

Lead Score Case Study

Summary Report

The Lead Score Case Study is the analysis done for a X Education Company for finding ways to increase the number of Industry Professionals joining their courses. The data provides an idea about the potential customers, how they go to know and visited the site, how much time they have spent on the website, conversion rate etc.

The steps used in analysis are as follows:

1. *Understanding and Reading the Data:* Read the data dictionary and data provided
2. *Cleaning the Data and EDA:* Dropped variables with high null values, treated the missing values both in categorical and numerical columns. EDA was done to understand the data orientation.
3. *Creating Dummy Variables:* Dummy variables were created for categorical variables.
4. *Train-Test Splitting of Data:* 70 % and 30% split was done
5. *Building the Model:* RFE and stats model was built and depending on the p-value and VIF values the variables were dropped.
6. *Evaluating Model:* ROC Curve was used to check the model fitting. Confusion matrix were used to find the accuracy, sensitivity, specificity of the logistic regression model.
7. *Prediction:* Prediction on test set was done with a cut-off probability of 0.3 which was found from the sensitivity-specificity curve and Precision-Recall curve.
8. *Sensitivity and Specificity:* The values of Sensitivity and Specificity obtained were 77% and 82% approximately on the Test Data.

Observations:

It was observed that the following variables were important in finding the potential customers:

- Total Time Spent on Website
- Lead Source: Welingkak chat and Olark Chat
- Lead Origin: Lead add form

Based on the probability found from the Logistic Regression Model, we have calculated the Lead Score and assigned to the Leads.

Considering the optimal Threshold obtained from the Precision-Recall curve, we decided that the Lead score with a probability value greater than or equal to 0.3 can be considered as a "Potential Leads".

This decision has led us to a **conversion rate** of around **77%** which is calculated based on the sensitivity of the model.

We have also found the model **accuracy** score to be **80%**.

Further, we have also found out that leads who are students or are unemployed tend to decrease the Conversion rate. The reason could be since these courses are designed for industry Professionals. They are not meant to be for Students. So, the company should work more on their website so that when such Leads go through the websites, they must be well aware of the courses or must have some industrial experience.

The company should also prioritise more on the working professionals instead of students. With these points the X Education Company have high chances of getting all the potential customers.