

Case Study Team 7 (Roll no: 11, 43, 55, 60, 86)

Problem Statement:

We have done Analysis for X-Education to understand and predict ways to get more industry professionals to join their courses. The data that was given to us provided a lot of information about the potential customers, how they visit the website, the time they spend there and the conversion rate.

The following are the steps used:

1. Cleaning data:

- The data was partially clean except for a few null values.
- The value 'Select' in some records suggests that the lead did not enter any data. So we can conclude that the 'Select' value is as good as null value. Hence, we replaced 'Select' with a null value.
- Few of the null values were changed to 'unknown' and 'No Tags' so as to not lose much data.
- Since there were many categories which were not having much records so we have clubbed them in a new column.

2. EDA:

- A quick EDA was done to check the condition of our data. It was found that some
- elements in the categorical variables were irrelevant. The numeric values seemed good
- and there were outliers present so we handled them by using capping technique.
- We have also analysed the converted lead percentage of each feature.

3. Dummy Variables:

- The dummy variables were created for categorical features.

4. Train-Test split:

- The Data was split into Training set and Testing set.
- The split was done at 80% and 20% for Train and Test data respectively.

5. Model Building:

- We have designed two models. One is Logistic Regression and the other is Decision Tree and among them we have selected the Logistic Regression model.
- Our final Logistic Regression Model is built with 13 features. Here the logistic regression model is used to predict the probability of conversion of a customer.
- The final model has accuracy score of 0.8991 i.e. 89.91% and sensitivity is 83.12%
- The top three categorical variables/dummy variables in the final model are Tags_Lateral student, Tags_Interested in Next batch, current_occupation_Housewife.

6. Model Evaluation:

- We evaluated our model using 3 methods: Confusion Matrix, K-Fold Cross Validation with 10 samples and Cross Validation with 15 samples.

7. Prediction:

- Prediction was done on the test data frame and we get an accuracy of 89.86%, sensitivity of 83.12% and specificity of 93.42%.

Conclusion:

1. API, Landing Page, Lead Import has less conversion percentage (less than 40%) but counts of the leads from them are considerable. The count of leads from the Lead Add Form is pretty low but the conversion rate is very high. Lead Import has very less count as well as conversion rate and hence can be ignored.

To improve the overall lead conversion rate, we need to focus on increasing the conversion rate of 'API' and 'Landing Page Submission' and also increasing the number of leads from 'Lead Add Form'.

2. The count of last activity and notable activity as "Email Opened" is max and conversion percentage of SMS sent as last activity is maximum.

We should focus on increasing the conversion rate of those having last activity as Email Opened by making a call to those leads and also try to increase the count of the ones having last activity as SMS sent.

3. We can say that Working professionals have high conversion percentage. The count of Number of Unemployed leads is more than any other category.

To increase overall conversion rate, we need to increase the number of Working Professional leads by reaching out to them through different social sites such as LinkedIn etc. Also, we need to work on increasing the conversion rate of Unemployed leads.

4. 'Tags', 'Will revert after reading the email ', 'Lateral student' and 'Closed by Horizon' have high conversion rate.

5. The count of leads from the Google and Direct Traffic is maximum. The conversion rate of the leads from Reference and Welingak Website is maximum.

To improve the overall lead conversion rate, we need to focus on increasing the conversion rate of 'Google', 'Olark Chat', 'Organic Search', 'Direct Traffic' and also increasing the number of leads from 'Reference' and 'Welingak Website'!

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.