

## Summary Report of the Approach.

### Business Objective:

- Help X Education to select the most promising leads (Hot Leads), i.e. the leads that are most likely to convert into paying customers. Currently typical lead conversion rate is 30%
- To build a logistic regression model to assign a lead score value between 0 and 100 to each of the leads. CEO of the company gives ballpark of the target lead conversion as 80%

### Approach

Following is the approach performed to reach the goals

- Uploading the Data
- Analysis of the data set like columns properties etc,
- EDA
- Data preparation/ Modelling
- Feature selection using RFE
- Model Building
- Predicting Lead conversion probability in Train dataset.
- Finding optimum threshold probability
- Plotting ROC
- Evaluating model on train dataset
- Making prediction on test dataset
- Evaluating model on Test dataset
- Lead score calculation on complete
- Determining important features.

### Matrics of Train Dataset:

- Accuracy: 92.22%
- Sensitivity: 88.21%
- Specificity: 95.13%
- Precision: 88.47%
- Recall: 91.69%

### Matrics of Test Dataset:

- Accuracy: 92.78%
- Sensitivity: 91.98%
- Specificity: 93.26%
- Precision: 89.15%
- Recall: 91.98%

**Important Features from the Dataset:**

Model co-efficient are used to determine the important features. Higher the value of the co-efficient higher the importance. Below are the top 3 features.

1. **Tags\_Closed by Horizzon**
2. **Tags\_Lost to EINS**
3. **Tags\_Will revert after reading the email**

**Challenges:**

1. Data skewness. Also, there were lot of null values and we have to individually check the columns are replace null values (impute the null values).
2. Accurate predictions in test data set.

**Learnings:**

1. Analysis of the dataframe columns are important as it helps us to get the insight of the data.
2. EDA is very important as doing EDA properly helps in removing the null values also it helps us in visualizing the data better
3. If we don't have enough insight of the columns, then we should let the algorithm decides according to p-value and multicollinearity of the columns if we need to drop the columns or not.
4. All metrics important to properly evaluate the model.
5. Business knowledge and requirements understanding is very important.