

SUMMARY

- X Education is a online learning platform which provide courses to industry professionals. The Company do marketing of there courses on various online platform and websites.
- The user who ever visit these platform they are redirect to the company official website for more detail and filling form for enquiry . Once the user (customer) fill the form they are marked as LEAD.
- The marketing team then contact these LEAD and give them offers to enrol them in courses.
- The Company currently has a LEAD conversion rate of 30%
- The Company has a target of LEAD conversion to 80%.
- The Company want to know the lead score of the user (customer) so that they can build some strategy to achieve there goal.
- We are provided with the company past data and our job is to build a logistic regression model to find out factor responsible for lead conversion .

Steps are being taken during the model creation.

1. **Data Cleaning:** In these we have cleaned the data by removing the null values and we have the option select which had to replace with nan- values as it not hold much important information so we later provided or change value with "Not given" to not loose data.
2. **EDA :** First we check the data imbalance in dataset here we observe that 38% of data is of Leads ones 62% of data is of Non-Leads .In EDA of categorical data we found most of the column are irrelevant, and no outlier or sudden fluctuation in data is not present in data when check the numerical data.
3. **Dummy variable:** Create dummy variable of all categorical variable and then rescale the numerical type variable to scale of 0 to 1 with help of Minmax- Scaler.
4. **Train-Test Split:** Here we have split the data into 70:30 ratio and used 70% of data for model training and 30% of data for model testing.
5. **Model Building :** Here first we have select top 15 variable based on there importance and ranking by using RFE method and than build the model having P-value < 0.05 and VIF < 5 by eliminating the variable .
6. **Model Evaluation :** First Confusion matrix is made to check model accuracy score, sensitivity and specificity and find out the optimum cut-off point based on ROC curve based on that ,found prediction on test set where we get sensitivity 78% and specificity 83% . Then also found the cut-off of Precision-Recall Curve which found about 0.41 and based on this cut of found the recall score of 76% on test data set

Business Recommendation:

So the top 5 variable which is responsible for lead conversion the most are:

- 1.Total Visit
- 2.Total time spent on website
- 3.Last activity—(Phone conversation)
- 4.Lead origin—(Lead add form.)
- 5.Current occupation—(Working professional)

So the X Education company should focus of these variable to increase there Lead Conversion rate and to achieve there goal of 80% lead conversion rate.