

Lead Score Case Study Report

Kaushal Nagrecha | Anant Sawant | Abhinav Pandey

Summary -

Problem Statement:

- **X Education** wants to **increase** its lead conversion from **30% to 80%**.
- The company wants to focus more on **'Hot Leads'**.

Objective:

- Design a model to assign lead scores based on the conversion likelihood.

Steps followed:

1. Data Cleaning -

The data was pretty clean, but there were some null values and some bogus `Select` values. In the data cleaning step, the null values were dropped, and the `Select` values were replaced by NaN; and all the NaNs and the NAs or the null values were then replaced with `not provided`.

2. EDA -

Relationships between variables was looked out for; a lot of categorical variables turned out to not be useful enough.

3. Dummy Variables -

Categorical variables with high correlation were converted into dummies, out of which the `not provided` ones were discarded.

4. Model Building -

1. Train Test Split -

70-30 Train-Test Split was made

2. Model Fitting -

RFE was carried out to choose limited number(15) of features to build the model on, then manually other irrelevant features were removed using VIF and p-values(VIF<5 and p-value<0.05 were kept, other variables were discarded)

5. Model Evaluation -

Confusion Matrix was made. Optimum cut off point was determined using the ROC curve to be 0.35, using which the values for accuracy, sensitivity and specificity were calculated to be ~80% each.

6. Prediction -

Optimum cut off of 0.35 was used to make the predictions on the test data

7. Precision and Recall -

Using Precision and Recall, a new optimum cut off value was found of 0.41, with a Precision value of 73% and a Recall value of 76%.

Findings -

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

1. The total time spent on the Website.
2. Total number of visits.
3. When the lead source was:
 1. Google
 2. Direct traffic
 3. Organic search
 4. Welingak website
4. When the last activity was:
 1. SMS
 2. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

Keeping these in mind **X Education** can identify the `**Hot Leads**` and maximize their conversion rate.