

Summary of Lead Scoring Case Study

MEMBERS

GOPIKA, NAGARJUNA, SHWETHA

CONTENTS

Summary	2
Steps Followed	2
1. Data Cleaning.....	2
2. EDA	2
3. Dummy variables	2
4. Train-Test Split	2
5. Model building	2
6. Model evaluation	2
7. Prediction.....	2
8. Precision recall	2
Final Analysis	3
Suggestion	3

SUMMARY

This analysis is done on X education's lead data to find out ways to improve the online course sales. The data provided gave us a fair understanding about the potential clients background, their conversion rate etc.,

STEPS FOLLOWED

1. DATA CLEANING

The data provided contained few null values which we changed it to 'Not provided' to save the data from losing wherever possible and were later removed during dummy process.

2. EDA

The EDA was carried out to check the relevancy of the categorical variables, outliers, etc.,

3. DUMMY VARIABLES

The dummy variables were created and in the process we removed the 'Not provided' elements that were created during the data cleaning process

4. TRAIN-TEST SPLIT

The split was done at 60% and 40% for train and test data.

5. MODEL BUILDING

RFE was done to attain the top 15 relevant variables and rest of the variables were removed depending on the VIF values and P-value (VIF = <5, P-value<0.05 were kept)

6. MODEL EVALUATION

A confusion matrix was created and ROC curve was used to find the accuracy, sensitivity and specificity of the data. (80%, 69% & 87% respectively)

7. PREDICTION

Prediction was done on the test data frame with the cutoff as 0.4 with accuracy, sensitivity and specificity of ~80%

8. PRECISION RECALL

This was used to recheck and a cutoff of 0.41 was found with precision around 78% and recall around 69% on the test data frame

FINAL ANALYSIS

The following variables mattered the most in the potential buyers/ Conversions:

1. The total time spent on the website
2. Total number of visits
3. Lead source:
 - a. Google
 - b. Direct traffic
 - c. Olark chat
 - d. Organic search
4. Last activity:
 - a. Olark Chat
 - b. SMS
5. Lead origin is Lead add Format
6. The current Occupation is Working Professional

SUGGESTION

1. The sales team should focus more on the profiles matching the criteria stated in the analysis to have maximum conversions and flourish the sales chart.
2. Make the website more interactive so that the potential buyers who spend more time on website have higher chance of buying the courses
3. Make the Olark chat process and SMS more engaging to have a higher conversion rate