

# Lead Score Case Study

## Summary Report

### Problem Statement:

The 'X Education' company sells online courses to industry professionals and acquires leads through website forms, referrals, and marketing campaigns. Their lead conversion rate is low(30%), and they want to identify potential leads, also known as 'Hot Leads,' to increase the conversion rate. They require a model that assigns a lead score to each lead based on their likelihood of converting into paying customers. The target lead conversion rate is around 80%.

### Data Cleaning:

- The unique columns were identified and dropped the columns that has single unique values which is least important for the analysis.
- Some of the categorical columns had value as "select", which is equivalent to null values and handled it by replacing it as "Not Available".
- The columns containing null values were identified and dropped the columns having more than 35% null values.
- The rows that had least null values (less than 1.5%) were dropped.

### Exploratory Data Analysis:

The below inferences were inferred from Data Analysis,

- The probability of lead getting converted is high when 'Lead Origin' is from 'Lead add form'.
- The Large number of leads come from Google and direct traffic but the referral sites convert most of the leads.
- Leads opting for emailing option have more probability of getting converted.
- The Conversion rate is higher when the information is sent through SMS
- Unemployed people have more conversion rate as well as more count.

### Data Modelling:

- The dummy variables were created and the Train - Test split was done at 70% and 30% respectively.
- **Model Building:**  
RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value.
- **Model Evaluation:**  
A confusion matrix was made. Later, the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 80% each.
- **Prediction:**  
Prediction was done on the test data frame and got the optimal cut-off of 0.35 and the final model gave the accuracy of 80% with 79% sensitivity and 80% specificity.
- **Precision – Recall:**  
This method was also used to recheck and a cut off of 0.41 was found with precision of 79% and Recall of 69% on the test data frame.

# Lead Score Case Study

## Summary Report

### Conclusion:

- The model accurately predicts lead conversion likelihood, with an 80% accuracy rate, using relevant variables identified through RFE and evaluated using a confusion matrix and ROC curve analysis, making it suitable for identifying "Hot Leads."
- With the help of this model if the sales team contact only the leads with high lead score, their conversion rate will increase to more than 80% as expected.
- The probability of lead getting converted is high when 'Lead Origin' is 'Lead add form'
- Large number of leads come from Google and direct traffic, but referral sites convert most lead conversions.
- Leads opting for emailing option have more probability of getting converted.
- Conversion rate is higher when the information is sent through SMS.