Lead Scoring
Case Study - Summary

By: Viral Shah, Vinoth, Vineela

Summary Report:

The detail report of how, we have performed the Lead Scoring Case Study.

We were given a problem where we had to build a model where the customer with higher lead score have higher chances of conversion. And lower lead score had lower chances of conversion. This would help the sales team to make a right call to the hot leads.

The Approach

Data Cleaning:

The data was required cleaning and also procession. After which we did analysis of columns to understand shape, data types, and content / quality of the data. One of the conditions which was given is converting 'select' values to 'NaN' as they don't serve any purpose. The columns that only had one value were considered as the column has no variance and won't contribute to final model.

Missing Value and outlier treatment:

The missing values of each column were analyzed and identified. The columns which had more than 25% missing values were not considered relevant for the model building; hence they were removed. Similar way, outlier analyzed for the numeric variables.

Binary Mapping:

The columns which have 'yes' and 'no' values had been converted to binary.

Model Building:

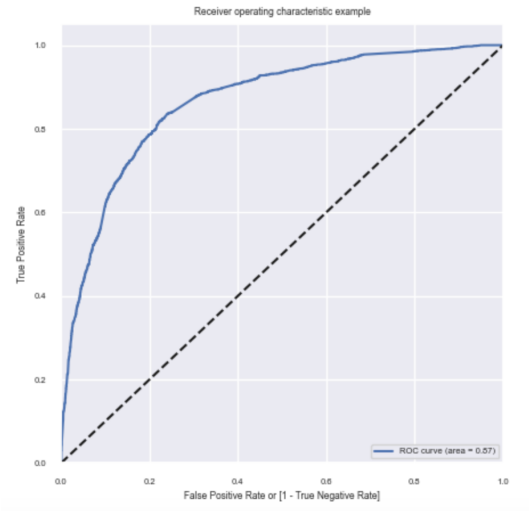
As the standard process, the data was split into test and train. Later, numeric variables excluding binary variables had been scaled using standard scaler. Recursive feature elimination was used to select top features. The P-value and VIF was analyzed to identified column which helped to identify columns which doesn't add much values. Which we had removed it as part of

the model building process. Later the model was re-build to provide accurate results.

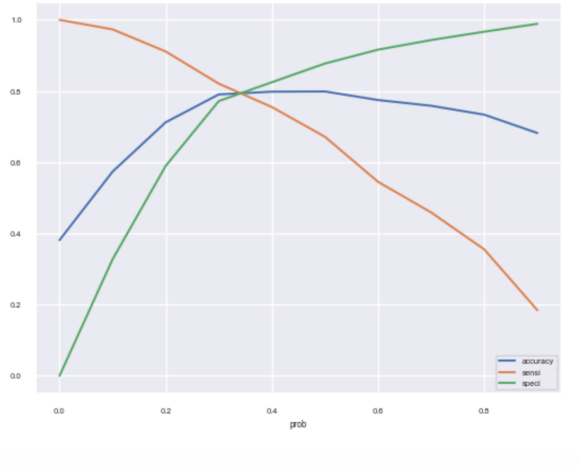
Evaluating Model:

We plotted the ROC curve to understand the performance of the model. We checked the specificity, sensitivity, precision and accuracy. Once the data found to be satisfactory, we checked the test data. And the result of the test data was similar to the train data. Almost similar. Around 80%.

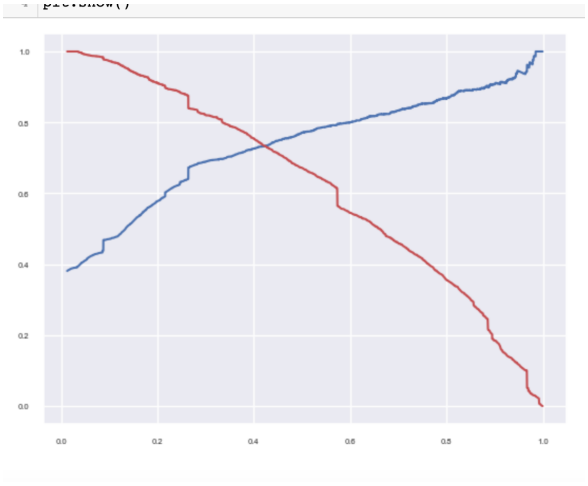
Ref plot images:



ROC Curve



Sensitivity – Specificity Plot:



Precision – Recall Plot: