

SUMMARY

This Analysis is for X Education Company to find more industry relevant professionals to get the course. These kind of lead generate from google, websites, references etc. so company is going to check these lead convert them in to hot lead and finding the probability (conversation rate) of the customer choosing the course based on technical and theoretical analysis

Steps are used in analysis:-

1. Cleaning Data sets:

In Data cleaning part finding the unique values and remove those values , finding missing values which have greater than equal to 40 drop those values. Finding outlier and treatment, imputing missing values in those columns which are useful for further analysis.

2. EDA:-

Here lot of elements are categorical variables, convert those object datatype elements into int64, float and unit8 etc. Univariate analysis, bivariate analysis of numerical and categorical data sets and find the insights from graph.

3. Dummy variables And test-train split:-

To get dummies for numerical values I use minmaxscaler & split the data sets into 70-30% train-test ratio.

4. Model building :-

The process of building the model based on main two factors vif and p values, vif values is must below the 5 and p value is 0 which show significant of variables.

5. Model Evolution:-

For model evaluation find confusion matrix, find sensitivity, specificity and find optimal cut-off of these three values and generate roc curve.

6. Prediction:-

Making prediction on data sets is useful for finding the cutoff values for prediction and is 0.38

7. Precision-Recall:-

This is trade off values of precision and recall is 0.42 which gives the precision value is 71% and Recall Values is 76%.

Conclusion:-

1. most of the time spends on website
2. last activity sms , email, chat conversation of most of the lead
3. choosing course because of better career options
4. lead generation from google , olark chat , organic searches , references
5. landing page from submission
6. Most are working professionals etc.
7. Lead score value is useful for better lead conversation rate and then contact them through email, phone call , sms etc.

