Summary

This Analysis is used for X education to find ways to get more industry professionals to join their course. The basic data provided to us gave us information about the customers who visit the site and how much time do they spend on the site and their conversion rate.

data preparation was done, which involved the following steps:

1. Data cleaning and preparation
- Combining three data frames
- Handling categorical variables
- Mapping categorical variables to integers
- Dummy variable creation
- Handling missing values
2. Test-train split and scaling
3. Model Building
- Feature elimination based on correlations
- Feature selection using RFE (Coarse Tuning)
- Manual feature elimination (using p-values and VIFs)
4. Model Evaluation
- Accuracy
- Sensitivity and Specificity
- Optimal cut-off using ROC curve
- Precision and Recall
5. Predictions on the test set

Standardisation of the scales of continuous variables.

The top three variables in the model which contributed most towards the probability of a lead getting converted were Lead Source, Lead Origin, What is your current occupation. The dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion were Lead Source, Last Notable Activity, What is your current occupation.

The model built helped us to get the leads which has high possibility of conversion.

- The model shows high close to 81% of accuracy.
- Threshhold can be selected from Accuracy, Specificity, Sensitivity and precision, Recall curves
- Model shows about 75% sensitivity and 85% specificity.
- Overall this model proves to be accurate.