LEAD-SCORING-CASE-STUDY-SUMMARY:

In the following case study a company named X Education wanted to make their lead conversion system more efficient. Earlier the conversion rate was around 30% which is needed to be pushed 80%, For this we need to analyse data to find most potential leads and work on them rather than wasting the workforce on not so useful ones. So basically, we started with the data inspection and then we moved to data cleaning in which we first checked total null values in every column ,after that some columns which provide similar(no unique values) information or no information were dropped like 'Asymmetrique Activity Score', 'Prospect ID' etc., then some rows too are removed which had too many values, Also the select and null values were made to a different category called unknown in this way we can juice out the data from these columns. Columns like total visit were handled by imputing median, also using boxplots outliers were removed from the data. Converted some columns like 'Search', 'Do Not Email' to binary values. After that dummy variables were created and unknow values were removed from it for the analysis. Train and test split on the data was performed and applied 70 -30, scaling of features such as 'totalvisits' was done .Then we build the model for logistic regression and performed RFE to select features, calculated variance inflation factor for the RFE selected features, after getting a good VIF we trained the model checked for accuracy (which should be above 80%), did check precision and recall too, created confusion matrix and then moved to check optimal value for classification using ROC and checked the accuracy of the model too, Got the optimal value as 0.35 for deciding the probability of conversion. Now we tested our model using test data and checked again for recall ,precision .Accuracy of the model was 80% Which was the target (or close to) Top features of our model were lead source ,occupation and leads profile which helped us in determining the most about the leads conversion score. Hence probabilities for the conversion were calculated hence our lead score for converting leads was calculated.