

Lead Scoring Summary

The X Education is targeting to get more industry professionals to join their course. The company has provided the data which provides lots of information about the leads who are likely to be converted. To Build list of the leads which can be used by the company to target potential leads.

Below mentioned are the steps followed:

1. **Data loading, Cleaning and Inspecting:** The data provided was clean to some extent. The 'select' option has to be replace with Nan as it could not give much information. Columns with more than 40% missing values are being dropped. And other columns with missing values are treated by imputing suitable values.
2. **Exploratory Data Analysis:** EDA has been performed on the data and checked each and every variables contributing to the conversion rate of leads. And how likely they are to get converted.
3. **Dummy Variables:** Converted Binary variable 'yes' or 'no as 1 or 0. Dummy Variables where created. MinMaxScaler is used for numeric values.
4. **Train Test Split:** The data was split as 70% and 30% as Train and Test data respectively.
5. **Model Building:** Feature selection was done using RFE to attain top 15 variables. Then rest of the variables were dropped based on p-value and VIF value.
6. **Model Evaluation:** The Confusion matrix was made . With cut off as 0.5 we have our Accuracy around 81% . Further on the optimal cut off value using ROC curve arrived at the accuracy, specificity and sensitivity was around 80% each.
7. **Prediction:** Prediction was done on test dataframe with optimal cutoff as 0.35 with accuracy, specificity and sensitivity was around 80%.
8. **Precision – Recall :** Trade off between precision and recall is 0.42 with precision = 79% and Recall = 71% on test dataframe.

The following are the variables which are to be considered to in most potential leads:

1. Lead sources : 'Welingak Website' and 'Reference'.
2. Leads should be Working Professionals
3. Leads whose Last Notable Activity is SMS Sent and olark chat conversation
4. Leads who spend most of the time on Website
5. Leads whose Last Notable Activity is Email Opened
6. Leads whose Last Activity_Had a Phone Conversation

We have to concentrate on above mentioned variables to convert a potential lead into course buyer.

To improve overall lead conversion rate, we need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.