

SUMMARY

This analysis is done for X Education and to find ways of getting to more professionals in the industry join their courses . The basic data provided us with a lot of information on how the customers potential visit the site, the time they spend there , how came to the site and rate of conversion .

The following are the steps used

1. DATA OF CLEANING :

The data were partially clean , except for a few values null and choice of selection had to be replaced by a null value as it didn't give us much information .

Few of the value zero is changed to 'not provided ' for not losing a lot of data . Although then they were removed while making mannequins . Since there were many from India and few from outside , the elements were changed to 'India', ' Out of India' and 'no planned ' .

2. EDA:

A quick EDA was performed to verify the status of our data .It was discovered that many the elements of categorical variables were irrelevant .The values numerical look good and not found values atypical .

3. VARIABLES DUMMY :

- The dummy variables were created and then the dummy variables with 'not provided '
- Items were removed . For numerical values we use Min Max Scaler.

4. TRAIN-TEST DIVISION :

The division was made at 70% and 30% for the data of the train and test , respectively .

5. CONSTRUCTION OF MODELS :

In the first place , it is made RFE to achieve the 15 main variables relevant . More later the rest of the variables were eliminated manually in function of the values of FIV and the value p. (The variables were maintained with IVF <5 and p value <0.05).

6. EVALUATION OF THE MODEL :

It is made one matrix of confusion .
Subsequently , the value of cutting optimal (using the curve ROC) was used to find the precision , sensitivity and specificity that emerged 80% each .

7. PREDICTION :

The prediction is performed in the frame of data for testing and a cut optimum of 0.35 with 80% precision , sensitivity and specificity .

8. ACCURACY - RECOVERY :

This method also is used to return to verify and found one cut of 0.41 with Precision around of 73% and remember about of 75% in the framework of data for testing .

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

- Total time spent on the website.

- Total number of visits .
- When the source of lead was:
- Google
- Direct traffic
- Organic search
- Welingak website
- When was the last activity :
- SMS
- Conversation chat Olark

When the origin of the customer potential is the format of addition of client potential .When your occupation today is as professional in assets.

Taking this into account , education X can prosper , as they have a very high probability of obtaining almost all the buyers potential change of opinion and buy their courses .