LEAD SCORE CASE STUDY

BY Anjali Menon

SUMMARY REPORT

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.

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 where 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.

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.