

## **Summary**

### **Problem Statement:**

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The process involves constructing and validating a predictive model for company X Education with the goal of identifying strategies to convert prospective users. By thoroughly examining and validating the data, we aim to draw conclusions that will guide us in targeting the right audience and ultimately enhance the conversion rate.

Let us discuss steps followed:

#### **1. EDA:**

- Quick check was done on Percentage of null value and we dropped columns with more than 40% missing values.
- We also saw that the rows with the null value would cost us a lot of data and they were important columns. So, instead we replaced it with mode and median.
- Then dropped the columns which are not giving any insights
- We also worked on numerical variable, outliers and dummy variables.

#### **2. Train-Test split & Scaling :**

- We split the data as 70% and 30% for train and test data respectively.
- We will do min-max scaling on the variables Like 'TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website'.

#### **3. Building Model**

- RFE was initially employed for feature selection.
- Following the RFE process, the top 15 relevant variables were identified.

- Subsequently, the remaining variables were manually eliminated based on their VIF values and p-values..
- A confusion matrix was created, and overall accuracy was checked which came out to be 81.01%.

#### 4. Model Evaluation

##### ▪ On Training Data

- The optimum cut off value was found using ROC curve. The area under ROC curve was 0.89.
- After Plotting we found that optimum cutoff was **0.34** which gave

Accuracy : 80.36%  
Sensitivity/Recall : 81.66%  
Specificity : 79.55%

##### ▪ Prediction on Test Data

- We get

Accuracy : 80.354%  
Sensitivity/Recall : 81.89%  
Specificity : 79.67%

## CONCLUSION

### TOP VARIABLE CONTRIBUTING TO CONVERSION:

- Total Visits
- Total Time Spent on Website
- Lead Origin:
  - Lead Add Form
- Lead source:
  - Direct traffic
  - Google
  - Welingak website

- Organic search
- Referral Sites

The model shows a strong capability to predict the conversion rate, giving confidence to the company to make informed decisions based on its predictions.