

Summary

Problem Statement:

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads.'
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Steps involved:

1. **Data Cleaning** - Data cleaning was done by dropping unwanted columns, handling duplicate values, handling null values and imputation of the values.
2. **Exploratory Data Analysis (EDA)** - EDA was performed to analyse the data. Univariate analysis was done by plotting boxplot on numerical columns and Bivariate analysis was done by plotting boxplot on both numerical and categorical columns. Finally, Heatmap was used to check the correlation between different variables.
3. **Data preparation** - Columns which have Yes, No values are mapped with 1,0 values. For categorical columns which has more than 2 options dummy variables were created dropping original column. For numeric values MinMaxScaler was used.
4. **Train-Test split** - The data was split into 70% and 30% for train and test data respectively.

5. **Building the model** - RFE was done to get the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value. Using GLM the model was built.
6. **Prediction** - From the final model probability of lead conversion is predicted, the lead is classified as Converted or not based on probability, if the probability is greater than 0.5 then lead is classified as converted else not converted. Accuracy of the model is calculated, which came to 77%
7. **Precision – Recall** - This method was also used to recheck and a cut off of 0.41 was found with Precision around 73% and recall around 75% on the test data frame.

Thus, X Education Company needs to focus on following key aspects to improve the overall conversion rate:

- Improve the user engagement on their website to help in higher conversion.
- Increase on sending SMS notifications since this helps in higher conversion.
- Get Total Visits increased by advertising.
- Improve the Olark Chat service since this is affecting the conversion negatively.