Lead Score Case Study Summary Report

X Education gets a lot of leads, its lead conversion rate is very poor. The typical lead conversion rate at X education is around 30%. X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Detailed summary report of Logistic Model:

➢ **Data Inspection and Cleaning:**

Provided data has 9240 entries with 37 Features, we have deleted columns with more than 40% on missing value.

For rest of the columns, we have imputed the missing values with required data so that it can be helpful for analysis. Checks for outliers and treat them by capping and flooring technique.

➢ **EDA and Data Preparation:**

Checked converted data imbalance, 38.53% converted leads found.

Performed Univariate and Bivariate Analysis where we gathered some insights from features like Lead Origin, Lead Source, Current Occupation, Total Time Spent, etc.

Then we have created dummy variables for categorical variable and map columns value with Yes\No to binary.

Split the dataset into Train and Test Data with 70- 30 ratio. Used feature Scaling for numerical data to normalise the value so that model interpretation will be easy.

➢ **Model Building and Evaluation:**

Used RFE technique for feature selection to select top 20 features out of 73 columns.

Then build a model and checks p-Value of all features, if it more than 0.05 simply drop that column one- by-one. Same step has been repeated till we get model where p-value looks a=significant and then calculate the VIF for reaming variables.

If the VIF value > 5 then dropped the column and again build the final Model. Mod8 is the final model with 12 features.

Then checked accuracy score which comes around 80%, Build the ROC curve with 0.88 curve area and gives cut-off at 0.37.

With this cut-off accuracy, specificity and sensitivity are around 80% and we have also calculated Precision and Recall cut-off that gives value less than 80%. Hence, we stick with optimal cut-off 0.37 and create Lead Score as per the cut-off.

➢ **Making Prediction on Test Dataset:**

Feature scaling and predict the test data based on final model i.e, mod8. Evaluate all the scores which is similar to train data i.e, 80%

Assigned Lead score to test data as well.

➢ **Conclusion and Recommendation:**

With out final we can find out top-features that are impacts leads conversion into Hot-Leads are:

- CurrentOccupation_Working Professional
- LastActivity_Had a Phone Conversation
- LastActivity_SMS Sent
- Total Time Spent on Website

Company should focus more on Working_Professional, there are high chances for them to convert into Hot Leads.
Also Leads who spent more time on Website or Had a Phone conversation or who has sent SMS are more likely to convert into Hot Leads.
Company should spend more on advertising as it can been seen some of the promising leads converted through Weblinks. And also promote some offers for Referral Leads.