Executive Summary of Lead Score Case Study

X Education is an online education company that offers courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once people land on the website, they may browse the courses, fill out a form for a course, or watch some videos. When people provide their email address or phone number, they are classified as a lead. The leads are then contacted by the sales team to convert them into customers. The typical lead conversion rate at X Education is around 30%

The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The following are the steps used:

1. Cleaning data:

The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed to 'India', 'Outside India' and 'not provided'.

2. EDA:

A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values are good and no outliers were found.

3. Dummy Variables:

The dummy variables were created and later on the dummies with ‘not provided’ elements were removed. For numeric values we used the MinMaxScaler.

4. Train-Test split:

The split was done at 70% and 30% for train and test data respectively.

5. Model Building:

Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with VIF < 5 and p-value < 0.05 were kept).

6. Model Evaluation:

A confusion matrix was made. Later on the optimum cut off value (using ROC curve)was used to find the accuracy, sensitivity and specificity which came to be around 80% each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.

8. Precision – Recall:

This method was also used to recheck and a cut off of 0.41 was found with Precision around 73% and recall around 75% on the test data frame.

Based on these observations, several recommendations were made, like:

• Features such as 'Lead Origin_Lead Add Form', 'Current_Occupation_Working Professional', and

'Total Time Spent on Website' have a high conversion rate and should be utilized more in lead

generation efforts.

• Working professionals should be aggressively targeted as they have a higher probability of

converting and are likely to have better financial situations to pay for services.

• Referral leads generated by old customers have a significantly higher conversion rate and should be

incentivized with discounts or other rewards to encourage more referrals.

• Increasing the frequency of media usage such as Google ads or email campaigns can save time and

increase the conversion rate.

• Leads whose 'Last Activity' is ‘SMS Sent’ or ‘Email Opened’ tend to have a higher conversion rate

and should be targeted more frequently.

• Analysing the behaviour of customers who spend more time on the website can help improve the

user experience and increase conversion rates, and company should focus on creating engaging

content and user-friendly navigation to encourage customers to spend more time on the website.

• Understanding the most popular specializations can help tailor course offerings and marketing

campaigns to specific groups of customers. Providing targeted content and resources for popular

specializations such as Marketing Management and HR Management can also help attract and retain

customers in those fields.