# Summary of Lead Score Case Study

By Balakrishna

Archana &

Ananth

Data soucing, data cleaning, EDA , Model Building, and Model Evaluation has been done for the company X Education to find out the ways to convert the potential leads or users. To target the particular set of group to increase the conversion rate, we have followed the below important steps sequentially.

1. Firstly we need to import all the necessary python and other libraries to read and analyse the data.
2. After we read the data, we check the dimension of the data, and the datatypes. This step is called data inspection.

**Data Cleaning:**

1. After the detailed inspection of the data, we observe the high percentage of null values for some of the attributes. We have dropped the columns which are having more than the 40% of null values.
2. We can see that there are 37% of missing values in Specialization, here lead might find blank so left the cell as it is with ‘Select’. We created Other category for this.
3. Similarly 36% of missing values have imputed for the Tags column. After clearing the missing values section moved to EDA.

**EDA**:

1. Started with Univariate and Bivariate analysis. Found that converted is the target variable, here we have 1 which means successfully converted, 0(zero) means not converted.
2. With the help of Univariate analysis, we have dropped some of the unnecessary columns.

3. We have created dummy variables and removed the attributes for which we have created the duplicates.

**Model Building & Evaluation:**

1. We calculated the generalised linear regression results after we import the statsmodel.api library. We have done this for 9 models, and 9th model is final one with final 12 variables after we reduced with the help of VIF and P-Value. Here we have used RFE for feature selection.
2. Confusion matrix was created and the overall accuracy was checked and it is around 81%.
3. Below are the accuracy, sensitivity, specificity for both trained and test sets.

   **Train Data:**

   i. Accuracy : 81.0 %
   ii. Sensitivity : 81.7 %
   iii. Specificity : 80.6 %

   **Test Data:**

   i. Accuracy : 80.4 %
   ii. Sensitivity : 80.4 %
   iii. Specificity : 80.5 %

4. The optimal cut off for both sensitivity, specificity and Precision and Recall are 0.34 and 0.44 respectively.

This model seems build with a good prediction rate and should be able to give confidence to the company for making calls based on this model.