LEAD SCORING CASE STUDY

Brief Summary of Case Study

Following steps were followed for the data analysis & model building:

1. Understanding the business problem & the dataset given:

- Before stepping into any cleaning & analysis part, we tried to understand the business problem first. This step is important as understanding the problem can give us clarity for problem-solving.

2. Data Cleaning:

- After importing the required libraries, we checked out the missing values in the datasets.
- After analyzing these columns we found that there are more than 70% of missing values and are not required for analysis. Therefore we decided to drop these columns.
- The remaining missing values were imputed for both the Continuous & Categorical Variables.

3. Exploratory Data Analysis:

- In this step, using the data, we made preliminary assessments about the population distribution of the variables.
- In univariate analysis, we assigned numerical variables to categories with 'yes' to 1 & 'no' to 0.
- 'Converted' is our target variable.
- Various inferences were drawn by plotting them.
- Based on univariate Analysis we observed many columns are not adding any information so we choose not to use this column for further analysis and we dropped them.

4. Data Preparation:

- Created dummy variables for some of the categorical variables.
- We split the dataset into train dataset, test dataset & scaled dataset.
- To check the various correlations between the variables, heatmaps are plotted & after finding some correlations, some of them were dropped.

We then standardized the independent features into fixed variables. This term is known as feature scaling.

5. Model Building :

- We created a model, assessed it using StatsModels with an RFE count of 15.
- Obtained the predicted values on the training dataset.
- We have checked optimal probability, accuracy, specificity, sensitivity, and also cut off by finding points
- We plotted a ROC curve & concluded that:
    - The curve is closer to the left side of the border than to the right side hence our model is having great accuracy.
    - The area under the curve is 80% of the total area.
- We also checked the precision and recall value from our final model and we took a tradeoff.
- From the curve, 0.4 is the optimum point we got.
- In the next step, a Prediction was made on the test set and we have found the predicted value.
- After finding the predicted value we got inferences that from the final test model accuracy sensitivity and specificity was in an acceptable range.
- We checked the precision & recall tradeoff using sensitivity, specificity & accuracy.

6. Conclusion :

- Important features we got during training of model which is responsible for good conversion rate are as follows
    - Total Time Spent On the website
    - Lead Source_Reference
    - Lead Source_ Social Media

- We got the Recall value greater than Precision Value also it is acceptable for the Business aspect**.**
- Sensitivity, Specificity, Accuracy we got from the test data set when compared with train data set it is in an acceptable range**.**
- The model can adjust as per company requirements and it will give a good result.