Summary

The current analysis by model building is done for an X-Education to find more probable leads to join their courses. The leads data provided states about the user profiles, occupation, source of origin, time spent, repeated visitors etc. Now let us discuss the steps followed and the conclusion reached by using the model created.

Data Cleaning: The data was checked for null and duplicate values and handled the missing values. The columns which were not useful for analysis or with more missing values was dropped after the EDA of the column. Imputation was done if necessary and also handled the outliers.

EDA: The quick EDA was done on each column before dropping them. Univariate Analysis is done to value count and distribution of variables etc. Bivariate Analysis used for correlation coefficients and the pattern between the variables. The logistic regression is used as classification technique for the model making and predicting.

Train-Test split: The split was done at 70% and 30% for train and test data respectively.

Model Building: The model is built by splitting the train and test sets into 70:30 ratio and ran RFE with the top 15 variables as output. Later the rest of the variables were removed based upon the VIF values more than 5 and and the P-value greater the 0.05.

Model Evaluation: The confusion matrix was made and later optimum cut off value (at 0.35) and ROC (AUC obtained is 89%) to find the accuracy, sensitivity and specificity which came to around 81%.

Model Prediction: The model prediction is done on the test data set with optimum cut off value as 0.35 and with accuracy, sensitivity of around 81%.

Conclusion: Based upon the model built, the following variable are to be considered to increase the probability of getting the conversion rate into customers: Total time spent on website, occupation targeting unemployed an

working professionals, Lead origin as Lead add form, Lead source as Google, direct traffic followed by organic search, Last activity as SMS sent, Olark chat, and email opened.