Summary

The Model is built for an X education company to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The X education company provided dataset with variables. The Model identify the potential lead. The target variable is “converted” which represents the lead to be successfully converted (1) or not (0).

The following points are used for model building:

1. We imported the necessary library.
2. We checked and calculated the null values in each column and dropped those column which have more than 30% null values and not necessary for the analysis.
3. We performed the EDA. We did Univariate and bivariate analysis on relevant columns.
4. We checked the unique values in column and created dummy variables for them.
5. We divided the variable in Train and test dataset and scaled it.
6. The RFE was performed and we done manual selection on variables depending upon the P and VIF values.
7. We continue did model building till the value of VIF and p-value come under the values 5 and 0.05 respectively.
8. After finalizing the model, we calculated the confusion matrix.
9. We draw the ROC curve and found the optimal cutoff.
10. We calculated accuracy, sensitivity and specificity for various probability cutoffs.
11. Predication was done on the test data set with accuracy, sensitivity and specificity of 48.03%, 70.31 % and 88.05% respectively
12. Optimal cutoff 41% and recall is around 94%.

Conclusion –

As we can see that, below are the Potential leads which can be preferred to increase chances of conversion:-

1. Leads with maximum Total Visits
2. Leads with maximum Total Time Spent on website
3. Leads with Lead Source as Google
4. Leads withLast Activity as SMS
5. Leads with Current Occupation as Working Professional