If the lead origin is Lead Add format with these in mind, X Education may succeed since they have a very high possibility of persuading nearly all prospective customers to alter their minds and purchase their courses.