

SUMMARY REPORT

Problem statement:

Identify the set of leads of X Education so that the lead conversion rate should go up and the sales team of the company focus more on communication with the potential leads rather than making calls to every customer.

Model outcome:

Optimum probability cut off: Around 0.35

Lead Score:

The lead score is calculated based on the probability of customer being converted. According to the final model, if the lead score is more than 35, then the customer is likely to be converted. Higher the lead score, higher the chance the lead/customer being converted.

Average Lead Score of the converted leads = 69

Average Lead Score for the not converted leads = 14

Features of the final model:

The significant features of the final model will be impact on the basis of their coefficients. Lead Source WELINGAK Website, Lead Source Reference, Total time spent on website has the highest coefficient, so will give a Positive Impacts. Where on the flip side Lead Score organic search and Lead Source Google has the lowest coefficient, will show high positive impact.

Important measures of the model:

On Train Set:

Accuracy: 0.808

Sensitivity: 0.815

Specificity: 0.804

On Test Set:

Accuracy: 0.813

Sensitivity: 0.805

Specificity: 0.819

Gini of the model: 0.88

Conclusion:

The model has good accuracy, sensitivity and specificity. Overall, the model performs well in the test set, what it had learnt from the train set.

Business recommendation for higher conversion rate:

Highly likely to be converted leads:

1. Lead score more than 68.
2. Total time spent on website more than 12 hrs.
3. Lead source WELINGAK Website and Reference.

Very less likely to be converted leads:

1. Customers opted for 'Do not email' option.
2. Lead score less than 15.
3. Total time spent on website less than 5 hrs.
4. Lead source Direct Traffic, Referral Sites, Organic Search and Google.

5. Last activity of the customers is any of 'Olark chat conversation', 'page visited on the website', 'Email bounced', 'Form submitted on website', 'Email link clicked'.

Learnings gathered:

1. Data preparation for modelling

- It is important to treat missing values and also get rid of the outliers present in the data.
- If there is huge data imbalance in the features, then it is better to either drop that particular feature or remove the imbalance by merging the imbalanced values to other values.
- All the features should be in the same scale.

2. Model building

- There shouldn't be any multicollinearity between the variables.
- Find the optimal probability cut off to get a balance between Sensitivity and Specificity with good Accuracy.
- The model should perform well in the test set in terms of Sensitivity, Specificity and Accuracy